Influence of Interactive Visualizations on Students' Learning Experience in Machine Learning

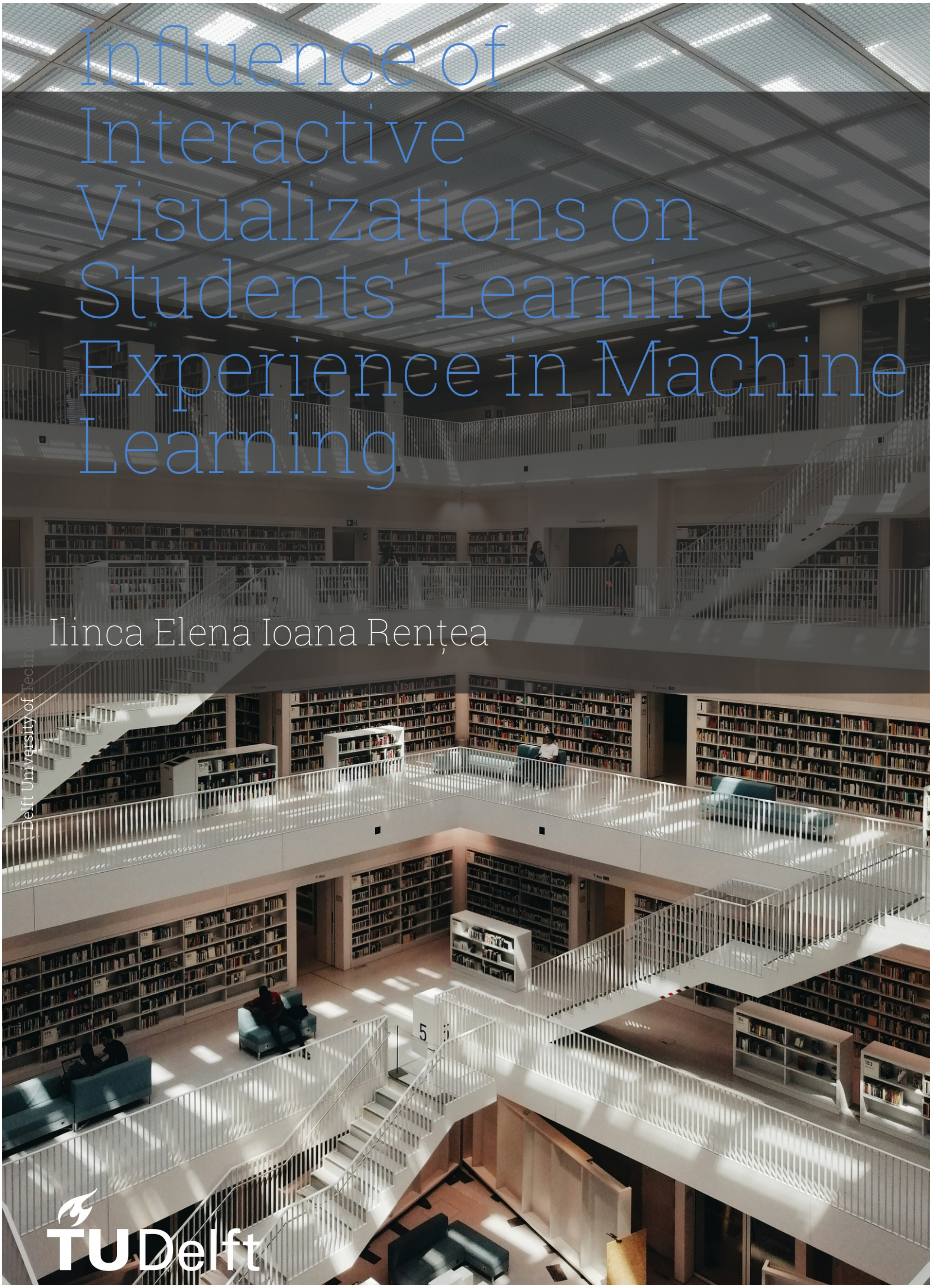
Ilinca Elena Ioana Renţea

Delft University of Technology

Influence of Interactive Visualizations on Students' Learning Experience in Machine Learning

Ilinca Elena Ioana Rențea

Delft University of Technology



Influence of Interactive Visualizations on Students' Learning Experience in Machine Learning

Ilinca Elena Ioana Rențea

to obtain the degree of Master of Science at the Delft University of Technology, to be defended publicly on Thursday June 27, 2024 at 9:00 AM.

Student number: 5058910 Project duration: November 13, 2023 – June 27, 2024 Thesis committee: Prof. dr. M. M. Specht, TU Delft, supervisor Dr. G. Migut, TU Delft, supervisor Dr. J. H. Krijthe, TU Delft

by

Influence of Interactive Visualizations on Students' Learning Experience in Machine Learning

by

Ilinca Elena Ioana Renţea

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday June 27, 2024 at 9:00 AM.

Student number:	5058910	
Project duration:	November 13, 2023 – June 27, 2024	
Thesis committee:	Prof. dr. M. M. Specht,	TU Delft, supervisor
	Dr. G. Migut,	TU Delft, supervisor
	Dr. J. H. Krijthe,	TU Delft

Contents

Machine learning(ML) is one of the most encountered topics and is a fast-growing technical field these days. It influences the field of computer science, but also healthcare, agriculture, education, media, and many others. Due to its growing popularity, much of the effort in the ML community is focused on ad- vancing the field through application development and research, leaving ML pedagogy and educational research less explored. Unlike more traditional subjects such as mathematics, physics, and statistics, which have well- established teaching methods and extensive research supporting their educational practices [6, 19, 53, 17, 21, 30], ML education lacks standardized instructional strategies. This gap highlights the need for dedicated research into effective ML teaching methods. ML topics are inherently complex and rely heavily on abstract mathematical concepts, making them challenging for students to understand [26]. For this reason, traditional teaching methods might not be sufficient, highlighting the need for research in this direction. One promising approach to address this challenge is the use of interactive visualizations. These tools allow students to manipulate and explore graphical representations of data and concepts, leading to a deeper understanding [10]. Interactive visualizations have been successfully used in subjects such as calculus [57, 27, 48], geometry [41], optimization problems [3], information retrieval [10], business analytics [55], and computer algorithms [11]. Since all of these subjects are science-related subjects, having elements in common with machine learning, the effectiveness of interactive visualizations in increasing understanding might transfer to ML as well. More than that, engaging students in the learning process is crucial for effective education [28], and it can be achieved using interactive visualizations. These tools can increase student engagement by fostering active learning, leading to a better understanding of complex topics [44, 43, 41]. Also, they encourage students' exploration and experimentation, resulting in deeper cognitive processing and improved learning outcomes, as supported by the ICAP framework [12]. Now that the potential of using interactive visualizations in teaching ML has been established, it is relevant to detail our research direction more clearly. The main focus of this research is to answer the following research questions:

1. How does the use of interactive visualizations in machine learning education affect students' knowledge gain on machine learning topics? 2. How does the use of interactive visualizations in machine learning education affect students' mo- tivation regarding machine learning topics?

The initial hypothesis is that the interactive visualizations will aid the students in understanding the machine learning concepts, based on similar previous research. More specifically, interactive visualiza- tions are believed to lead to a higher knowledge gain and motivation, than the ones achieved through using static visualizations. The novelty of this research lies in developing and empirically evaluating interactive visualizations focused on teaching ML. To illustrate their impact, two non-trivial ML topics were chosen: gradient de- scent and principal component analysis(PCA). On one hand, Gradient descent is a non-trivial concept

Introduction

1

with existing visualizations that lack formal evaluation [31, 62]. On the other hand, PCA is a non-trivial topic, previously recognized as difficult to teach[65]. The thesis includes a scientific article, presented in chapter 2, and supplementary materials, pre- sented in chapter 3. The scientific article is limited to the structure of such an article and its limitations. Therefore, chapter 3 aims to help the reader better understand the process of the research, by providing a more thorough literature review of the research direction, detailing the topics chosen for visualization, the development process for the visualizations, and the expert evaluation of the visualizations. Lastly, the appendices cover additional materials developed during the project.

Scientific Article

3

Are interactive visualizations in machine learning education helping students?

Ilinca Rențea Delft University of Technology Delft, The Netherlands

ABSTRACT

With the fast integration of Machine Learning(ML) into several in- dustries, the motivation to develop effective pedagogical strategies for teaching this complex and evolving field has become critical. Machine Learning, once mainly a topic in Computer Science Bach- elor programs, is now widely integrated into various majors and introduced at earlier educational stages, including high school and secondary school. However, the reduced research focus on ML ped- agogy results in a lack of standard teaching methods compared to other science-related subjects, which have established norms for topic introduction, teaching tools, and assessment methods. With inspiration from other fields, this research aims to look into us- ing interactive visualizations in teaching ML topics, more specifi- cally in teaching Gradient Descent and Principal Component Anal- ysis (PCA). The research includes the development and evaluation of Jupyter Notebooks for introducing these visualizations to stu- dents. The targeted student population is composed of Computer Science and Engineering Bachelor students who have not yet fol- lowed any Machine Learning courses but have the necessary back- ground knowledge, namely calculus, linear algebra, and statistics knowledge. The evaluation of this teaching method measures the knowledge gained and the motivation of the students, compared to a static version of the materials. The results of the evaluation have shown a significant effect of interactive visualizations on knowl- edge gained related to PCA. The evaluation has not identified any difference in knowledge gain for gradient descent and learning motivation for both topics. With these results, we contribute to the body of evidence-based teaching methods in Machine Learn- ing and identify further research needed to generalize the effect of interactive visualizations as a teaching method for teaching ML basic concepts.

CCS CONCEPTS

• Applied computing →Computer-assisted instruction.

KEYWORDS

machine learning, education, interactive visualizations, knowledge gain, motivation, controlled experiment

1 INTRODUCTION

Machine learning (ML) is one of the most rapidly growing techni- cal and research topics at the moment [20], having applications in various fields, such as healthcare [40], agriculture [23], education [21], and many more. Due to its popularity, much of the effort in the ML community is focused on advancing the field through ap- plication development and research, reducing the emphasis on ML pedagogy and educational research.

Gosia Migut

Delft University of Technology Delft, The Netherlands

Given the evident need for this research, displaying what teach- ing machine learning typically involves is crucial. Learning ob- jectives associated with this subject are understanding machine learning algorithms, implementing and applying them in spe- cific use cases, evaluating the performance of such algorithms, and analyzing their performance and limitation, as suggested by [44]. However, these learning objectives can differ depending on the intended audience, since the knowledge can be applied in dif- ferent manners and contexts [41]. More than that, machine learn- ing topics are considered to be difficult because they rely on math- ematics and abstract concepts[15]. On one hand, teaching non-trivial mathematical concepts has been accompanied by using interactive visualizations. These are graphical representations that allow users to manipulate and ex- plore the visualized information. The interactivity of visualizations enables the exploration of the underlying knowledge, resulting in a deeper understanding[8]. These tools were used in mathemati- cal subjects such as calculus[16, 31, 38], geometry[24] and opti- mization problems [4]. Besides being effective in teaching mathe- matics, interactive visualizations have also been applied in teach- ing information retrieval[8], business analytics[37], and computer algorithms[9]. On the other hand, students' engagement in the learning pro- cess is crucial for effective education[17]. In this context, interac- tive visualizations have been used to increase students' engage- ment, since they foster active learning, leading to a better under- standing of complex topics [24, 26, 28]. Additionally, interactive tools encourage exploration and experimentation, leading to deeper cognitive processing and improved learning outcomes according to the ICAP framework[10]. However, despite promoting motiva- tion and learning, interactive visualizations could introduce an ad- ditional cognitive load that is not related to the tasks that students need to perform. [43] Therefore, the research questions this paper aims to explore and answer are the following: How does the use of interactive visualizations in machine learn- ing education affect students' knowledge gain on machine learning topics? How does the use of interactive visualizations in machine learn- ing education affect students' motivation about machine learning topics? The initial hypothesis is that the interactive visualizations will aid the students in understanding the machine learning concepts, based on similar previous research. This hypothesis relies on the effectiveness of interactive visualizations in other similar subjects, but also on their popularity in machine learning courses. Similarly,

the hypothesis for the second question is that the interactive visu- alizations will positively influence the students' motivation related to machine learning topics. The novelty of the current research does not lay in the use of interactive visualizations in the context of teaching machine learn- ing, but rather in developing these visualizations and testing their effectiveness with students. We use two ML non-trivial topics to illustrate the influence of interactive visualizations on students' knowledge gain and motiva- tion. The first topic is gradient descent, which was chosen because it was previously visualized [19, 46], but there is no evaluation of the performance of these visualizations. Whereas, PCA was cho- sen because it is considered to be a rather hard topic to teach to students, as suggested by Westfall [49]. Even if the difficulty was mentioned in the context of behavioral sciences students, it might also apply in the context of computer science students, highlight- ing the importance of tackling this topic. Besides development, this research is also focused on measuring the knowledge gained and the motivation of the students when ex- posed to interactive visualizations, in comparison to when exposed to static visualizations. The research paper will first touch upon the related work in sec- tion 2, then describe the methodology of developing the visualiza- tions and evaluating them in section 3, followed by the results of the evaluation in section 4. Lastly, the discussion of the findings is included in section 5, followed by the conclusion and future work in section 6.

2 RELATED WORK

Despite the importance of machine learning education, this research topic is only now beginning to take shape. One of the first steps taken in this direction is the agenda for future research developed by Shapiro and Fiebrink [41]. This agenda offers plenty of possible research questions and ideas peers in the field could use as inspira- tion for their research. From the large number of research direc- tions provided, the conceptualization and reasoning of students about machine learning algorithms and the parameters for these algorithms served as the starting point for our research. [41]. This point of the agenda highly focuses on exploring how students un- derstand different concepts, which can be partially answered by exploring the usage of interactive visualizations. Other researchers, like Skripchuk et al. [42], have focused on identifying common errors in open-ended ML projects. Their study primarily discusses code-related issues such as inadequate hyper- parameter tuning and improper use of test data during model eval- uation. However, the research focuses on errors in applying ML algorithms rather than students' understanding of the algorithms themselves. The research field also features previous work focusing on the use of visualizations for teaching ML[11, 27, 47, 50]. The first ex- ample of relevant work done in this direction provides interactive visualizations to high school students [11]. In this context, visual- izations are used for describing data collection, visualization, and processing, but also analysis, classification, and regression algo- rithms. All mentioned concepts are adapted to a daily life scenario, namely rain and weather prediction. However, this research does

not evaluate the proposed method, leaving it unclear how effective or not this method would be when applied. Another relevant work for the current research is a convolutional neural network(CNN) visualizer[47]. The authors of this paper describe the process of developing the interactive visualization, while also including an observational study to evaluate the proposed solution. The evalu- ation is mainly focusing on the users' perception of the given vi- sualization, rather than their true knowledge gain. The qualitative study's results included positive feedback from the participants, who found the method helpful in their study process, attractive, and easy to use. Another relevant example is the What-If Tool [50] which aims at helping users probe, visualize, and analyze differ- ent machine learning systems. This method is particularly relevant for the current research as it represents an interactive system that highly uses visualizations since it focuses on minimal coding. The system's target user is represented by people working in differ- ent companies that interact with machine learning. However, the idea behind the system could also be used in an academic scenario. Interactive visualizations are also used in portraying fairness con- cepts related to machine learning algorithms and models, as stud- ied by Mashhadi et al [27]. Their research focuses on 6 open-source fairness tools, by conducting a qualitative review and four focus groups. Through their research, they show the importance of inte- grating interactive visualizations for teaching fairness in machine learning and artificial intelligence courses. One reported insight is that the previously mentioned What-If Tool [50] shows greater transparency due to its interactive design. Interactive visualizations were also previously used in teach- ing science topics[29]. This teaching tool has been successfully ap- plied in mathematics[1, 4, 16, 24, 31, 33, 38], algorithms education [2, 17, 36], statistics [22], information retrieval [8], and engineer- ing education [39]. One key finding of previous research is that interactive visualizations aid in a better understanding of mathe- matical concepts [16, 24]. More than that, existing research shows that animations might be helping students learn faster than with static visualizations[9]. Lastly, existing research looked into the effect of different levels of interactivity on understanding and applying knowledge related to Signals and Systems in Electrical Engineering [35]. The results of the study showed that higher levels of interactivity did not have an effect on the learning outcome in understanding conceptual knowl- edge and applying procedural knowledge and had a negative effect on understanding procedural knowledge.

3 METHODOLOGY 3.1 Gradient Descent and PCA

To assess whether interactive visualizations positively impact the knowledge gained by students and their motivation, we first iden- tify the key concepts of these two topics. On one hand, Gradient Descent is an optimization algorithm used in Machine

Learning mainly for regression problems. This algorithm is used for searching through a large continuously pa- rameterized hypothesis space when the error can be differentiated with regards to the hypotheses[30]. It starts with a random value, updating the parameters of the hypotheses based on the gradient of the error function and the set learning rate.

2

(a) Gradient Descent with one-variable func- tion.

On the other hand, PCA is a dimensionality reduction algorithm used in Machine Learning. This technique derives new dimensions that are uncorrelated linear combinations of the original dimen- sions of a given dataset, maximizing the retained variance of the dataset. The generated dimensions are the eigenvectors of the co- variance matrix of the dataset. These dimensions represent the prin- cipal components, ordered by decreasing eigenvalues, which are directly proportionate to the retained variance.

3.2 Interactive Visualizations Development

Starting from the concepts previously mentioned, 2 Jupyter Note- books were created for the chosen topics1. The Gradient Descent Notebook includes textual information, inspired by existing books and courses in Machine Learning [30, 32], and 3 interactive visu- alizations. The first visualization, shown in figure 1a, shows a one- variable scenario to introduce the students to the idea of gradient descent, and it allows them to set the starting point, learning rate, chosen function, and whether the gradient of the last update is shown graphically. The second visualization, shown in figure 1b, allows the students to manipulate the values of 6 variables, aim- ing to help them understand the mechanics of gradient descent by trying to imitate the behavior of the algorithm. The third visualiza- tion, from figure 1c, displays a scenario with 2 variables, where the student can set the starting coordinates, learning rate, and epochs computed. The second and third visualizations are set within the scope of machine learning specifically since they display a regres- sion scenario with a given dataset. Similarly, the Notebook introducing Principal Components Anal- ysis(PCA) includes textual information, inspired by books and courses in Machine Learning [6, 48], and 3 interactive visualizations. The first visualization, shown in figure 2a, displays a dataset, together with a line on which points are projected, but also the associated re- construction error and variance of the performed transformation. The students in this case can choose the slope of the line used for the transformation. The second visualization, displayed in figure

1https://github.com/ieirentea/Interactive-Visualizations-ML

(b) Gradient Descent with 6-variable func- tion.

Figure 1: Gradient Descent visualizations.

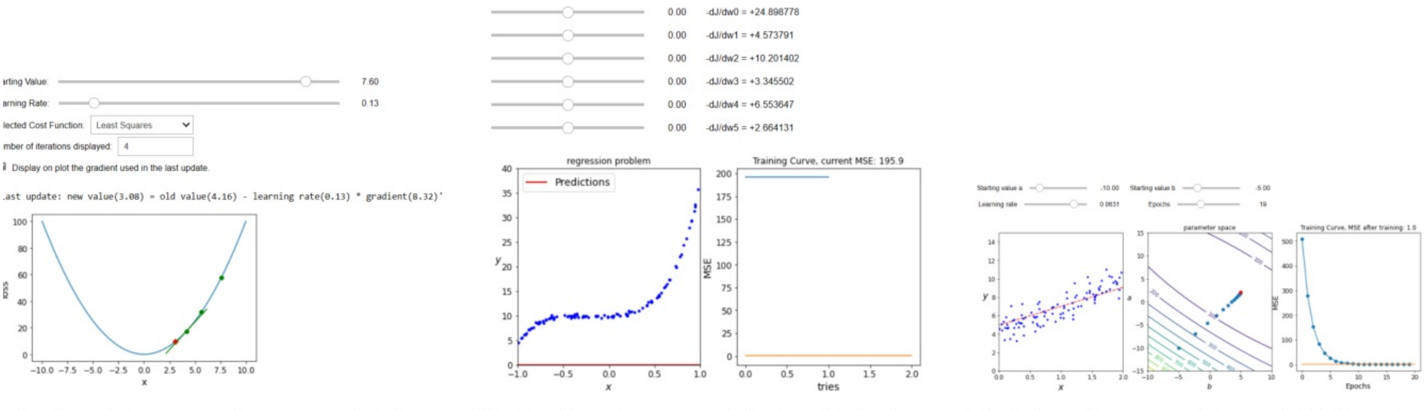
2b, allows students to scale the 2 dimensions of the dataset, ob- serving how the principal components are affected. The third vi- sualization, shown in figure 2c, illustrates how data is transformed by the covariance matrix. The student can choose how many times the data is transformed and should note that the new points are along the principal component with the highest variance. This last visualization aims to make the covariance matrix a less abstract concept. Lastly, cognitive load theory[45] was taken into account while developing the interactive visualizations. Since this teaching tool could sometimes introduce additional cognitive load [43], the visu- alizations' complexity was minimized while retaining the level of knowledge they delivered.

3.3 Experiment Design

The initial hypothesis needed to be tested using an experiment fo- cusing on the knowledge gain and motivation of students. This sec- tion will focus on describing the methodology of the experiment. Participants. The experiment recruited 40 first-year Computer Science and Engineering Bachelor students who did not follow any Machine Learning courses in their previous education and were the least likely to know gradient descent and PCA already. How- ever, at the beginning of the experiment, participants were asked if they were already familiar with these two concepts to ensure that no participant had the knowledge before the experiment. Procedure. The participants were divided into two groups, namely Group A (the control group, N=20), and Group B (the treatment group, N=20). The group assignment was done randomly. Each group received the same survey which included the pre-test and post-test questions, together with the Reduced version of IMMS (Instructional Materials Motivation Survey), for both topics. The procedure of giving a pre-test and a post-test to measure the knowl- edge gain is similar to the one applied by previous research [14, 34]. The pre-test and the post-test were built to match the instruc- tional learning outcomes (ILOs) of the materials, according to the constructive alignment framework[5]. These outcomes were de- fined using Bloom's taxonomy[7], gradient descent being associ- ated with understanding and applying the algorithm, and PCA with

3

(c) Gradient Descent with 2-variable function.



(a) Projection of data points on a chosen line, reconstruction error, and variance.

remembering and understanding it. More than that, the experi- ment was conducted in a quiet, distraction-free space, to ensure the optimum study environment for the participants. Materials. For the experiment, Group B received written ma- terials adapted from the bachelor-level Machine Learning course from TU Delft, accompanied by the interactive visualizations de- veloped for this research. Group A received the same written ma- terials, this time accompanied by static visualizations adapted from the developed interactive ones. The static visualizations were ob- tained by saving multiple instances of possible interactions with the visualizations. Due to the interactive nature of the second visu- alization of gradient descent, it was omitted from the static version of the materials. Measurements. To measure the knowledge gained by the stu- dents, they were asked to fill in a pre-test, follow the given ma- terials, and fill in a post-test. This procedure had the goal of mea- suring the knowledge gain of the students rather than the knowl- edge level, similar to the methodology proposed by [13, 18]. Both tests included the same questions, in increasing order of difficulty, as suggested by existing research [3]. The knowledge gain is mea- sured using the following formula [12]:

= −

, where pre and post are the scores of the knowledge tests. After the post-test, students were given a questionnaire to mea- sure their motivation regarding the received materials. The selected questionnaire is the Reduced Instructional Materials Motivation Survey [25], comprised of 12 questions that measure attention, rel- evance, confidence, and satisfaction associated with the materials. This survey was chosen due to its structured and validated devel- opment, together with its compatibility with the context of this experiment, namely a self-directed instructional setting.

4 RESULTS 4.1 Knowledge gain

The first concept measured by this experiment is the knowledge gained by reading and interacting with the educational materials. The knowledge gain was calculated using equation 1, based on the scores of the pre and post-tests. The score for the pre and post-tests

(b) Effect of scaling dimensions on principal components.

Figure 2: Principal Component Analysis visualizations.

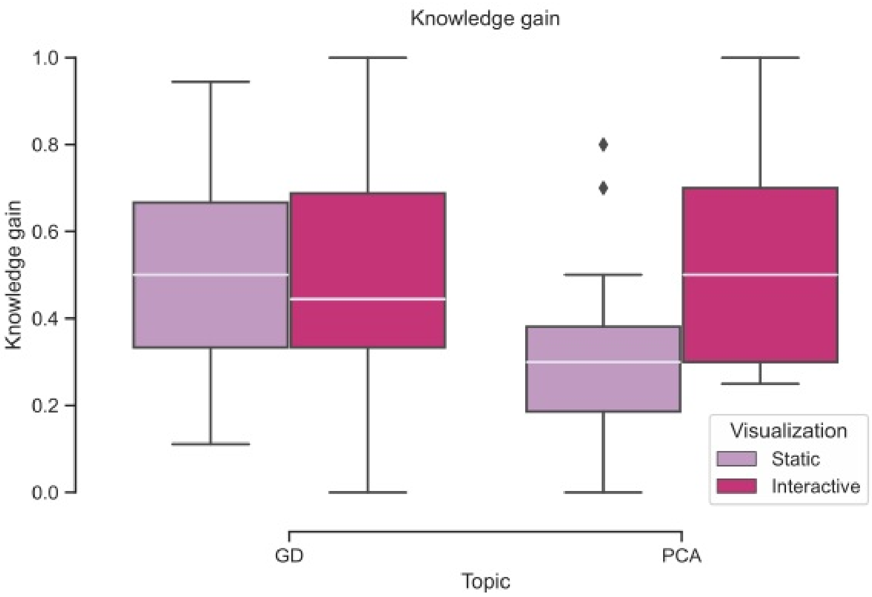
1 − (1)

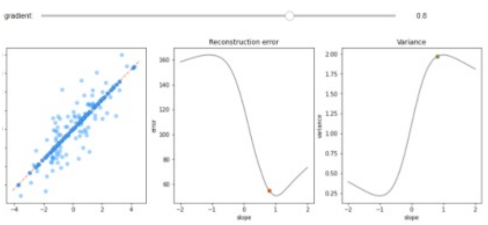
(c) Transformation of data points by covari- ance matrix.

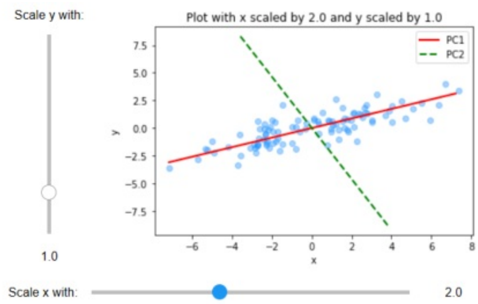
was calculated by assessing the correctness of the answers, where each question received an equal weight for the final score. In figure 3, you can see the aggregated knowledge gain for each topic, comparing the two groups, where static is the control group and interactive is the treatment group.

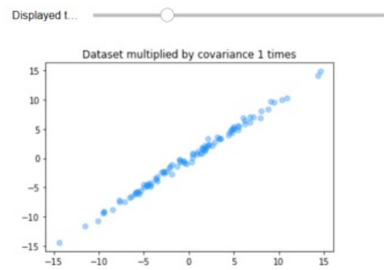
Figure 3: Knowledge gain results. The calculation of the scores is based on equation 1

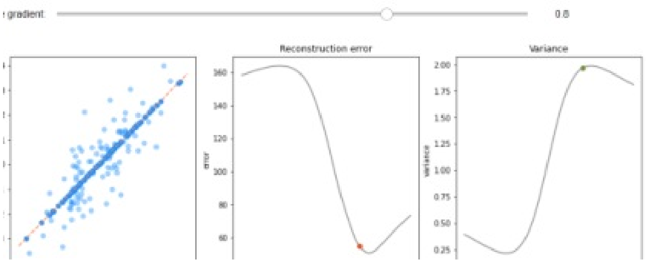
Knowledge Gain gradient descent. Before comparing the knowl- edge gained by the two groups, the scores from the pre-test were compared. Pre-test scores were non-normally distributed (Shapiro- Wilk: interactive $W = 0.5763$, $p < .001$; static $W = 0.6637$, $p < .001$). Therefore, a Mann-Whitney U-test was conducted and showed no significant difference between the pre-test scores of the two groups ($U = 190.5$, $p = .7639$), namely the static($M = 0.1028$, $SD = 0.1759$) and the interactive($M = 0.1333$, $SD = 0.2520$) groups. The normalized knowledge gain of the two groups was com- pared using a t-test due to the normally distributed data(Shapiro- Wilk: interactive $W = 0.9582$, $p = .5091$; static $W = 0.9633$, $p = .6114$) and equal variances(Levene's test: $F = 1.3963$, $p = .2447$). The inde- pendent samples t-test showed no significant difference in knowl- edge gain between the interactive($M = 0.5014$, $SD = 0.2917$) and static visualizations($M = 0.4915$, $SD = 0.2192$) for the gradient de- scent topic($t(df) = 0.1216$, $p = .9038$).











Knowledge Gain PCA. Similarly, the pre-test scores were com- pared before comparing the knowledge gained by the students. Pre- test scores were non-normally distributed (Shapiro-Wilk: interac- tive W = 0.5929, p < .001; static W = 0.751, p < .001). Therefore, a Mann-Whitney U-test was conducted and showed no significant difference (U = 156, p = .1626) between the pre-test scores of the interactive(M = 0.0532, SD = 0.1003) and static (M = 0.1050, SD = 0.1276) visualizations groups regarding PCA. The normalized knowledge gain of the two groups was com- pared using a t-test due to the normally distributed data(Shapiro- Wilk: interactive W = 0.9246, p = .1215; static W = 0.9331, p = .1769) and equal variances(Levene's test: F = 0.7702, p = .3857). The independent samples t-test revealed a very significant differ- ence in knowledge gain between the interactive(M = 0.5274, SD = 0.2268) and static visualizations(M = 0.3162, SD = 0.1957) for the PCA topic(t(df) = 3.1522, p = .0032).

Table 1: Statistical Test Results for Knowledge Gain and Mo- tivation

Variable	Group	Mean	SD	t/U	p-value
Knowledge Gain (GD)	Interactive	0.50	0.29	0.1216	.9038
	Static	0.49	0.22		
Knowledge Gain (PCA)	Interactive	0.52	0.23	3.15	.0032**
	Static	0.32	0.20		
Motivation (GD)	Interactive	3.67	0.61	221.5	.569
	Static	3.52	0.69		
Motivation (PCA)	Interactive	3.48	0.90	-0.193	.8485
	Static	3.52	0.57		

4.2 Motivation

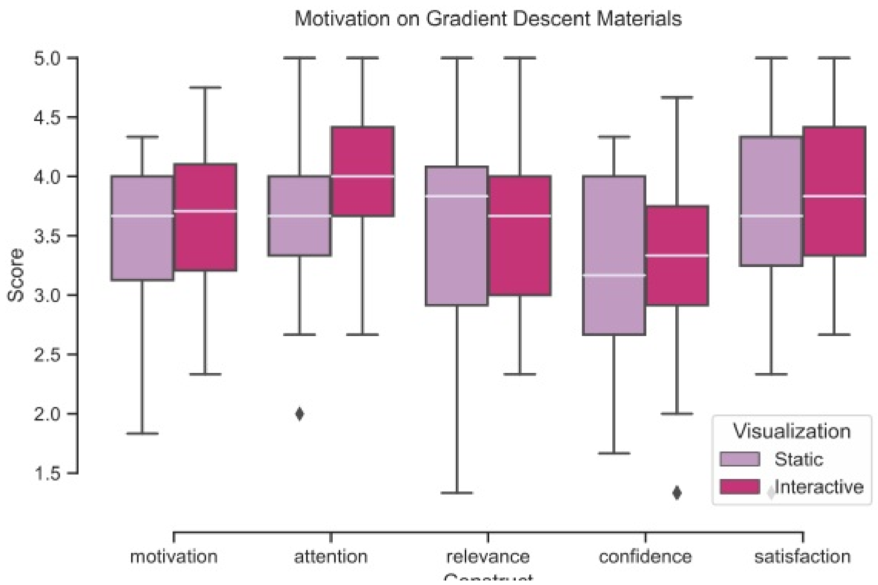
The second construct measured is the students' motivation regard- ing the instructional materials. The Reduced Instructional Materi- als Motivation Survey results were aggregated using the accompa- nying recommendations, namely calculating one score for the over- all motivation, but also one score for each of the four constructs. The aggregated results for the gradient descent and PCA mate- rials are displayed in figure 4, and 5 respectively. In both graphs, the overall motivation score is labeled 'motivation'. Motivation Gradient Descent. A non-parametric Mann-Whitney U-test was used to compare the motivation scores regarding gra- dient descent materials due to the non-normal distribution of the static group scores (Shapiro-Wilk: interactive W = 0.9716, p = .7876; static W = 0.8960, p = .0348). No significant difference was found between interactive (M = 3.6750, SD = 0.6099) and static (M = 3.525, SD = 0.6920) groups (U = 221.5, p = .5690). Motivation PCA. A Welch's t-test was performed to compare the motivation regarding PCA materials due to unequal variances of the two groups (Levene's test: F = 7.8823, p = .0078) and normally distributed data(Shapiro-Wilk: interactive W = 0.9271, p = .1355; static W = 0.9674, p = .7002). No significant difference in motivation regarding PCA was found between the interactive (M = 3.4792, SD = 0.8987) and static (M = 3.525, SD = 0.5698) groups (t(df) = -0.1926, p = .8485).

Figure 4: Motivation regarding gradient descent materials. Motivation is the overall aggregated RIMMS score and the rest are the 4 constructs measured by the survey, each aggre- gated separately.

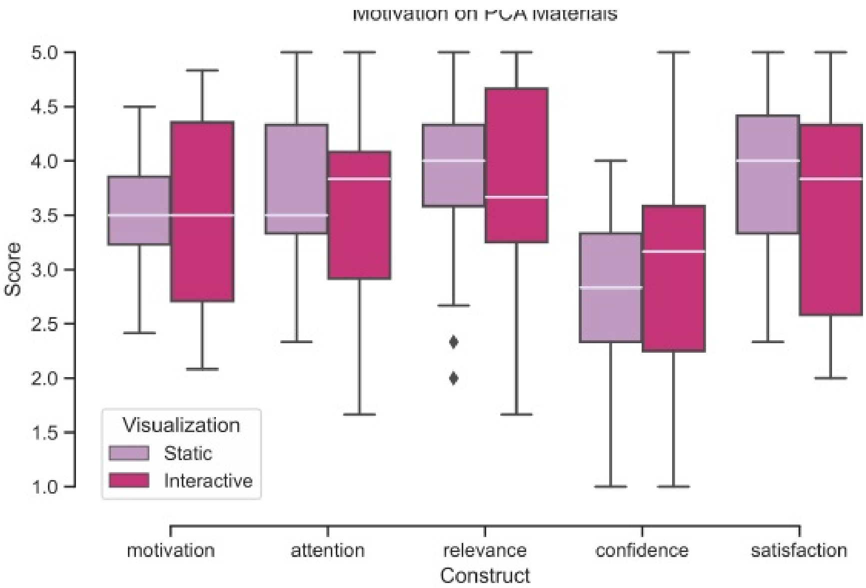
Figure 5: Motivation regarding PCA materials. Motivation is the overall aggregated RIMMS score and the rest are the 4 constructs measured by the survey, each aggregated sepa- rately.

5 DISCUSSION 5.1 Knowledge gain

The results of the conducted study reveal no significant difference in knowledge gained by students between the interactive and static visualizations regarding gradient descent. This outcome indicates that the interactivity of the visualizations did not influence stu- dents' understanding of the presented materials. On the other hand, the study revealed a very significant dif- ference in knowledge gained by students regarding PCA. This re- sult indicates that the added interactivity positively impacted the understanding of the presented materials, helping students better grasp the concepts related to PCA. Therefore, the interactive visualizations had different effects on the two chosen topics. The difference between the outcomes could



Variable	Group	Mean	SD	t/U	p-value
Knowledge Gain (GD)	Interactive Static	0.50 0.49	0.29 0.22	0.1216	.9038
Knowledge Gain (PCA)	Interactive Static	0.52 0.32	0.23 0.20	3.15	.0032**
Motivation (GD)	Interactive Static	3.67 3.52	0.61 0.69	221.5	.569
Motivation (PCA)	Interactive Static	3.48 3.52	0.90 0.57	-0.193	.8485



be a result of the different natures of the two topics. Gradient de- scent is a topic based on calculus concepts, closer related to pre- vious knowledge of the students, while PCA is a topic based on linear algebra concepts, harder to understand and conceptualize, as mentioned by [49]. Based on the results, the interactive visu- alizations might help students better understand non-trivial and more abstract topics. The results of the study contradict the results of the existing research looking into the effect of different levels of interactivity [35]. However, the topics of the mentioned research are within the scope of Electrical Engineering, which are not directly linked to Machine Learning topics.

5.2 Motivation

The results of the conducted research reveal no significant differ- ence in motivation levels regarding the constructed materials for gradient descent and PCA. Therefore, the introduction of interac- tive visualizations does not have a clear effect. One possibility is that the introduction of interactive visualizations does not have an effect on the motivation of the students. Another possibility is that the effect is too small to be observed using the current sample size, namely 20 students per group. However, it is important to mention that both versions of visualizations received relatively high moti- vation scores, having an average of around 3.5, which lies between moderately and mostly true. The lack of a significant difference between interactive and static visualizations could originate from the two versions having the same effect on the motivation of students. However, motivation is measured through the RIMMS survey, which relies on the subjec- tive answers of the students. Consequently, the results reflect the opinions and perceptions of the students on the instructional ma- terials. A larger sample size is needed to generalize the results of this survey and to ensure that potential biases are accounted for.

6 CONCLUSION & FUTURE WORK

Future Work. The first limitation of the current research is the se- lected sample of the study. The experiment included 40 Computer Science and Engineering bachelor students. Expanding the num- ber of participants could strengthen the validity and applicability of the results in a more general context. Also, the research could be expanded by studying the effect of interactive visualizations on non-computer science students. Another important limitation of the presented research study is the short-term nature of the controlled experiment. Due to time limitations, the controlled experiment fully focuses on short-term knowledge gain. However, in the future, it would be important to study the effects of interactive visualizations on long-term mem- ory, since that is the main target of education. More than that, for the current experiment, the setting was a controlled one, where the interventions were minimal and targeted on certain topics. However, the results of the experiment could be extended by conducting an in-the-wild experiment at one or mul- tiple universities. This would imply inserting similar Notebooks, maybe more than those presented in the current research, in Ma- chine Learning courses, and observing their effect on students'

learning experience, mainly through knowledge level and motiva- tion. Additionally, the current research only measures the effect of interactive visualizations for teaching gradient descent and PCA. These two topics are only a subset of the possible topics of ma- chine learning. Therefore, further research is needed to explore the potential of these tools in the context of other topics. Lastly, the focus of our experiment was on Computer Science Bachelor students. However, the effect of interactive visualizations should be also measured with students from other levels of educa- tion, such as pre-university, but also Master's students. More than that, the effectiveness of these tools might differ for non-Computer Science students, which motivates the need for this research direc- tion as well. Conclusion. This research looked into the effect of introduc- ing interactive visualizations in teaching machine learning top- ics, namely gradient descent and PCA. Two sets of materials ex- plaining the two concepts were created, one including static visu- alizations and one including interactive ones. The materials were used in a randomized controlled experiment with 40 TU Delft Com- puter Science and Engineering Bachelor students. The experiment focused on measuring the knowledge gain and motivation related to each topic. The results of the study showed a significant dif- ference in knowledge gained by students regarding PCA, interac- tive visualizations positively impacting the outcome. However, all other statistical tests returned no significant difference between the interactive and static visualizations. Based on the results, interactive visualizations could be a useful tool in teaching certain ML topics. Due to no negative effect ob- served, such visualizations could be introduced in machine learn- ing courses from computer science programs. However, the cur- rent research is only the first step towards validating the effect of using interactive visualizations in machine learning education, additional research being needed to validate the effect of such visu- alizations regarding other topics commonly presented in machine learning.

REFERENCES

[1] Mojeed Kolawole Akinsola and IA Animasahun. 2007. The Effect of Simulation- Games Environment on Students Achievement in and Attitudes to Mathematics in Secondary Schools. Online Submission 6, 3 (2007). [2] Nouf M Al-Barakati and Arwa Y Al-Aama. 2009. The effect of visualizing roles of variables on student performance in an introductory programming course. In Proceedings of the 14th annual ACM SIGCSE conference on Innovation and tech- nology in computer science education. 228–232. [3] Lina Anaya, Nagore Iriberri, Pedro Rey-Biel, and Gema Zamarro. 2022. Under- standing performance in test taking: The role of question difficulty order. Eco- nomics of Education Review 90 (2022), 102293. [4] Svetlana Asmuss, Natalja Budkina, et al. 2019. On usage of visualization tools in teaching mathematics at universities. Engineering for Rural Development 18 (2019), 1962–1969. [5] John Biggs. 1996. Enhancing teaching through constructive alignment. Higher education 32, 3 (1996), 347–364. [6] Christopher M Bishop and Nasser M Nasrabadi. 2006. Pattern recognition and machine learning. Vol. 4. Springer. [7] Benjamin S Bloom, Max D Engelhart, EJ Furst, Walker H Hill, and David R Krath- wohl. 1956. Handbook I: cognitive domain. New York: David McKay (1956). [8] Peter Brusilovsky, Jae-wook Ahn, and Edie Rasmussen. 2010. Teaching informa- tion retrieval with web-based interactive visualization. Journal of Education for Library and Information Science (2010), 187–200. [9] Michael D Byrne, Richard Catrambone, and John T Stasko. 1999. Evaluating ani- mations as student aids in learning computer algorithms. Computers & education 33, 4 (1999), 253–278.

6

[10] Michelene TH Chi and Ruth Wylie. 2014. The ICAP framework: Linking cog- nitive engagement to active learning outcomes. Educational psychologist 49, 4 (2014), 219–243. [11] Siddharth Chittora and Anna Baynes. 2020. Interactive visualizations to intro- duce data science for high school students. In Proceedings of the 21st Annual Conference on Information Technology Education. 236–241. [12] Elaine Christman, Paul Miller, and John Stewart. 2024. Beyond normalized gain: Improved comparison of physics educational outcomes. Physical Review Physics Education Research 20, 1 (2024), 010123. [13] Michael Delucchi. 2014. Measuring student learning in social statistics: A pretest- posttest study of knowledge gain. Teaching Sociology 42, 3 (2014), 231–239. [14] Timothy Ellis. 2004. Animating to build higher cognitive understanding: A model for studying multimedia effectiveness in education. Journal of Engineer- ing Education 93, 1 (2004), 59–64. [15] Christian Herta, Benjamin Voigt, Patrick Baumann, Klaus Strohmenger, Christoph Jansen, Oliver Fischer, Gefei Zhang, and Peter Hufnagel. 2019. Deep teaching: materials for teaching machine and deep learning. In HEAD'19. 5th International Conference on Higher Education Advances. Editorial Universitat Politècnica de València, 1153–1131. [16] Andrea Hoffkamp. 2011. The use of interactive visualizations to foster the under- standing of concepts of calculus: design principles and empirical results. ZDM 43 (2011), 359–372. [17] Christopher D Hundhausen and Sarah A Douglas. 2002. Low-fidelity algorithm visualization. Journal of Visual Languages & Computing 13, 5 (2002), 449–470. [18] Maria-Blanca Ibanez, Angela Di-Serio, and Carlos Delgado-Kloos. 2014. Gam- ification for engaging computer science students in learning activities: A case study. IEEE Transactions on learning technologies 7, 3 (2014), 291–301. [19] Lili Jiang. 2020. A visual explanation of gradient descent methods (momen- tum, AdaGrad, RMSProp, adam). https://towardsdatascience.com/a-visual- explanation-of-gradient-descent-methods-momentum-adagrad-rmsprop- adam-f898b102325c [20] Michael I Jordan and Tom M Mitchell. 2015. Machine learning: Trends, perspec- tives, and prospects. Science 349, 6245 (2015), 255–260. [21] Danijel Kučak, Vedran Juričić, and Goran Đambić. 2018. MACHINE LEARNING IN EDUCATION-A SURVEY OF CURRENT RESEARCH TRENDS. Annals of DAAAM & Proceedings 29 (2018). [22] Shailesh S Kulkarni, Bin Mai, S Yasaman Amirkiaee, and Hakan Tarakci. 2019. Dynamic interactive visualizations: Implications of seeing, doing, and playing for quantitative analysis pedagogy. INFORMS Transactions on Education 19, 3 (2019), 121–142. [23] Konstantinos G Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. 2018. Machine learning in agriculture: A review. Sensors 18, 8 (2018), 2674. [24] Hai-Ning Liang and Kamran Sedig. 2010. Can interactive visualization tools en- gage and support pre-university students in exploring non-trivial mathematical concepts? Computers & Education 54, 4 (2010), 972–991. [25] Nicole Loorbach, Oscar Peters, Joyce Karreman, and Michaël Steehouder. 2015. Validation of the Instructional Materials Motivation Survey (IMMS) in a self- directed instructional setting aimed at working with technology. British journal of educational technology 46, 1 (2015), 204–218. [26] Feiyu Lu, Difeng Yu, Hai-Ning Liang, Wenjun Chen, Konstantinos Papangelis, and Nazlena Mohamad Ali. 2018. Evaluating engagement level and analytical support of interactive visualizations in virtual reality environments. In 2018 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, 143– 152. [27] Afra Mashhadi, Annuska Zolyomi, and Jay Quedado. 2022. A Case Study of Integrating Fairness Visualization Tools in Machine Learning Education. In CHI Conference on Human Factors in Computing Systems Extended Abstracts. 1–7. [28] Richard E Mayer. 2004. Should there be a three-strikes rule against pure discov- ery learning? American psychologist 59, 1 (2004), 14. [29] Kevin W McElhaney, Hsin-Yi Chang, Jennifer L Chiu, and Marcia C Linn. 2015. Evidence for effective uses of dynamic visualisations in science curriculum ma- terials. Studies in Science Education 51, 1 (2015), 49–85. [30] Tom M Mitchell. 1997. Machine learning. [31] Elena Nardi. 2014. Reflections on visualization in mathematics and in mathe- matics education. Mathematics & mathematics education: searching for common ground (2014), 193–220. [32] Andrew Ng. 2000. CS229 Lecture notes. CS229 Lecture notes 1, 1 (2000), 1–3. [33] Hitoshi Nishizawa, Takayoshi Yoshioka, Martti E Pesonen, and Antti Viholainen. 2012. Interactive worksheets for learning the connection between graphic and symbolic object representations. In Proc. 17th Asian Technology Conference in Mathematics. [34] Barbara M Olds, Barbara M Moskal, and Ronald L Miller. 2005. Assessment in engineering education: Evolution, approaches and future collaborations. Journal of Engineering Education 94, 1 (2005), 13–25. [35] Mrinal Patwardhan and Sahana Murthy. 2015. When does higher degree of interaction lead to higher learning in visualizations? Exploring the role of 'In- teractivity Enriching Features'. Computers & Education 82 (2015), 292–305.

[36] Henry W Robbins, Samuel C Gutekunst, David B Shmoys, and David P Williamson. 2023. GILP: An Interactive Tool for Visualizing the Simplex Algo- rithm. In Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1. 108–114. [37] Dilal Saundage, Jacob L Cybulski, Susan Keller, and Lasitha Dharmasena. 2016. Teaching data analysis with interactive visual narratives. Journal of Information Systems Education 27, 4 (2016), 233–248. [38] Kamran Sedig and Mark Sumner. 2006. Characterizing interaction with visual mathematical representations. International Journal of Computers for Mathemat- ical Learning 11 (2006), 1–55. [39] Levent Sevgi. 2006. Modeling and simulation concepts in engineering education: virtual tools. Turkish Journal of Electrical Engineering and Computer Sciences 14, 1 (2006), 113–127. [40] K Shailaja, Banoth Seetharamulu, and MA Jabbar. 2018. Machine learning in healthcare: A review. In 2018 Second international conference on electronics, com- munication and aerospace technology (ICECA). IEEE, 910–914. [41] R Benjamin Shapiro and Rebecca Fiebrink.

2019. Introduction to the special sec- tion: Launching an agenda for research on learning machine learning. , 6 pages. [42] James Skripchuk, Yang Shi, and Thomas Price. 2022. Identifying Common Er- rors in Open-Ended Machine Learning Projects. In Proceedings of the 53rd ACM Technical Symposium on Computer Science Education-Volume 1. 216–222. [43] Alexander Skulmowski and Kate Man Xu. 2022. Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. Educational psychology review 34, 1 (2022), 171–196. [44] Elisabeth Sulmont, Elizabeth Patitsas, and Jeremy R Cooperstock. 2019. What is hard about teaching machine learning to non-majors? Insights from classifying instructors' learning goals. ACM Transactions on Computing Education (TOCE) 19, 4 (2019), 1–16. [45] John Sweller. 2011. Cognitive load theory. In Psychology of learning and motiva- tion. Vol. 55. Elsevier, 37–76. [46] Teach LA Dev Team. [n. d.]. Gradient Descent Visualiser. https://uclaacm. github.io/gradient-descent-visualiser/ [47] Zijie J Wang, Robert Turko, Omar Shaikh, Haekyu Park, Nilaksh Das, Fred Hohman, Minsuk Kahng, and Duen Horng Polo Chau. 2020. CNN explainer: learning convolutional neural networks with interactive visualization. IEEE Transactions on Visualization and Computer Graphics 27, 2 (2020), 1396–1406. [48] Andrew R Webb, Keith D Copsey, and Gavin Cawley. 2011. Statistical pattern recognition. Vol. 2. Wiley Online Library. [49] Peter H Westfall, Andrea L Arias, and Lawrence V Fulton. 2017. Teaching princi- pal components using correlations. Multivariate behavioral research 52, 5 (2017), 648–660. [50] James Wexler, Mahima Pushkarna, Tolga Bolukbasi, Martin Wattenberg, Fer- nanda Viégas, and Jimbo Wilson. 2019. The what-if tool: Interactive probing of machine learning models. IEEE transactions on visualization and computer graphics 26, 1 (2019), 56–65.

As mentioned in the introduction, this section aims to help the reader better understand the process of the conducted research. It provides more detail than the scientific article includes since the former follows the structure of a scientific article. Therefore, section 3.1 will provide a more thorough analysis of the existing research in the field of machine learning education, but also in adjacent fields. Next, section 3.2 focuses on the theory behind the chosen topics, namely section 3.2.1 describes gradient descent, and section 3.2.2 describes the principal component analysis. More than that, section 3.3 describes the developed visualizations for gradient descent, included in section 3.3.1, and for principal component analysis, included in section 3.3.2. Afterwards, sections 3.4 and 3.5 focus on the expert evaluations and on the development of the user study.

3.1. Related Work On top of the related work already mentioned in Chapter 2, more literature was explored during the first stage of the research. Firstly, existing research exploring the use of interactive visualizations in other related fields, such as mathematics or computer science, was considered. More than that, existing literature about the use of dynamic simulations in active learning or inquiry-based learning was analyzed. Another topic relevant to the current research is student engagement through the use of visualizations. Lastly, current methodologies of evaluation of Machine Learning teaching methods, but also of teaching methods in general were considered.

3.1.1. Interactive Visualizations in Mathematics Machine Learning concepts are closely related to mathematics since they use certain knowledge from calculus, linear algebra, and statistics. Based on this link, it is relevant to look into the use of interactive visualizations in teaching mathematics. First of all, previous research has looked into the use of visualization tools in university-level educa- tion[3, 48], but also in pre-university education[41, 27, 2]. Based on the mentioned research, such tools are effective in both levels of education, when they are correctly designed to fit the intended audience. More than that, mathematics is a field that includes several subfields with very different topics. Therefore, researchers have tackled the application of visualization tools in calculus [57, 27, 48], but also geometry[41], proofs[61] and optimization problems[3]. The mentioned applications of interactive visualizations have promising outcomes, showing the potential of these tools in teaching mathematics. The focus on calculus is relevant for the current research since gradient descent is a topic that heavily relies on calculus, namely on derivation. More than that the focus on optimization problems[3] can also motivate the use of interacted visualizations in teaching machine learning topics. Therefore, the positive outcomes of such tools in teaching mathematics motivate the potential of the tools in teaching machine learning topics.

3.1.2. Interactive Visualizations in Other Disciplines Besides exploring the use of interactive visualizations in mathematics, researchers have also looked into the usage of these tools in other disciplines related to machine learning. Examples of such disci-

Supplemental Materials

3

plines are algorithms education[5], hierarchical clustering[58], statistics [39] and engineering education [59]. From this list, statistics are closely related to machine learning, since they sit at the base of some models. Also, these visualizations are accompanied by visualizations for the optimal facility location problem, an optimization problem[39]. More than that, the hierarchical clustering explorer[58] is related to machine learning, having some concepts in common. Similar to the mathematics visualizations, the positive outcomes of other subjects motivate the potential of this teaching tool.

3.1.3. Teaching Techniques Active Learning. The first relevant teaching technique is active learning. This technique relies on intro- ducing instructional methods that involve the student and motivate them to reflect upon their knowledge and how they relate to other pieces of information [20, 52, 45]. This technique has been declared suc- cessful in increasing exam scores and reducing failure rates[20], increasing student engagement and learning outcomes[52], and helping students better understand and remember the materials[45]. More than that, active learning has been combined with interactive visualizations in teaching computer sci- ence concepts[56]. This approach has received positive feedback when applied in a classroom context. Discovery Learning. Another teaching technique related to this research is discovery learning. This technique is a constructivist educational approach in which students explore, experience, and discover concepts independently, inquiring about the information rather than receiving it directly[23]. However, past research has shown that discovery learning has the best outcomes when an instructor guides it[44]. This type of learning has been combined with simulations in past research [54, 13]. More than that, previous research has shown that combining these two techniques achieves better learning outcomes than only using discovery learning alone[13]. Inquiry-based Learning. A third type of teaching technique is inquiry-based learning. This type of learning is an approach that highlights the role of the student in the learning process, encouraging exploration of knowledge by investigation and answering questions. This approach is well-defined, with clear phases including orientation, conceptualization, investigation, conclusion, and discussion [51]. Along with other fields, inquiry-based learning has been used in teaching algorithms[40]. More than that, simulations have been combined with this technique and it lead to positive learning outcomes [47], simulations helping students understand the underlying knowledge of certain models.

3.1.4. Engagement through Visualizations Most research focusing on interactive visualizations measures the learning outcomes of using these tools[47, 13, 45]. However, there is also research that is interested in measuring the engagement and motivation levels of the students[41, 52, 43, 32]. More than that, a review [25] highlights the existing trend of measuring students' engagement in technology-mediated learning. This review motivates the need to measure engagement in learning activities, besides learning outcomes.

3.1.5. Evaluation of Machine Learning Teaching Methods One of the topics relevant to expand in this section is the setup of experiments from this area of re- search. Multiple research papers detail their methodology and can serve as inspiration for this paper's experiment. Some existing research chose to evaluate the methods they propose through qualitative measures. Ebert et al proposed a strategy for creating learning materials that should motivate students by provid- ing an interesting learning experience. [16] The methods presented in the paper were evaluated by conducting interviews with the participants. Hasselmann and Lurkin shaped a bachelor-level course aimed at presenting searching algorithms applied to board games. [24] The course presented in the paper was evaluated through observations drawn from informal interactions and the atmosphere ob- served during the course. The level of knowledge of students was measured using the quality of the final project and the results of the competition run at the end of the course. Shouman et al. discussed two practical machine learning courses, one introductory and one advanced. [60] The courses were evaluated using surveys to collect feedback from the students. Also, they included graded homework and group projects to measure the knowledge of students who followed the courses. Dogan presents course project materials that aim to stop students from overfitting the already existing machine learning algorithms and to create their own solutions. [15] The course was run for two semesters and included a small amount of homework and project materials to evaluate the students. Also, the course included a reflection opportunity for both students and mentors using the Gibbs Reflective Cycle [22]. This

reflection method was also used to extract some indirect feedback on the course and its materials. More than that, some researchers measure the knowledge gained through different evaluation meth- ods. Kazmi designs an introductory course on machine learning and optimization for energy engineer- ing students. [33] The course is designed to evaluate the learning outcomes through Kaggle compe- titions and a project. Similarly, Weerts and Pechenizkiy developed a course for responsible machine learning for engineering students. [64] The presented course included evaluation methods individual assignments, 3 quizzes, and a final group project. The quizzes would have been particularly relevant to the current research, but they were not included in the research paper. Kinnaird presents a course focusing on improving coding and algorithmic-related skills, assuming to statistical background knowl- edge. [37] The course uses assignments and projects to assess the student's knowledge level since it is fully focused on understanding the coding of machine learning algorithms, rather than how and why they work a certain way. Similarly, Brown details an introductory course on data science that focuses on programming. [9] An interesting aspect of this research is how the learning outcomes are divided into different levels of understanding and how these levels each influence the final grade of the student. The researcher defined different skills the students should acquire by the end of the course, each with 3 levels of difficulty.

3.1.6. Evaluation of General Teaching Methods Since there is no standard way of evaluating machine learning teaching methods in particular, it is also relevant to look into general practices regarding evaluation. For this reason, it is also important to analyze the existing methods of measuring the effect of introducing new teaching methods. First of all, Delucchi measures the knowledge gained regarding social statistics through the use of a pre-test and a post-test [14]. In a similar manner, Ibanez studied the knowledge gain when gamification is introduced in computer science topics, namely programming, through the same combination of tests [29]. More than that, an article describing the state of assess- ment in engineering courses identifies randomized controlled trials [8] as an efficient way of comparing two methods of teaching, namely the standard way and the newly introduced method [50]. This method was applied by Ellis in their study about multimedia effectiveness in education together with the pre-test and post-test combination, focusing on the knowledge gained by the participants [18]. Besides the knowledge gain focus, there is also research introducing and validating the use of surveys to measure student motivation. Keller introduced the Instructional Materials Motivation Sur- vey(IMMS) which focuses on attention, relevance, confidence, and satisfaction, comprising 36 ques- tions each associated with one of the aforementioned categories [34]. Based on this survey, Loorbach et al introduce the Reduced Instructional Materials Motivation Survey(RIMMS) that includes only 12 questions from the initial set [42]. The subset of questions were validated and they were proven to be preferred over the initial set of questions, being effective in measuring the four constructs in a self- directed instructional setting.

3.2. Description of Topics Chosen for Visualization This section includes a description of the machine-learning topics chosen for building interactive visu- alizations. The concepts were studied from several different sources to ensure the highest level of completeness in describing the subjects.

3.2.1. Gradient Descent Gradient descent is the first relevant topic for the current research. This concept is usually taught in machine learning courses in the context of linear regression, classification, as well as deep learning. Gradient descent is an algorithm "for searching through a large or infinite hypothesis space that can be applied whenever the hypothesis space contains continuously parameterized hypotheses and the error can be differentiated with respect to these hypothesis parameters"[46]. The search is performed by starting with a random value that is updated based on the learning rate and the gradient of the chosen cost function. The update is proportional to the gradient term, having a small update if the gradient is small. The general form of the update [49] is represented in 3.1, where α is the learning rate and θ is the parameter that needs to be optimized.

Depending on the chosen cost function, the gradient descent update can take several shapes. For example, in linear regression, the most commonly used cost function is the least squares function, represented in 3.2, giving the least squares regression model. This model is associated with the update rule represented in 3.3.

Generally, there are two versions of gradient descent[49], namely batch gradient descent and stochastic gradient descent. Both versions update the parameter until convergence, but the manner of doing it differs. The batch gradient descent updates the parameters considering all points in the training set. This implies that the algorithm has to compute the error for all data points before each update. On the other hand, stochastic gradient descent updates the parameter considering only one point at a time. Therefore, stochastic gradient descent is a better alternative when dealing with a large training set, and manages to optimize the parameter faster. However, since the stochastic version only considers one point at a time, it may never converge to the true minimum, oscillating around the minimum of the cost function. In practice, there is also another version of gradient descent that combines the batch and stochas- tic. It is called mini-batch gradient descent[49], which considers a small batch of data points when computing the update. The main issues that can occur while using this algorithm are that the convergence might be slow, and if the cost function includes several local minima, there is no guarantee that the algorithm will find the global minimum[46].

3.2.2. Principal Component Analysis(PCA) Principal Component Analysis(PCA) is a technique used for deriving new variables that are uncor- related linear combinations of the original variables of a dataset. These variables are generated in decreasing order of importance based on the amount of variance they account for. As mentioned by Web and Copsey PCA is generally used for defining a smaller group of variables that can describe the data reasonably well compared to the original variables[63]. The method of principal component analysis identifies first the principal component associated with the highest principal value, which represents the variance. The first principal component is calculated by finding the line through the data which keeps the highest variance and lowest distance to the points. These two measures are inversely proportionate and any of them can be used in the analysis. The second principal component is the line orthogonal to the first principal component that keeps the most variance. This process is repeated until all principal components are found, the number depending on the dimensions of the initial set of data points. The principal components can be easily identified by computing the eigenvectors and their eigen- values for the covariance matrix of the data set. The set of eigenvectors and eigenvalues are sorted in descending order of the eigenvalues because these represent the variance. After identifying the principal components the variance is studied. If the first few principal compo- nents account for most of the variance, then the rest can be discarded, resulting in dimensionality reduc- tion. The number of principal components used can be determined by using a set level of preserved variance, such as between 70% and 90% [63], or by using the elbow method. The aforementioned method analyses the point where the variance associated with the components falls abruptly before settling at small values. PCA also depends on some factors related to the data set. First of all, the algorithm requires the covariance matrix, which is estimated using the sample data. More than that, the range of values for each variable can affect the principal component analysis. Therefore, before computing the principal components, the data needs to be standardized by transforming it to have zero mean and unit variance.

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \quad (3.1)$$

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{n} (h_\theta(x^{(i)}) - y^{(i)})^2 \quad (3.2)$$

$$\theta_j := \theta_j + \alpha(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)} \quad (3.3)$$

3.3. Developed visualizations After describing the theory behind the chosen topics in section 3.2, it is relevant to provide details about the visualizations that were created. Each visualization is associated with a list of learning objectives and a description. Learning objectives are crucial in this process because they should dictate the design of the final visualization. The learning objectives are defined according to Bloom's taxonomy of educational objectives, more specifically according to the taxonomy presented by Krathwohl [38]. From this revised version of the original taxonomy, the focus is on the cognitive process dimension and the knowledge dimension. The cognitive process dimension divides the learner's outcomes into 6 categories: remember, understand, apply, analyze, evaluate, and create. The knowledge dimension divides knowledge into 4 categories, namely factual, conceptual, procedural, and metacognitive. The taxonomy also focuses on affective and psycho-motor outcomes, but these are not relevant to the current research. On top of defining clear learning objectives for each topic, the bachelor-level course in Machine Learning from the Technical University Delft was analyzed. The materials from this course, together with the learning objectives represent the starting point of the visualizations and developed Jupyter Notebooks, by influencing the presented content.

3.3.1. Gradient Descent Visualizations The first step in deriving a proper interactive visualization is to derive the learning objectives linked to the given topic. Regarding gradient descent, the learning objectives are the following:

• LO1: Explain the gradient descent algorithm and its components. (Understand Conceptual Knowl- edge) • LO2: Infer potential problems of gradient descent and their causes. (Understand Conceptual Knowledge) • LO3: Execute gradient descent update steps. (Apply Procedural Knowledge)

To aid the students in understanding the gradient descent algorithm, three interactive visualizations were developed. Before discovering and interacting with these visualizations, the students need to read some textual information about the theory behind gradient descent, and the mathematical formulation of this algorithm. After reading this information, the students interact with the following visualizations: Gradient Descent for functions with one variable. The first step to understanding gradient de- scent as a general algorithm is to understand it in its most simple form, namely with a one-variable function. This is the reasoning behind having this visualization the first one shown to the student. The visualization shows the behavior of gradient descent with a set function, learning rate, starting point, and the number of iterations displayed, as it can be observed in Figure 3.1. The student can interact with this visualization by changing the value of the variables mentioned above, with the mention that the students can select functions from a pre-defined list. On top of setting values for variables and observing their behavior, the students can opt to display or hide the gradient used in the last update calculation. Since mathematical calculations might be one of the non-intuitive parts of gradient descent, the students see the last performed update in a mathematical form at all times. Lastly, this visualization aims to help the student understand the gradient descent as a general searching algorithm, rather than a machine learning algorithm. This is also the reason why the first visualization is not linked to a specific data set or machine learning problem. Gradient Descent for functions with more than 2 variables. The next visualization included in the Jupyter Notebook displays some data points and the training curve of the algorithm, as it can be seen in Figure 3.2. This visualization allows the user to imitate the gradient descent algorithm by manipulating the value of the 6 presented variables. The visualization was taken from the course TI3145TU Machine Learning and Introduction to AI taught by Tom Viering and Neil Yorke-Smith. This visualization aims to help the student actively think about the behavior of the gradient descent algorithm and try to imitate it. Ideally, the student would look at the displayed gradients for each of the variables and decide to change the largest ones, until they find a configuration that resembles the distribution of the data closely enough. In this visualization, the training curve is the student's training curve, showing the mean squared error at each update of the variables.

(a) Visualization for gradient descent applied on least squares function, with starting value = 7, learning rate = 0.13, 4 shown iterations, including the gradient used in the last update calculation.

(a) Initial instance of the visualization, with MSE = 195.9. (b) Instance of the visualization after around 25 updates, with MSE = 0.6.

Figure 3.2: Gradient descent: visualization for function with 6 variables

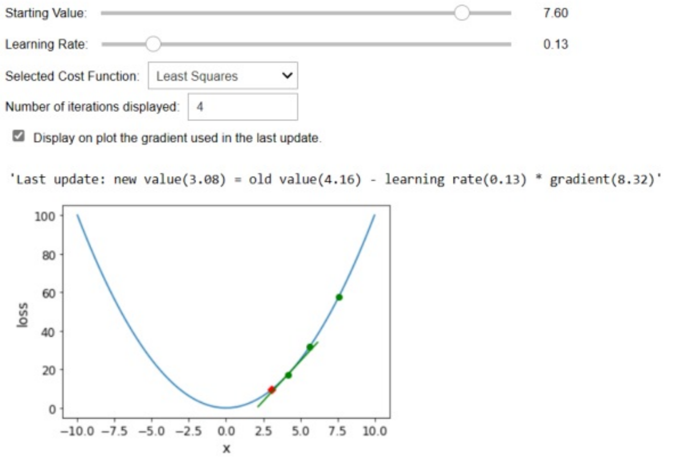
Gradient Descent for functions with 2 variables. The third visualization included in the Notebook is a visualization showing 3 subplots, namely one plot showing a dataset together with a regression line, one plot showing the evolution of the gradient descent for the given parameters through a 3d-plot of the loss function, and one plot showing the training curve, mentioning the mean squared error after training. The visualization can be seen in Figure 3.3 and it was taken from the previously mentioned course TI3145TU Machine Learning and Introduction to AI taught by Tom Viering and Neil Yorke-Smith. This visualization aims to help the student understand how the gradient descent can be applied in more than 1-dimensional scenarios. Similar to the first visualization, the student can observe the influence of the learning rate and starting point on the accuracy of the algorithm within a set number of epochs.

3.3.2. PCA Visualizations The learning objectives designed for PCA are the following:

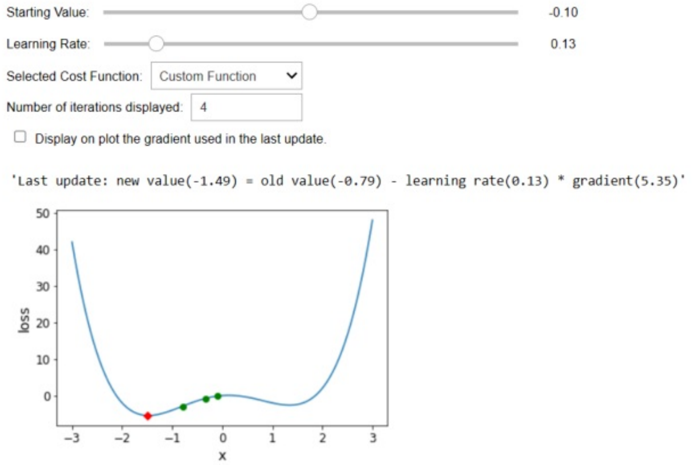
• LO4: Recall Principal Component Analysis elements. (Remember Factual Knowledge) • LO5: Explain Principal Component Analysis and its components. (Understand Conceptual Knowl- edge) • LO6: Infer equivalence between variance and error Principal Component Analysis versions. (Un-

Figure 3.1: Gradient descent: visualization for one-variable function

(b) Visualization for gradient descent applied on a custom function, with starting value = -0.1, learning rate = 0.13, 4 shown iterations, without displaying the gradient used in the last update calculation.

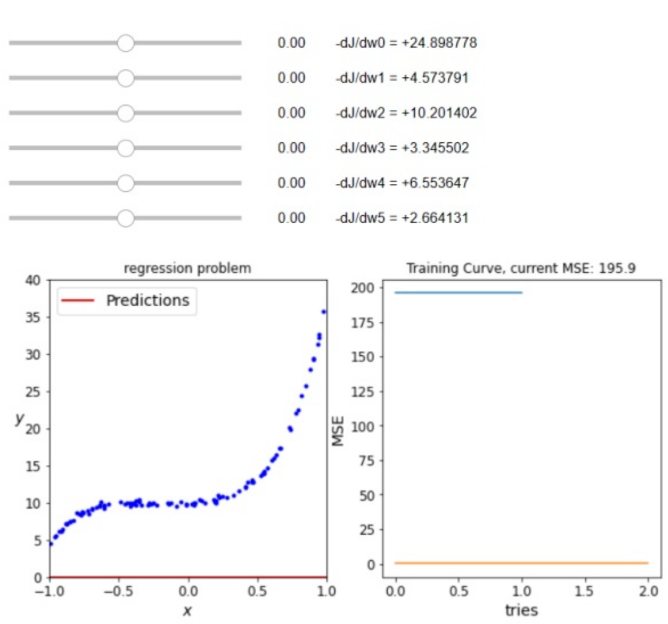


(a) Visualization for gradient descent applied on least squares function, with starting value = 7, learning rate = 0.13, 4 shown iterations, including the gradient used in the last update calculation.

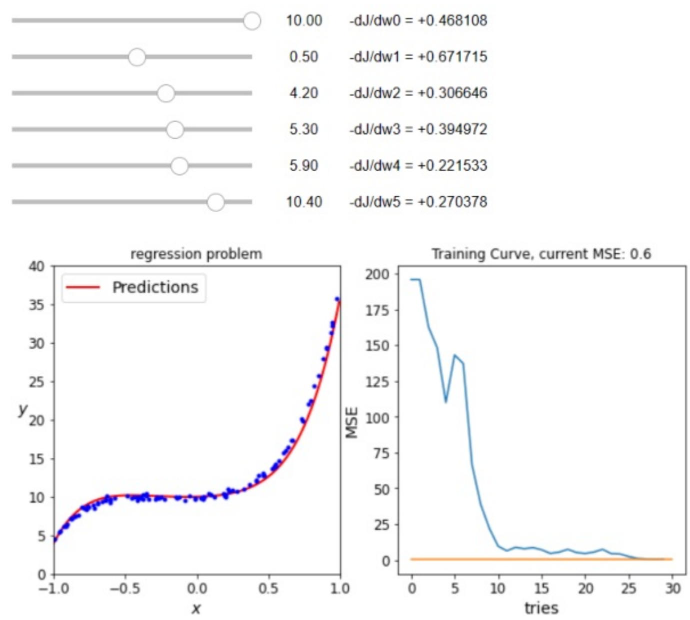


(b) Visualization for gradient descent applied on a custom function, with starting value = -0.1, learning rate = 0.13, 4 shown iterations, without displaying the gradient used in the last update calculation.

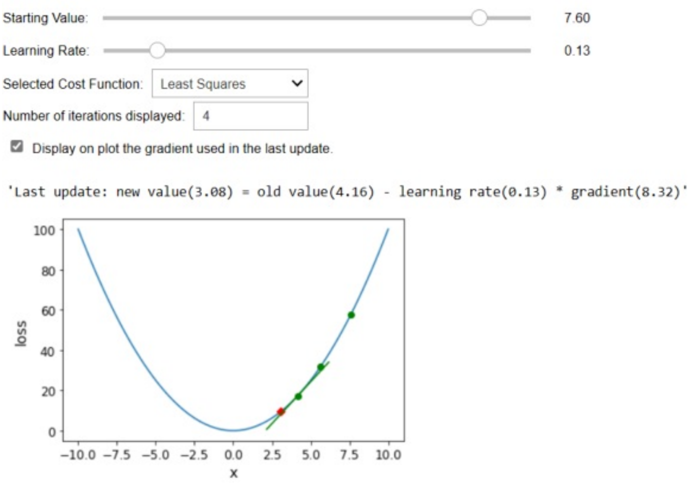
Figure 3.1: Gradient descent: visualization for one-variable function

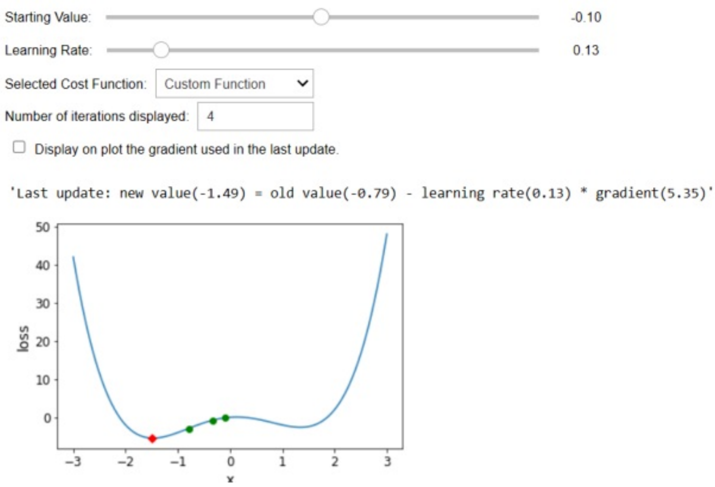


(a) Initial instance of the visualization, with MSE = 195.9



(b) Instance of the visualization after around 25 updates, with MSE = 0.6





(a) Visualization for gradient descent for the given dataset with mean squared error, with starting values a = -10, b= -5, learning rate = 0.01, and 5 epochs.

(b) Visualization for gradient descent for the given dataset with mean squared error, with starting values a = -10, b= -5, learning rate = 0.0631, and 19 epochs.

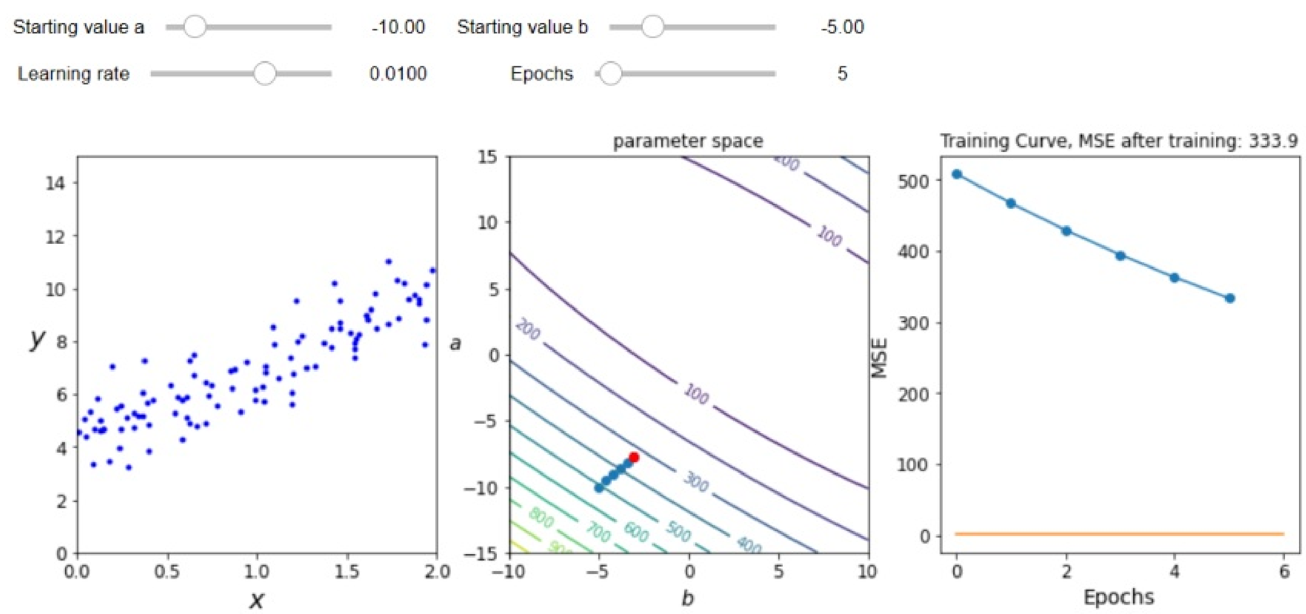
(c) Visualization for gradient descent for the given dataset with mean squared error, with starting values a = 0.9, b= 0.8, learning rate = 0.0631, and 19 epochs.

Figure 3.3: Gradient descent: visualization for function with 6 variables

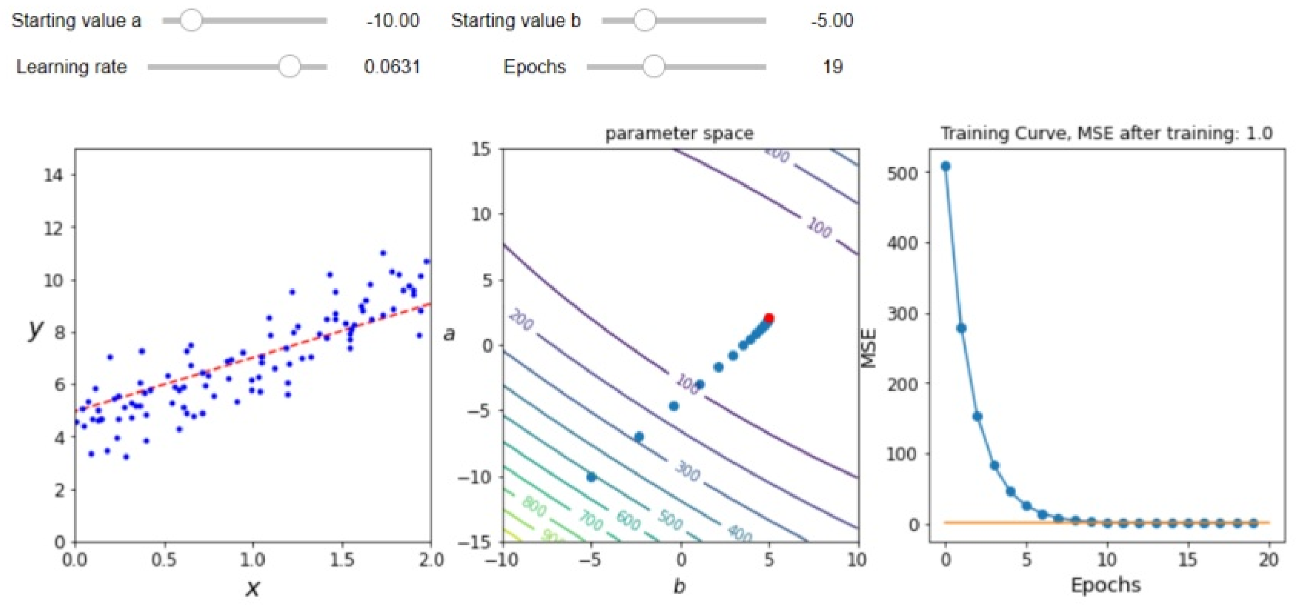
derstand Conceptual Knowledge)

Similar to the process from the gradient descent part, three visualizations were developed to explain the PCA algorithm and concepts to students. Before interacting with the visualizations, the students need to read some textual information that briefly introduces the algorithm, and explains the mathe- matical derivation that backs up this method. The textual information was derived using the Machine Learning course previously mentioned, as well as the materials recommended by the course [63, 7]. Most of the text focuses on the mathematical derivation, since this element of understanding principal components analysis is considered by the students to be quite difficult, as suggested by the lecturers of Machine Learning. More than that, the visualizations were developed following a discussion with a Machine Learning professor. After reading the information, the students can interact with the following visualizations:

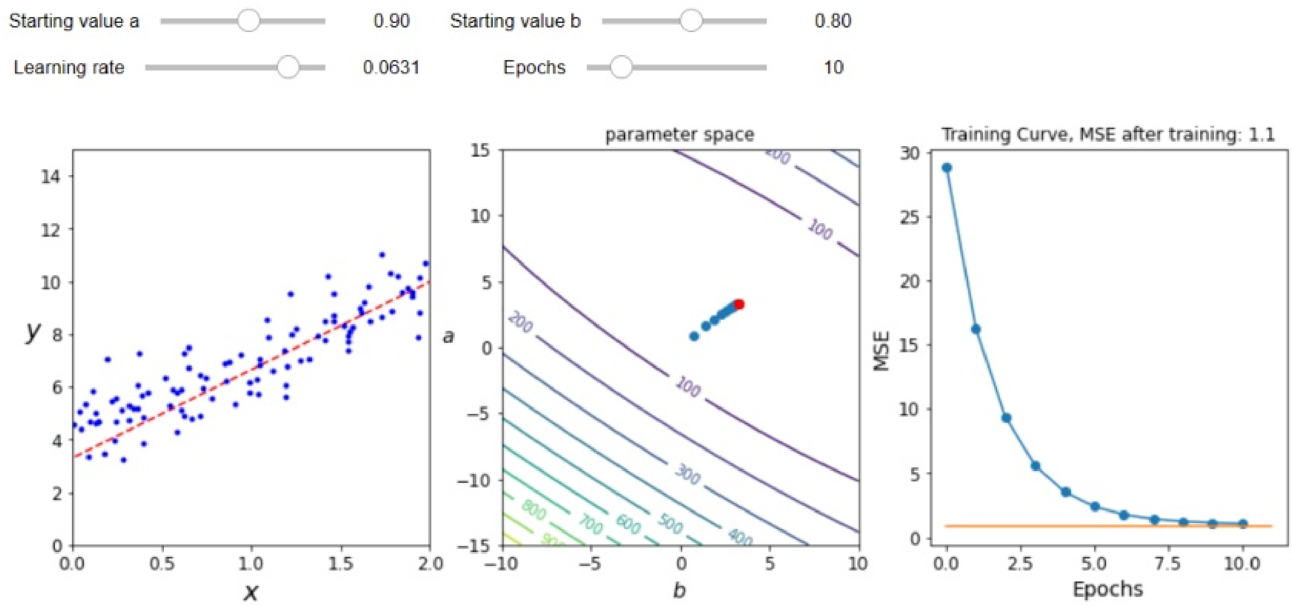
d visualizations



(a) Visualization for gradient descent for the given dataset with mean squared error, with starting values a = -10, b= -5, learning rate = 0.01, and 5 epochs.



(b) Visualization for gradient descent for the given dataset with mean squared error, with starting values a = -10, b= -5, learning rate = 0.0631, and 19 epochs.



(c) Visualization for gradient descent for the given dataset with mean squared error, with starting

3.3. Developed visualizations 18

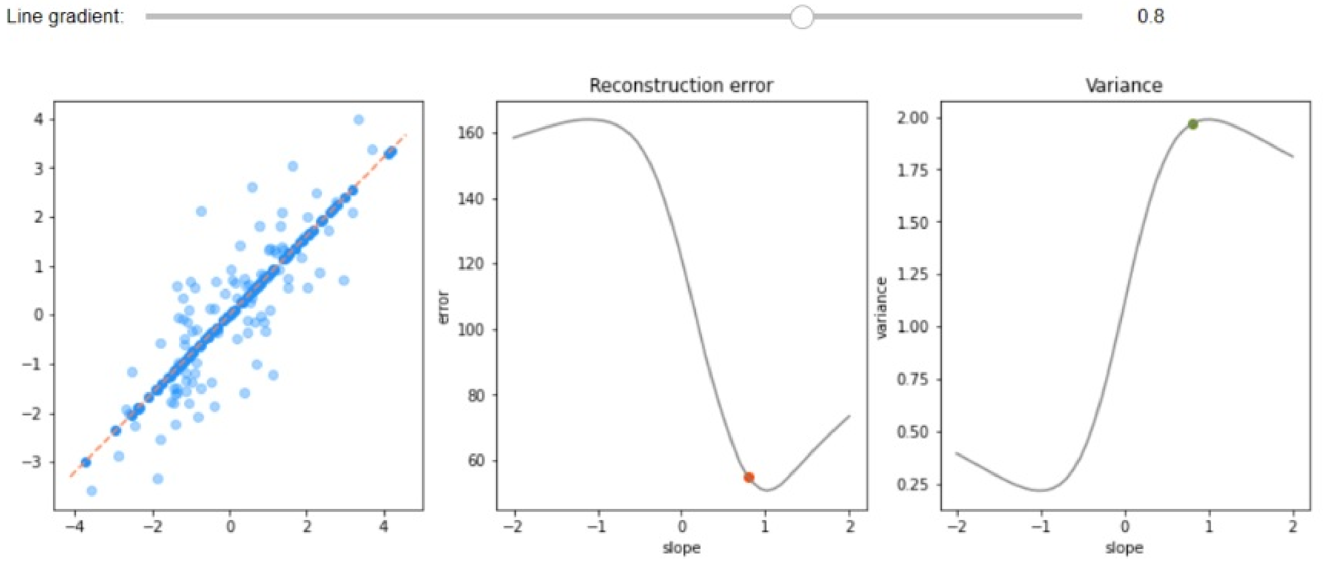
Projection of dataset on a line, together with the error and variance. The first visualization includes 3 graphs, as it can be seen in Figure 3.4. The first graph displays a dataset, together with a line on which the points are projected. In this graph, the initial data points are the ones with a faded color, and the projected points are the ones with a stronger color. The second and third ones display the error and variance of the projected points. In these, the current projection's error and variance are displayed through a point. The student can interact with the visualization by changing the slope of the line on which the points are projected. The aim of this visualization is for the students to understand how points get projected, and how the variance and error are influenced by the chosen lines. More than that, in the textual infor- mation from the beginning of the Notebook, the variant of choosing to minimize the error was left out on purpose. The reason behind this is that the students should be able to infer this information by looking at the graphs of error and variance and observing that the 2 functions are inversely proportionate.

Figure 3.4: PCA: projection on a chosen line, including variance and error graphs

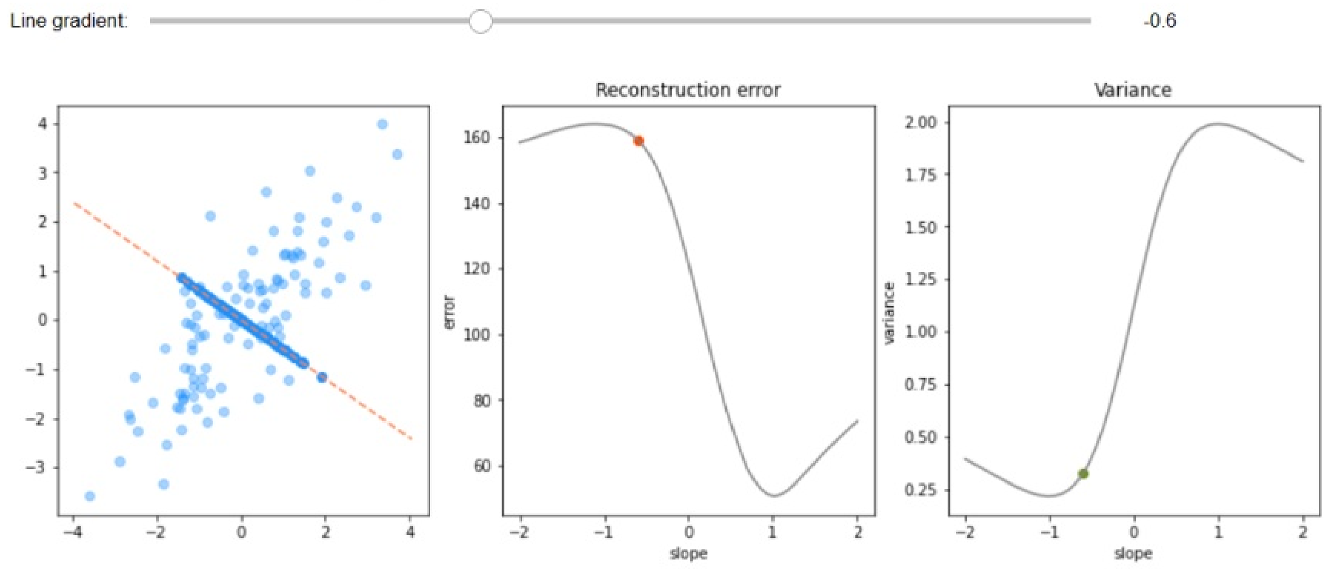
Principal components with scaled dimensions. The second visualization included in the Note- book displays the same dataset as the one from the first visualization. Together with the data, it displays the principal components obtained by applying the PCA algorithm, as shown in Figure 3.5. The aim of this visualization is to highlight the importance of standardizing the data, as mentioned in Section 3.2.2. In this case, the student can interact with the visualization by scaling each of the two variables of the initial dataset. Besides this graph, the visualization also includes the initial version, where variables are scaled by a factor of 1, at all times, similar to the one shown in Figure 3.5a. This way, the student can scale the variables and compare the newly generated principal components with the initial ones. Visualizing effect of transforming the dataset by the covariance matrix. The third visualization initially includes a graph of the previously used dataset. This visualization shows the effects of trans- forming the data by the covariance matrix and it aims to help the student better understand the theory behind this algorithm. The students can interact with the visualization by moving the slider and deciding how many times the data is transformed. Ideally, after transforming it once, they will observe that the data resembles the first principal component of the data.

(a) Instance of the visualization with slope = 0.8

(b) Instance of the visualization with slope = -0.6



(a) Instance of the visualization with slope = 0.8



(b) Instance of the visualization with slope = -0.6

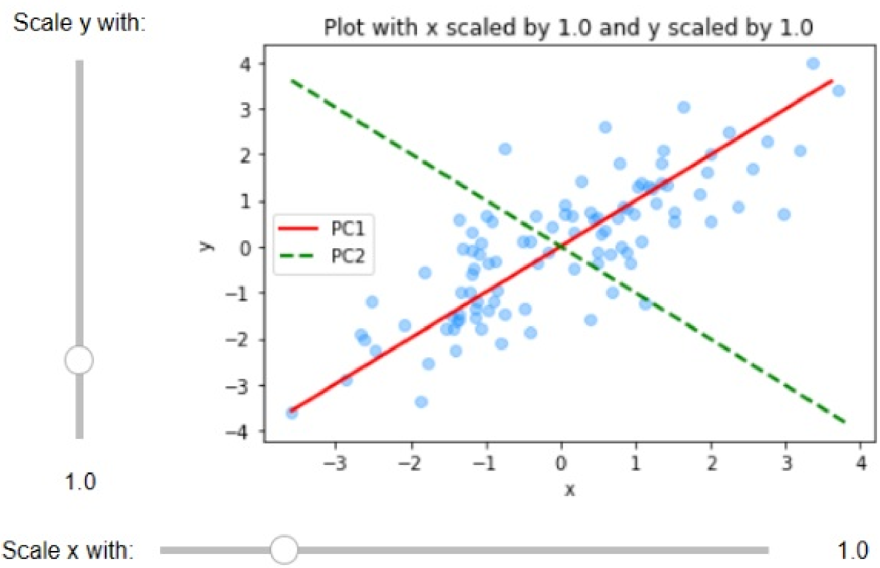
Figure 3.5: PCA: scaling features and generating principal components

3.4. Expert Evaluation of Visualizations After developing the 2 sets of visualizations, these were checked by selected Machine Learning course staff members. This procedure was held out to ensure that the visualizations correctly match the learn- ing objectives and to obtain potential feedback points before carrying out the evaluations. First of all, the gradient descent visualizations were checked. Based on the obtained feedback, a few modifications were made to the Gradient Descent Notebook. First of all, the textual information from the Notebook and some of the knowledge assessing questions focused on the different types of gradient descent algorithms, such as batch or stochastic. However, the visualizations did not include these topics. Therefore, differentiating between different types of gradient descent algorithms was removed both from the Notebook and from the assessment questions, which lead to more simple and straightforward materials. More than that, the Principal Components Analysis visualizations 2 and 3 were developed based on ideas from staff members, therefore they were not checked a second time. However, the first visu- alization was checked. The function line for the error and variance was added based on the received feedback. Also, clear labels of axes were displayed.

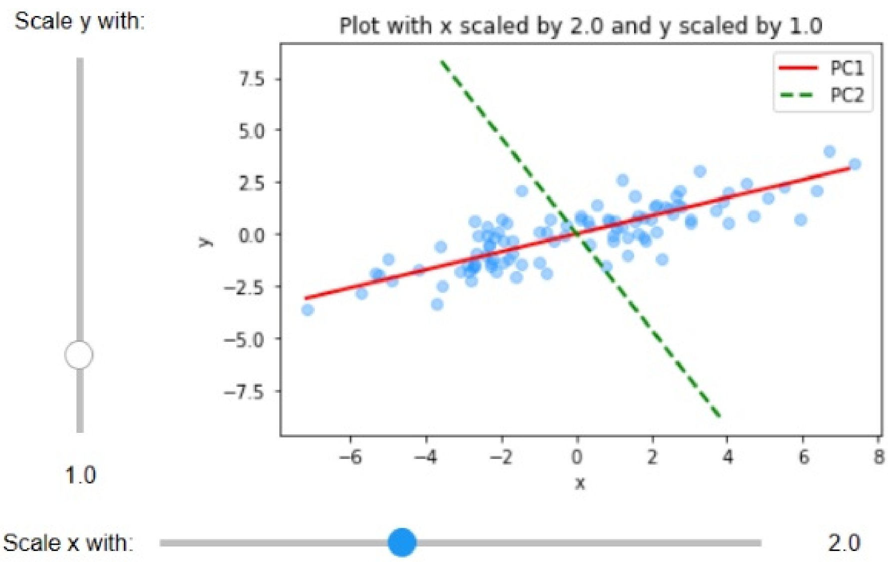
(a) PCA with non-scaled data (x scaled by 1 and y by 1).

(b) PCA with scaled data (x scaled by 2 and y by 1).

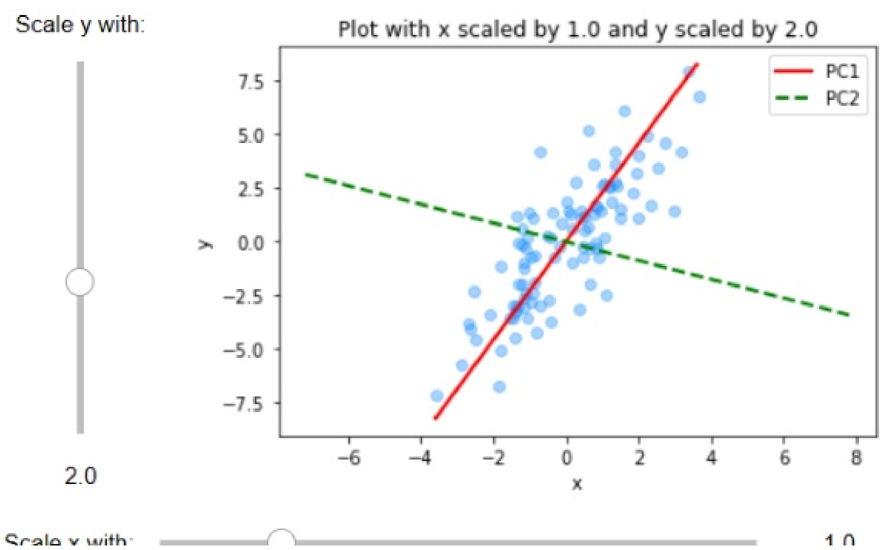
(c) PCA with scaled data (x scaled by 1 and y by 2).



(a) PCA with non-scaled data (x scaled by 1 and y by 1).



(b) PCA with scaled data (x scaled by 2 and y by 1).



3.5. User Evaluation 20

It is important to mention that these are only the most important points of feedback received while the development process benefited from more improvement points. Besides requesting feedback from current ML lecturers, the materials were also given to students who had already followed ML courses during a trial version of the experiment. During this trial, the students read the developed materials and interacted with the visualizations, while providing points of improvement. After this trial, conclusions were added to the Notebooks, possible errors were corrected, such as describing more clearly the third visualization for PCA. This trial was conducted to ensure the best quality for the experiment that followed.

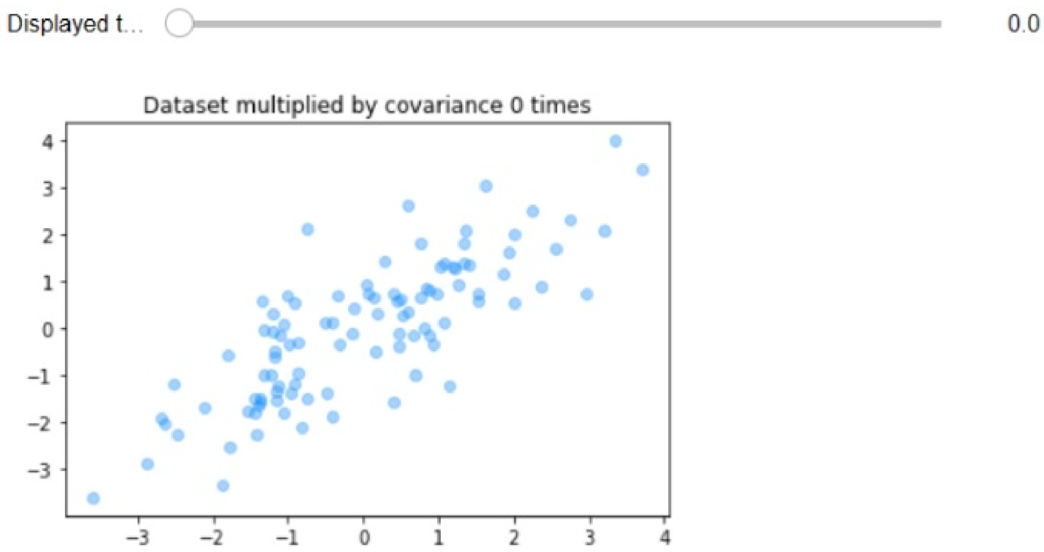
3.5. User Evaluation This section provides additional information about the user evaluation that was performed to study the effects of interactive visualizations in teaching certain ML concepts. Section 3.5.1 details the knowl- edge assessment methods that were used during the experiment, while section 3.5.2 focuses on the motivation assessment of students. Also, section 3.5.3 describes the human research ethics proce- dure.

(a) Initial dataset.

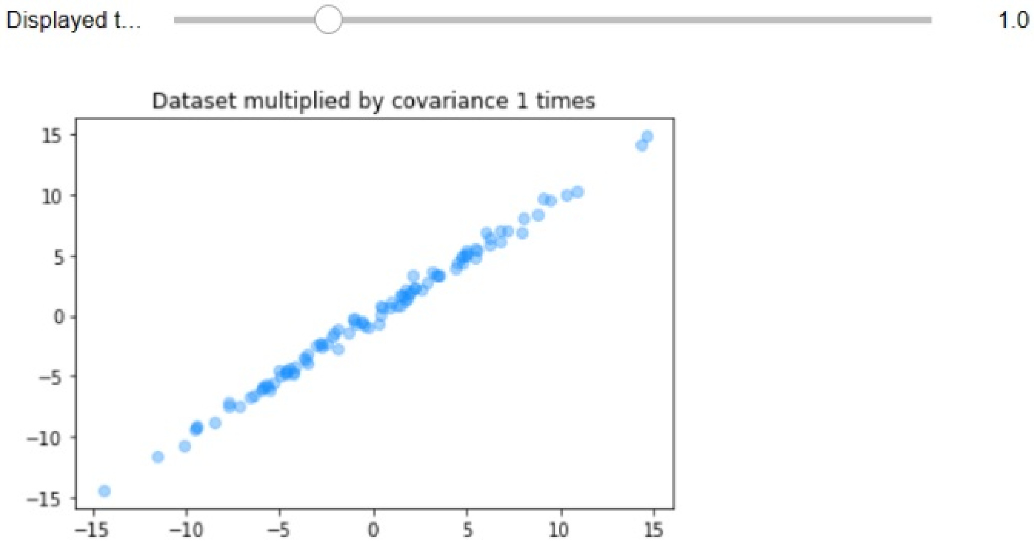
(b) Dataset transformed by the covariance matrix once.

(c) Dataset transformed by the covariance matrix twice.

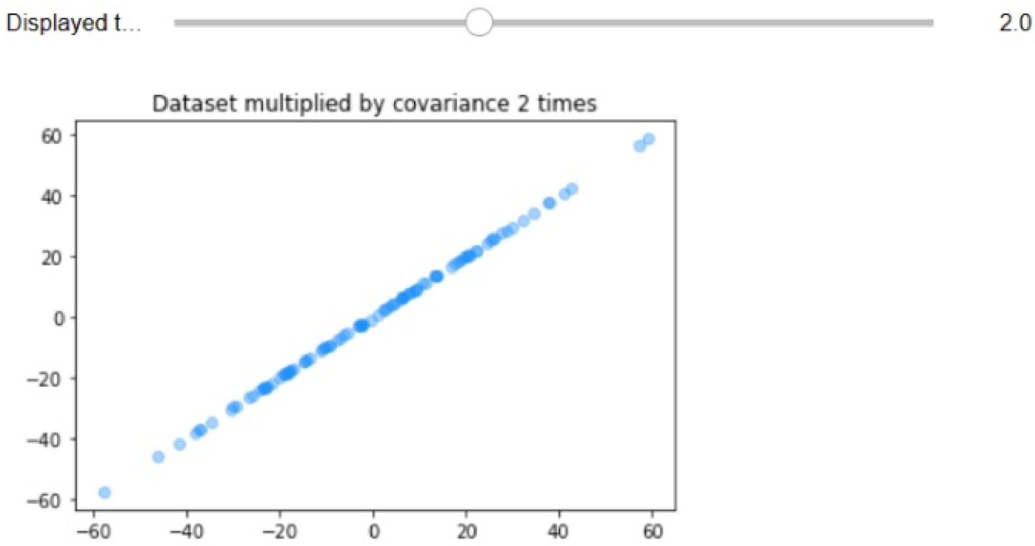
Figure 3.6: PCA: Covariance matrix transformation.



(a) Initial dataset.



(b) Dataset transformed by the covariance matrix once.



(c) Dataset transformed by the covariance matrix twice.

3.5. User Evaluation 21

3.5.1. Knowledge Assessment As previously mentioned, the first aspect measured during the experiment is the knowledge gained by students after exploring the provided materials. This aspect is measured by performing a combination of a pre-test and a post-test, as described in section ??, from the scientific article. Both tests contain the same knowledge assessment questions, namely the ones included in Appendix B. The next two subsections will go into more detail for each of the chosen topics, namely gradient descent and principal component analysis.

Learning Objec- tive

Topic

LO1 What is the primary objective of gradient descent in machine learning?

What is the role of the learning rate in gradient descent? LO2 What is a possible consequence of choosing a too-large learning rate? What is a possible consequence of choosing a too-small learning rate?

Gradient Descent

LO3

LO4 Which of the following statements about PCA is/are incorrect?

LO5 Which eigen pairs would you pick if you want the new data to have 3 dimensions? (given already computed eigen pairs) Principal component analysis (PCA) is the process of computing the principal components (PCs) and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest. Choose the correct statement: - Performing the PCA means: Which statement about PCA is false?

PCA

LO6 Which of the following statements is/are true? PCA is a dimensionality reduction algorithm that minimizes/ maximizes error and variance.

Gradient Descent Questions As mentioned in Section 3.3.1, the visualizations and knowledge assessment questions were built on the basis of the following learning objectives:

• LO1: Explain the gradient descent algorithm and its components. (Understand Conceptual Knowl- edge) • LO2: Infer potential problems of gradient descent and their causes. (Understand Conceptual Knowledge) • LO3: Execute gradient descent update steps. (Apply Procedural Knowledge)

As it can be seen in 3.1, each learning objective can be associated with a few questions used in the experiment. The questions and their possible answers (for multiple choice questions) are included in Appendix A.1. The first three questions aim to assess whether the students understand the gradient descent algo- rithm and its processes. To answer these questions, students need to reflect on the textual materials provided, but also on their interaction, especially with the first visualization. The next five questions focus on whether students can infer potential problems of gradient descent and their causes. In the given materials, there is no mention of these scenarios, disadvantages, or consequences, but they

Which of the following best describes the gradient descent algorithm?

Write two disadvantages of using the gradient descent algorithm. Describe a scenario where gradient descent may get stuck in a local minimum. Give 1 reason why gradient descent is suitable to optimize the mean squared error.

You have the loss function f(x) = 4*x**3 + 10*x**2+5, learning rate 0.1, and current point x = 2. What is the value of the parameter after performing one update?

Table 3.1: Learning objectives per topic and associated questions

Question

Topic	Learning Objective	Question
Gradient Descent	LO1	What is the primary objective of gradient descent in machine learning? Which of the following best describes the gradient descent algorithm? What is the role of the learning rate in gradient descent?
	LO2	What is a possible consequence of choosing a too-large learning rate? What is a possible consequence of choosing a too-small learning rate? Write two disadvantages of using the gradient descent algorithm. Describe a scenario where gradient descent may get stuck in a local minimum. Give 1 reason why gradient descent is suitable to optimize the mean squared error.
	LO3	You have the loss function f(x) = 4*x**3 + 10*x**2+5, learning rate 0.1, and current point x = 2. What is the value of the parameter after performing one update?
PCA	LO4	Which of the following statements about PCA is/are incorrect?
	LO5	Which eigen pairs would you pick if you want the new data to have 3 dimensions? (given already computed eigen pairs) Principal component analysis (PCA) is the process of computing the principal components (PCs) and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest. Choose the correct statement: - Performing the PCA means: Which statement about PCA is false?
	LO6	Which of the following statements is/are true? PCA is a dimensionality reduction algorithm that minimizes/ maximizes error and variance.

Topic	Learning Objective	Question
Gradient Descent	LO1	What is the primary objective of gradient descent in machine learning? Which of the following best describes the gradient descent algorithm? What is the role of the learning rate in gradient descent?
	LO2	What is a possible consequence of choosing a too-large learning rate? What is a possible consequence of choosing a too-small learning rate? Write two disadvantages of using the gradient descent algorithm. Describe a scenario where gradient descent may get stuck in a local minimum. Give 1 reason why gradient descent is suitable to optimize the mean squared error.
	LO3	You have the loss function f(x) = 4*x**3 + 10*x**2+5, learning rate 0.1, and current point x = 2. What is the value of the parameter after performing one update?
PCA	LO4	Which of the following statements about PCA is/are incorrect?
	LO5	Which eigen pairs would you pick if you want the new data to have 3 dimensions? (given already computed eigen pairs) Principal component analysis (PCA) is the process of computing the principal components (PCs) and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest. Choose the correct statement: - Performing the PCA means: Which statement about PCA is false?
	LO6	Which of the following statements is/are true? PCA is a dimensionality reduction algorithm that minimizes/ maximizes error and variance.

have to critically reason about the gradient descent algorithm and its components. Lastly, the students need to execute one gradient descent update in question 9. This aims to test whether they understood the mathematical formula that was given to them and whether they know how to apply it in a specific example. Besides being included in the textual information, the formula is also displayed in the first visualization in the mathematical updates for each iteration. These questions were derived both using previous exams of CSE2510 Machine Learning and TI3145TU Machine Learning and Introduction to AI, courses offered at TU Delft. However, some of them, such as the first ones were derived specifically to test specific learning objectives, without being inspired by the mentioned courses.

Principal Component Analysis Questions As mentioned in Section 3.3.2, there is a clear list of 3 learning objectives tackled during the experiment, namely:

• LO4: Recall Principal Component Analysis elements. (Remember Factual Knowledge) • LO5: Explain Principal Component Analysis and its components. (Understand Conceptual Knowl- edge) • LO6: Infer equivalence between variance and error Principal Component Analysis versions. (Un- derstand Conceptual Knowledge)

In the same manner, questions for the pre and post-test were derived to assess these learning objectives, and their connection is shown in Table 3.1. Also, the full questions for this topic, together with their possible answers, can be seen in Appendix A.2. The first question is related to the first learning objectives, assessing whether students recall how Principal Component Analysis works and what is its application and outcome. Along that, questions 2, 4, and 5, are closely related to explaining the principal component analysis and its components. The difference between these questions and the first question is that the answers were not specifically men- tioned in the materials and they require the student to understand the mechanics behind the algorithm. Lastly, question 3 measures whether the students can infer the equivalence between maximizing the variance and minimizing the error. This learning objective is backed up by the first visualization from the Notebook, since they can see both the variance and the error graphs and they should realize that these two are inversely proportionate. All previously mentioned questions were taken from the CSE2510 Machine Learning course. The decision to inspect the questions and include them in the materials was taken due to the higher level of experience of the mentioned course and its staff members. However, each question was examined to match the contents of the Notebooks, and based on this analysis, one answer was removed from the fourth question.

3.5.2. Motivation Assessment Besides considering the knowledge gained by the students by reading and understanding the materials, the motivation of students was also considered. The current approach of the research combines what is being done by others, namely focusing on the level of understanding or students' experience. To measure the motivation of students, the Instructional Materials Motivation Survey(IMMS) was analyzed, choosing to use the reduced version(RIMMS) proposed and validated by Loorbach et al [42]. This survey is based on the ARCS model [35] and it focuses on the motivation of students regarding instructional materials. Initially, the survey was aimed at in-person instructional materials, but it has since been used in the context of computer-based remote instructional materials [4, 36]. The survey focuses on the four constructs targeted by the ARCS model, namely attention, relevance, confidence, and satisfaction, containing a total of 36 questions. The creators of the reduced version claimed that the length of the initial version might lead to fatigue or boredom [42], motivating the creation of the reduced version. This version only includes 12 questions, 3 questions per construct, and after the validation, it proved to be more efficient in measuring the construct than the original version. Due to this improvement, the reduced version of the survey was chosen for the current research. This decision was especially important due to the already long process of going through the materials and understanding them. This survey instructs people to rate a list of specific statements on a scale from not true to very true, as it can be seen in Appendix B.1. After collecting the answers, they are matched to numbers from 1 to

5. These numbers are used to calculate the average score for each construct and the average score overall for motivation.

3.5.3. Human Research Ethics The conducted experiment involves humans as they are the main target and focus. Following the rules and practices of TU Delft, "all research involving Human Research Subjects – including Master's theses – requires approval from the Human Research Ethics Committee (HREC) before it can go ahead." [1]. For this reason, before starting the previously described experiment, an application was submitted to HREC and approval was obtained from the committee to conduct the research. The main reason for this procedure is to ensure that the participants are exposed to a risk as low as possible. In the given setting, risks can include collecting sensitive data, mishandling the data, or being pressured to participate in the experiment. As stated in the HREC application forms, the current experiment does not collect any personally identifiable data, since it is not relevant to the research, it stores all data in places approved by the current practice of TU Delft, and the participants are free to choose if they want to take part in the experiment or whether they want to opt out of the experiment. More than that, to ensure the minimal risk nature of the experiment, Microsoft Forms is used to gather the responses from the survey, and to store them.

[1] URL: https://www.tudelft.nl/strategie/integriteitsbeleid/human-research-ethics.

[2] Mojeed Kolawole Akinsola and IA Animasahun. "The Effect of Simulation-Games Environment on Students Achievement in and Attitudes to Mathematics in Secondary Schools." In: Online Submission 6.3 (2007).

[3] Svetlana Asmuss, Natalja Budkina, et al. "On usage of visualization tools in teaching mathematics at universities". In: Engineering for Rural Development 18 (2019), pp. 1962–1969.

[4] Hermann Astleitner and Manuela Hufnagl. "The effects of situation-outcome-expectancies and of ARCS-strategies on self-regulated learning with web-lectures". In: Journal of educational mul- timedia and hypermedia 12.4 (2003), pp. 361–376.

[5] Nouf M Al-Barakati and Arwa Y Al-Aama. “The effect of visualizing roles of variables on student performance in an introductory programming course”. In: Proceedings of the 14th annual ACM SIGCSE conference on Innovation and technology in computer science education. 2009, pp. 228– 232.

[6] Alan J Bishop. “What values do you teach when you teach mathematics?” In: Teaching Children Mathematics 7.6 (2001), pp. 346–346.

[7] Christopher M Bishop and Nasser M Nasrabadi. Pattern recognition and machine learning. Vol. 4. 4. Springer, 2006.

[8] Robert F Boruch. Randomized experiments for planning and evaluation: A practical guide. Vol. 44. Sage, 1997.

[9] Sarah M Brown. “Participatory Live Coding and Learning-Centered Assessment in Programming for Data Science”. In: Proceedings of the Second Teaching Machine Learning and Artificial Intel- ligence Workshop. PMLR. 2022, pp. 73–77.

[10] Peter Brusilovsky, Jae-wook Ahn, and Edie Rasmussen. “Teaching information retrieval with web-based interactive visualization”. In: Journal of Education for Library and Information Science (2010), pp. 187–200.

[11] Michael D Byrne, Richard Catrambone, and John T Stasko. “Evaluating animations as student aids in learning computer algorithms”. In: Computers & education 33.4 (1999), pp. 253–278.

[12] Michelene TH Chi and Ruth Wylie. “The ICAP framework: Linking cognitive engagement to active learning outcomes”. In: Educational psychologist 49.4 (2014), pp. 219–243.

[13] Ton De Jong and Wouter R Van Joolingen. “Scientific discovery learning with computer simula- tions of conceptual domains”. In: Review of educational research 68.2 (1998), pp. 179–201.

[14] Michael Delucchi. “Measuring student learning in social statistics: A pretest-posttest study of knowledge gain”. In: Teaching Sociology 42.3 (2014), pp. 231–239.

[15] Gulustan Dogan. “Teaching Machine Learning with Applied Interdisciplinary Real World Projects”. In: The Third Teaching Machine Learning and Artificial Intelligence Workshop. PMLR. 2023, pp. 12–15.

[16] Jan Ebert et al. “Hearts Gym: Learning Reinforcement Learning as a Team Event”. In: The Third Teaching Machine Learning and Artificial Intelligence Workshop. PMLR. 2023, pp. 16–21.

[17] Andrew Elby. “Helping physics students learn how to learn”. In: American Journal of Physics 69.S1 (2001), S54–S64.

[18] Timothy Ellis. “Animating to build higher cognitive understanding: A model for studying multimedia effectiveness in education”. In: Journal of Engineering Education 93.1 (2004), pp. 59–64.

[19] Megan Loef Franke and Elham Kazemi. “Learning to teach mathematics: Focus on student think- ing”. In: Theory into practice 40.2 (2001), pp. 102–109.

References

[20] Scott Freeman et al. “Active learning increases student performance in science, engineering, and mathematics”. In: Proceedings of the national academy of sciences 111.23 (2014), pp. 8410– 8415.

[21] Joan Garfield and Dani Ben-Zvi. “How students learn statistics revisited: A current review of re- search on teaching and learning statistics”. In: International statistical review 75.3 (2007), pp. 372– 396.

[22] Graham Gibbs. “Learning by doing: A guide to teaching and learning methods”. In: Further Edu- cation Unit (1988).

[23] David Hammer. “Discovery learning and discovery teaching”. In: Cognition and instruction 15.4 (1997), pp. 485–529.

[24] Ken Hasselmann and Quentin Lurkin. “Stimulating student engagement with an AI board game tournament”. In: arXiv preprint arXiv:2304.11376 (2023).

[25] Curtis R Henrie, Lisa R Halverson, and Charles R Graham. “Measuring student engagement in technology-mediated learning: A review”. In: Computers & Education 90 (2015), pp. 36–53.

[26] Christian Herta et al. “Deep teaching: materials for teaching machine and deep learning”. In: HEAD’19. 5th International Conference on Higher Education Advances. Editorial Universitat Politèc- nica de València. 2019, pp. 1153–1131.

[27] Andrea Hoffkamp. “The use of interactive visualizations to foster the understanding of concepts of calculus: design principles and empirical results”. In: ZDM 43 (2011), pp. 359–372.

[28] Christopher D Hundhausen and Sarah A Douglas. “Low-fidelity algorithm visualization”. In: Jour- nal of Visual Languages & Computing 13.5 (2002), pp. 449–470.

[29] Maria-Blanca Ibanez, Angela Di-Serio, and Carlos Delgado-Kloos. “Gamification for engaging computer science students in learning activities: A case study”. In: IEEE Transactions on learning technologies 7.3 (2014), pp. 291–301.

[30] D Mills Jamie. “Using computer simulation methods to teach statistics: A review of the literature”. In: Journal of Statistics Education 10.1 (2002).

[31] Lili Jiang. A visual explanation of gradient descent methods (momentum, AdaGrad, RMSProp, adam). June 2020. URL: https://towardsdatascience.com/a- visual- explanation- of- gradient-descent-methods-momentum-adagrad-rmsprop-adam-f898b102325c.

[32] Deepti Prit Kaur, Archana Mantri, and Ben Horan. "Enhancing student motivation with use of aug- mented reality for interactive learning in engineering education". In: Procedia Computer Science 172 (2020), pp. 881–885.

[33] Hussain Syed Kazmi. "Teaching machine learning to energy engineers through end-to-end deci- sion making". In: https://openreview. net/forum? id= 4rOjJXlcU0. 2021.

[34] John M Keller. Motivational design for learning and performance: The ARCS model approach. Springer Science & Business Media, 2009.

[35] John M Keller. "Motivational design of instruction". In: Instructional design theories and models: An overview of their current status 1.1983 (1983), pp. 383–434.

[36] John M Keller. "Using the ARCS motivational process in computer-based instruction and distance education". In: New directions for teaching and learning 1999.78 (1999), pp. 37–47.

[37] Katherine M Kinnaird. "Teaching Computational Machine Learning (without Statistics)". In: Euro- pean Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. PMLR. 2021, pp. 17–22.

[38] David R Krathwohl. "A revision of Bloom's taxonomy: An overview". In: Theory into practice 41.4 (2002), pp. 212–218.

[39] Shailesh S Kulkarni et al. "Dynamic interactive visualizations: Implications of seeing, doing, and playing for quantitative analysis pedagogy". In: INFORMS Transactions on Education 19.3 (2019), pp. 121–142.

[41] Hai-Ning Liang and Kamran Sedig. "Can interactive visualization tools engage and support pre- university students in exploring non-trivial mathematical concepts?" In: Computers & Education 54.4 (2010), pp. 972–991.

[42] Nicole Loorbach et al. "Validation of the Instructional Materials Motivation Survey (IMMS) in a self- directed instructional setting aimed at working with technology". In: British journal of educational technology 46.1 (2015), pp. 204–218.

[43] Feiyu Lu et al. "Evaluating engagement level and analytical support of interactive visualizations in virtual reality environments". In: 2018 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE. 2018, pp. 143–152.

[44] Richard E Mayer. "Should there be a three-strikes rule against pure discovery learning?" In: Amer- ican psychologist 59.1 (2004), p. 14.

[45] Joel Michael. "Where's the evidence that active learning works?" In: Advances in physiology education (2006).

[46] Tom M Mitchell. Machine learning. 1997.

[47] Yvonne G Mulder, Ard W Lazonder, and Ton de Jong. "Simulation-based inquiry learning and computer modeling: Pitfalls and potentials". In: Simulation & Gaming 46.3-4 (2015), pp. 322– 347.

[48] Elena Nardi. "Reflections on visualization in mathematics and in mathematics education". In: Mathematics & mathematics education: searching for common ground (2014), pp. 193–220.

[49] Andrew Ng. "CS229 Lecture notes". In: CS229 Lecture notes 1.1 (2000), pp. 1–3.

[50] Barbara M Olds, Barbara M Moskal, and Ronald L Miller. "Assessment in engineering educa- tion: Evolution, approaches and future collaborations". In: Journal of Engineering Education 94.1 (2005), pp. 13–25.

[51] Margus Pedaste et al. "Phases of inquiry-based learning: Definitions and the inquiry cycle". In: Educational research review 14 (2015), pp. 47–61.

[52] Michael Prince. "Does active learning work? A review of the research". In: Journal of engineering education 93.3 (2004), pp. 223–231.

[53] Edward F Redish. "Implications of cognitive studies for teaching physics". In: American Journal of Physics 62.9 (1994), pp. 796–803.

[54] Lloyd P Rieber. "Supporting discovery-based learning within simulations". In: Cognitive effects of multimedia learning. IGI Global, 2009, pp. 217–236.

[55] Dilal Saundage et al. "Teaching data analysis with interactive visual narratives". In: Journal of Information Systems Education 27.4 (2016), pp. 233–248.

[56] Dino Schweitzer and Wayne Brown. "Interactive visualization for the active learning classroom". In: Proceedings of the 38th SIGCSE technical symposium on Computer science education. 2007, pp. 208–212.

[57] Kamran Sedig and Mark Sumner. "Characterizing interaction with visual mathematical represen- tations". In: International Journal of Computers for Mathematical Learning 11 (2006), pp. 1–55.

[58] Jinwook Seo and Ben Shneiderman. "Knowledge discovery in high dimensional data: case stud- ies and a user survey for an information visualization tool". In: IEEE Transactions on Visualization and Computer Graphics 12.3 (2006), pp. 311–322.

[59] Levent Sevgi. "Modeling and simulation concepts in engineering education: virtual tools". In: Turk- ish Journal of Electrical

Engineering and Computer Sciences 14.1 (2006), pp. 113–127.

[60] Omar Shouman, Simon Fuchs, and Holger Wittges. “Experiences from teaching practical ma- chine learning courses to master’s students with mixed backgrounds”. In: Proceedings of the Second Teaching Machine Learning and Artificial Intelligence Workshop. PMLR. 2022, pp. 62– 67.

[62] Teach LA Dev Team. Gradient Descent Visualiser. URL: https://uclaacm.github.io/gradie nt-descent-visualiser/.

[63] Andrew R Webb, Keith D Copsey, and Gavin Cawley. Statistical pattern recognition. Vol. 2. Wiley Online Library, 2011.

[64] Hilde Jacoba Petronella Weerts and Mykola Pechenizkiy. “Teaching responsible machine learn- ing to engineers”. In: Proceedings of the Second Teaching Machine Learning and Artificial Intel- ligence Workshop. PMLR. 2022, pp. 40–45.

[65] Peter H Westfall, Andrea L Arias, and Lawrence V Fulton. “Teaching principal components using correlations”. In: Multivariate behavioral research 52.5 (2017), pp. 648–660.

A.1. Gradient Descent Questions

1. What is the primary objective of gradient descent in machine learning?

(a) Maximizing the accuracy of the model (b) Minimizing the loss function

(c) Maximizing the number of features (d) Minimizing the number of iterations (e) I don’t know

2. Which of the following best describes the gradient descent algorithm?

(a) It finds the maximum of a function by moving opposite to the gradient direction. (b) It randomly searches for the minimum of a function.

(c) It finds the minimum of a function by moving in a zig-zag pattern. (d) It finds the minimum of a function by moving opposite to the gradient direction. (e) I don’t know

3. What is the role of the learning rate in gradient descent?

(a) It decides the maximum value of the loss function. (b) It determines the number of iterations needed for convergence.

(c) It determines the size of the steps taken during optimization. (d) It controls the number of features in the model. (e) I don’t know

4. What is a possible consequence of choosing a too-large learning rate? 5. What is a possible consequence of choosing a too-small learning rate? 6. Write two disadvantages of using the gradient descent algorithm. 7. Describe a scenario where gradient descent may get stuck in a local minimum. 8. Give 1 reason why gradient descent is suitable to optimize the mean squared error. 9. You have the loss function $f(x) = 4 \ast x3 + 10 \ast x2 + 5$, learning rate 0.1, and current point x = 2. What is the value of the parameter after performing one update?

A.2. Principal Component Analysis Questions

1. Which of the following statements about PCA is/are incorrect?(multiple answers allowed)

(a) The component we find become the new dimensions of our data. (b) The principal components are ordered from those that account for the largest variance to the ones that account for the smallest variance in the data. (c) The principal components are given by the eigenvectors of the data matrix. (d) PCA returns a subset of the original dimensions.

Knowledge Assessment

A

(e) I don’t know.

2. For a given dataset represented by feature vectors of length 6 (6 features per datapoint), the PCA technique is performed and resulted in the following eigen pairs:

Which eigenpairs would you pick if you want the new data to have 3 dimensions?(exactly 3 an- swers allowed)

(a) eigen pair 1 (b) eigen pair 2

(c) eigen pair 3 (d) eigen pair 4 (e) eigen pair 5

(f) eigen pair 6

3. Which of the following statements is/are true? PCA is a dimensionality reduction algorithm which:(multiple answers allowed)

(a) maximizes variance (b) minimizes variance

(c) maximizes reconstruction error (d) minimizes reconstruction error (e) I don't know

4. Principal component analysis (PCA) is the process of computing the principal components (PCs) and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest. Choose the correct statement:

-Performing the PCA means:

(a) projecting the data in the direction of each PC, and scaling it by a factor of **λ** (eigenvalue) of that direction (eigenvector). (b) projecting the data in the direction of each PC, without scaling it by a factor of **λ** (eigenvalue) of that direction (eigenvector). (c) translating the data to be zero mean and normalize to have unit variance. (d) I don't know

5. Which statement about PCA is false?

(a) The larger the number of principal components for PCA, the smaller the reconstruction error. (b) Principal components founded by PCA are always orthogonal.

(c) Reducing the dimensionality using PCA must not be done as preprocessing step before applying a classification model. (d) Two principal components are often used for visualizations purposes using PCA. (e) I don't know

B.1. Reduced Instructional Materials Motivation Survey(RIMMS) Following the procedure from [42], the students were asked to rate each statement on a scale from not true to very true(not true, slightly true, moderately true, mostly true, very true)

1. The quality of the writing helped to hold my attention. 2. The way the information is arranged on the pages helped keep my attention. 3. The variety of reading passages, exercises, illustrations, etc, helped keep my attention on the lesson. 4. It is clear to me how the content of this material is related to things I already know. 5. The content and style of writing in this lesson convey the impression that its content is worth knowing. 6. The content of this lesson will be useful to me. 7. As I worked on this lesson, I was confident that I could learn the content. 8. After working on this lesson for a while, I was confident that I would be able to pass a test on it. 9. The good organization of the content helped me be confident that I would learn this material. 10. I enjoyed this lesson so much that I would like to know more about this topic.

11. I really enjoyed studying this lesson. 12. It was a pleasure to work on such a well-designed lesson.

Motivation Assessment

30

B

B

As previously mentioned, the research included the development of 2 Jupyter Notebooks each accom- panied by 3 interactive visualizations and tackling one of the two topics. In this section, the exact materials that were developed and used in the experiments are displayed.

Interactive Visualizations

31

C

]:

```python
from GD_utils import GD_2D, GD_3D, GD_manual
plot = GD_2D()
plot
```

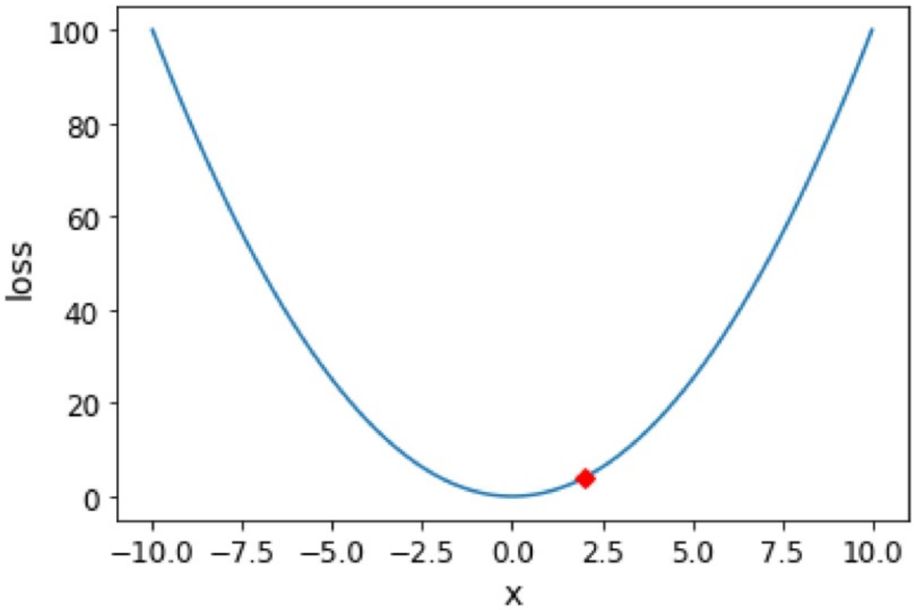
Starting Value: 2.0(

Learning Rate: 0.1(

Selected Cost Function: Least Squares

Number of iterations displayed: 1

☐ Display on plot the gradient used in the last update.



```
plot = GD_manual()
plot
```

0.00 -dJ/dw0 = +24.898778

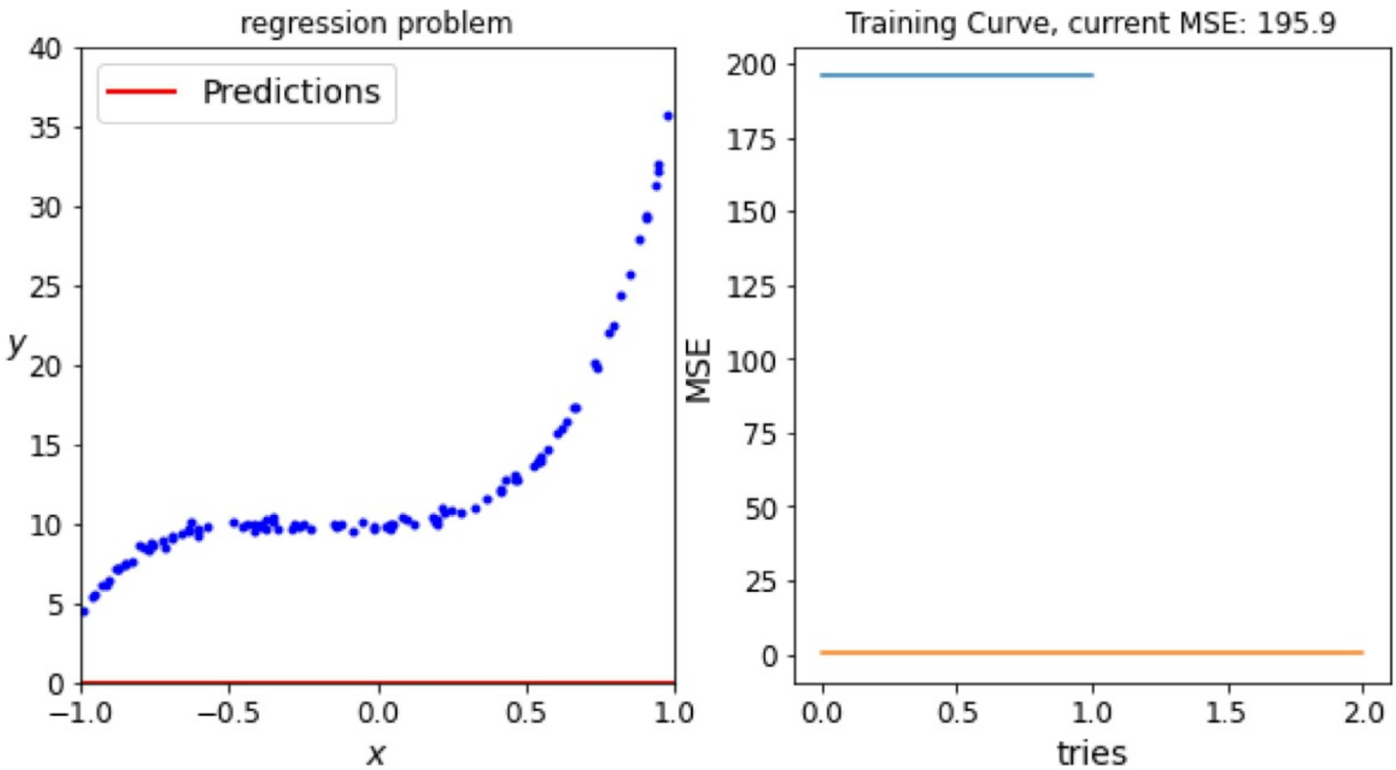
0.00 -dJ/dw1 = +4.573791

0.00 -dJ/dw2 = +10.201402

0.00 -dJ/dw3 = +3.345502

0.00 -dJ/dw4 = +6.553647

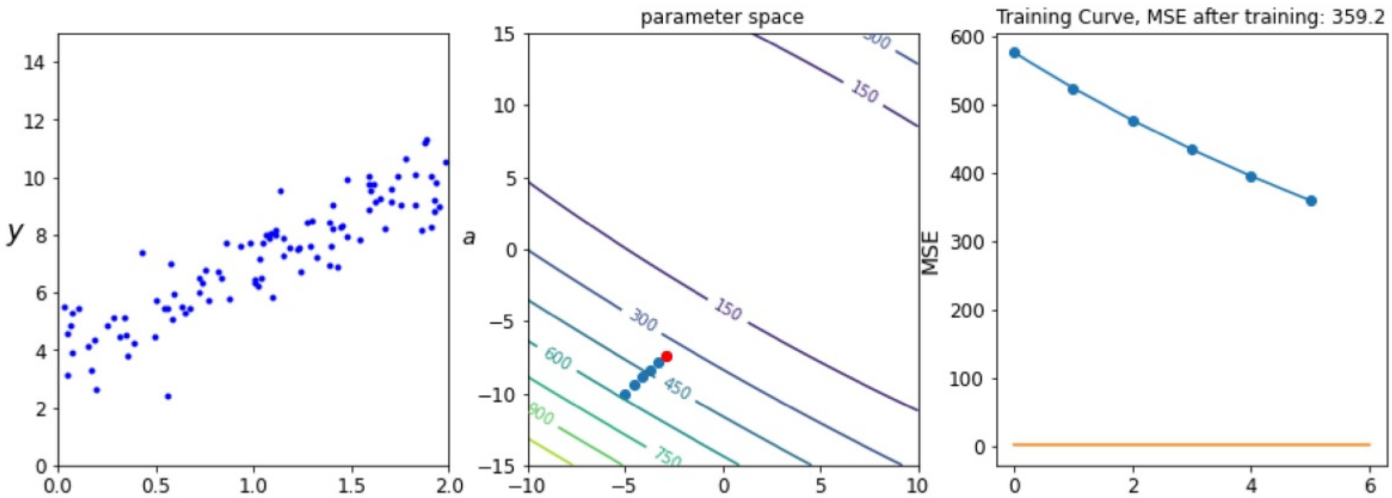
0.00 -dJ/dw5 = +2.664131



```
plot = GD_3D()
plot
```

Starting value a -10.00 Starting value b -5.00

Learning rate 0.0100 Epochs 5



the dataset) is an eigenvector of the covariance matrix, and λ is the eigenvalue. As you remember from Linear Algebra, the eigenvector of a transformation is the vector which after applying the transformation is only being scaled.

It is also important to mention that the eigenvalues are proportionate to the variance, namely the eigenvector with the highest eigenvalue retains the highest amount of variance from the initial dataset, and so on.

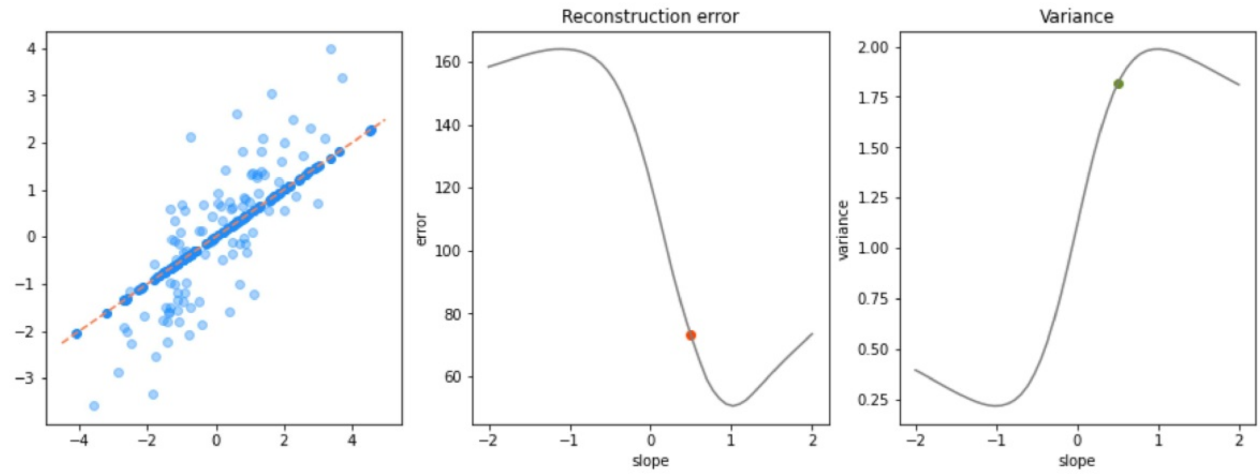
Running the following code will generate a visualization showing a dataset, together with a line.

You can interact with the visualization by changing the slope of the line. When you change the line, you will see in the graphs on the right the effect on the reconstruction error and variance associated with each transformation.

As you interact with the visualization, can you see any link between the variance and error of the transformed dataset?

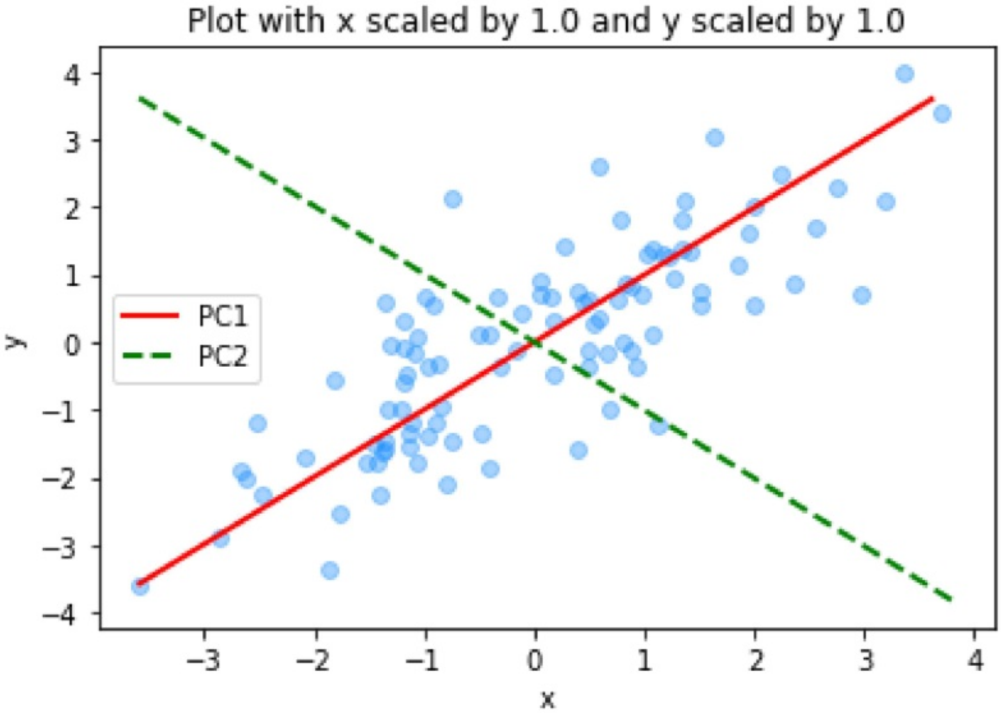
[1]:
```
from PCA_utils import pca_vis1, pca_scaling, pca_cov
plot = pca_vis1()
plot
```

Line gradient:



```
1 plot = pca_scaling()
2 plot
```

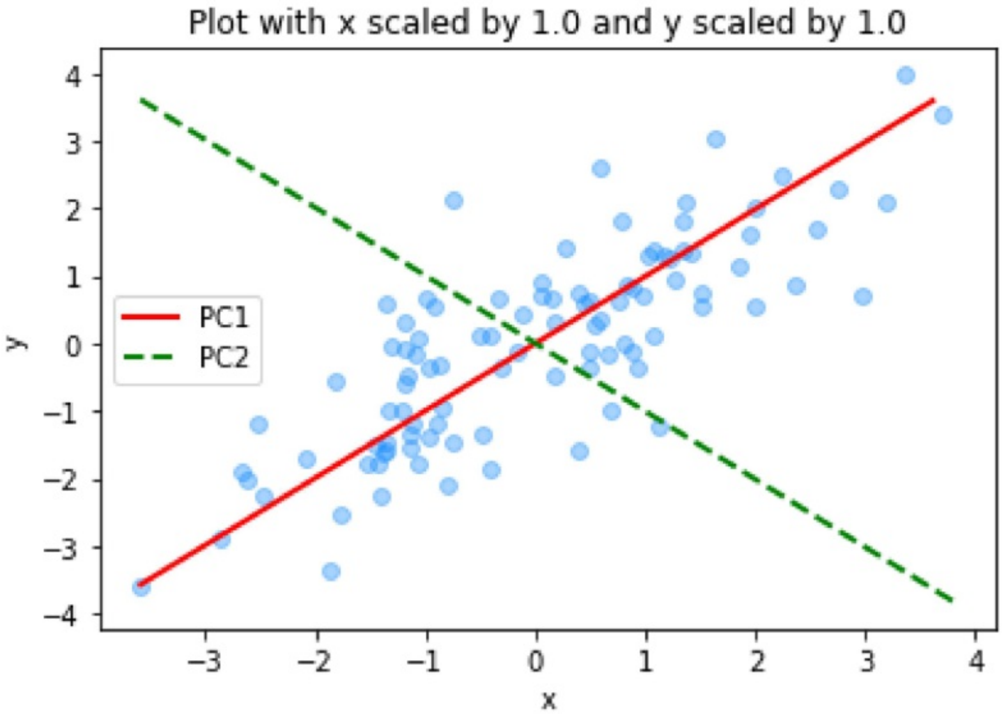
Scale y with:

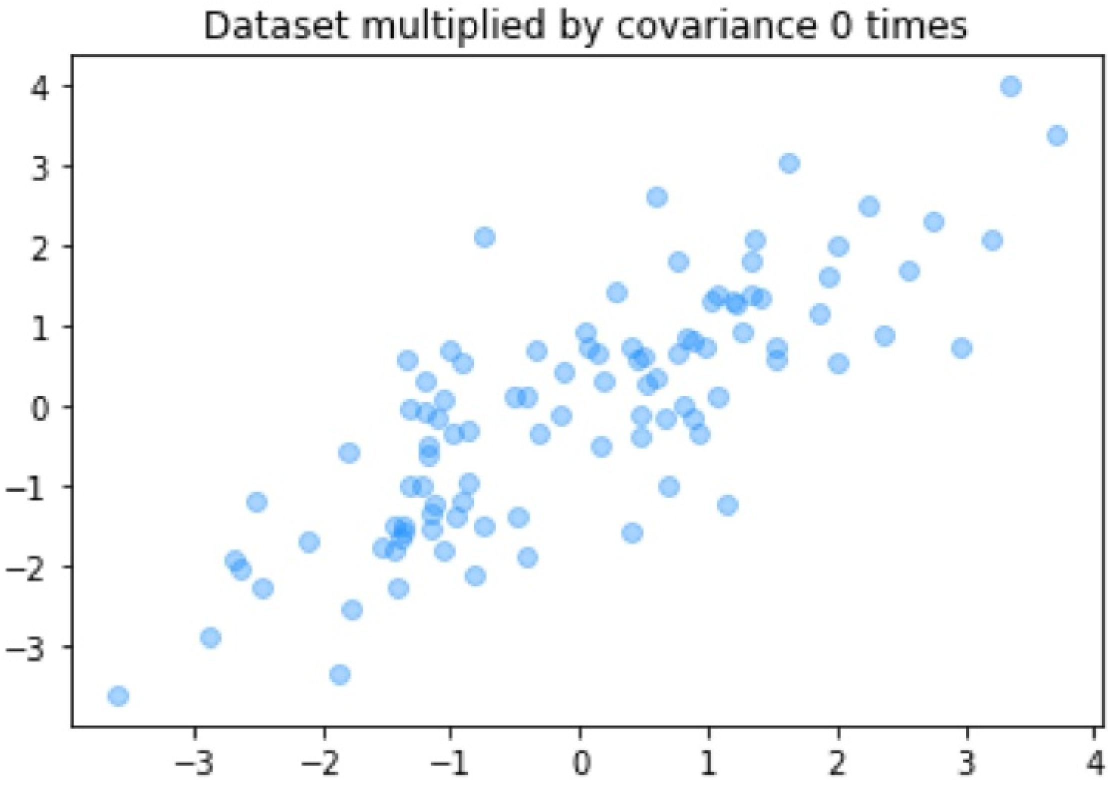


1.0

Scale x with: 1.0

None





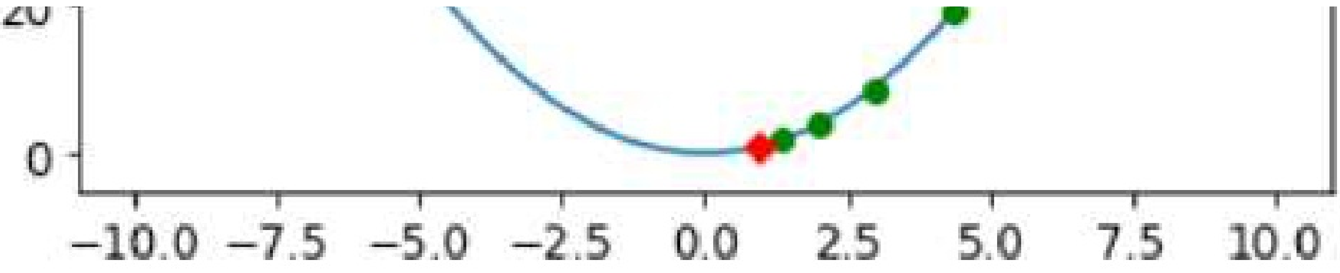
The conducted controlled experiment required a control group that received a version of the previous Jupyter notebooks including static visualizations. In this section, the exact materials that were given to the control group are shown.

Static Visualizations

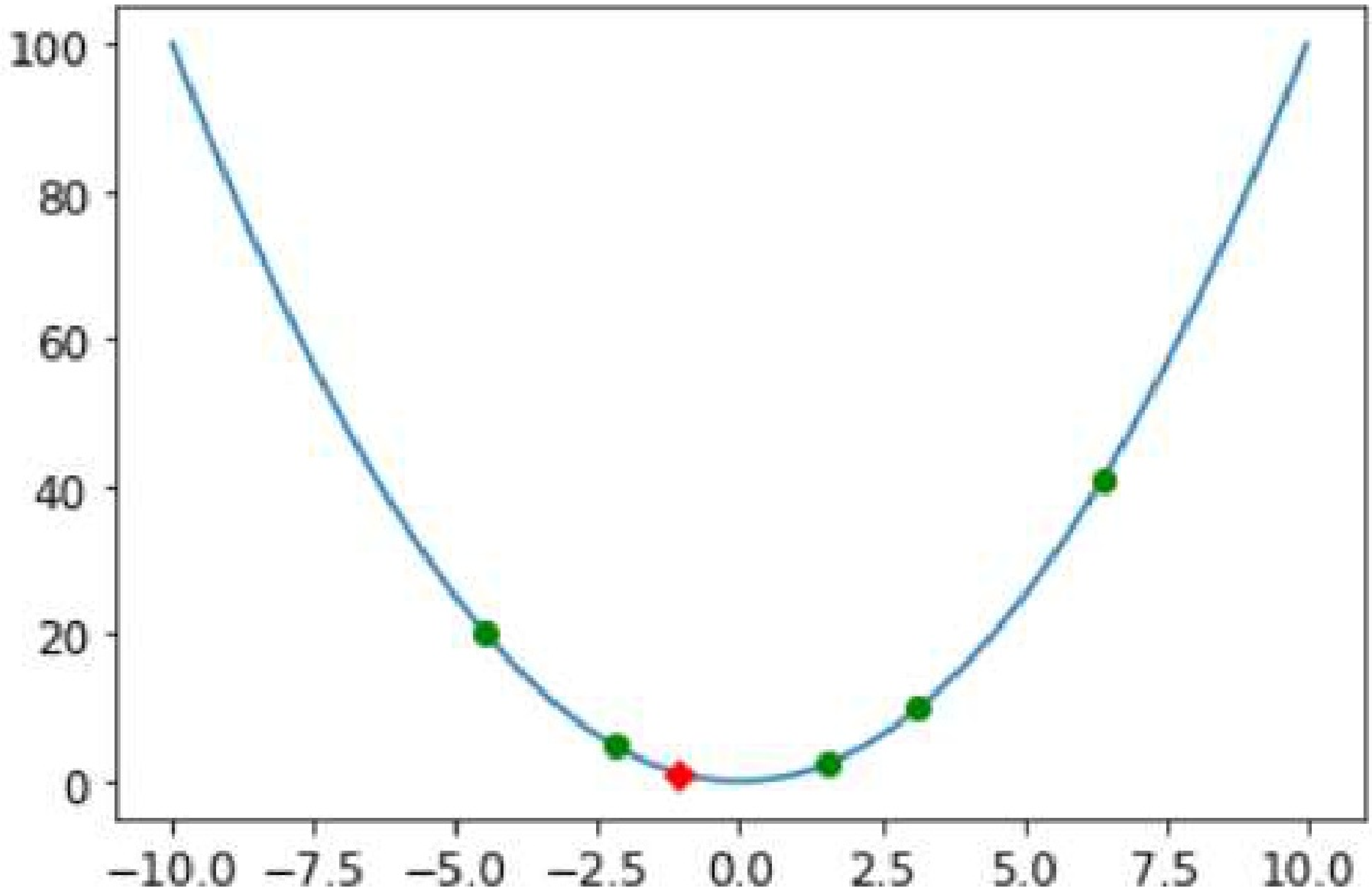
40

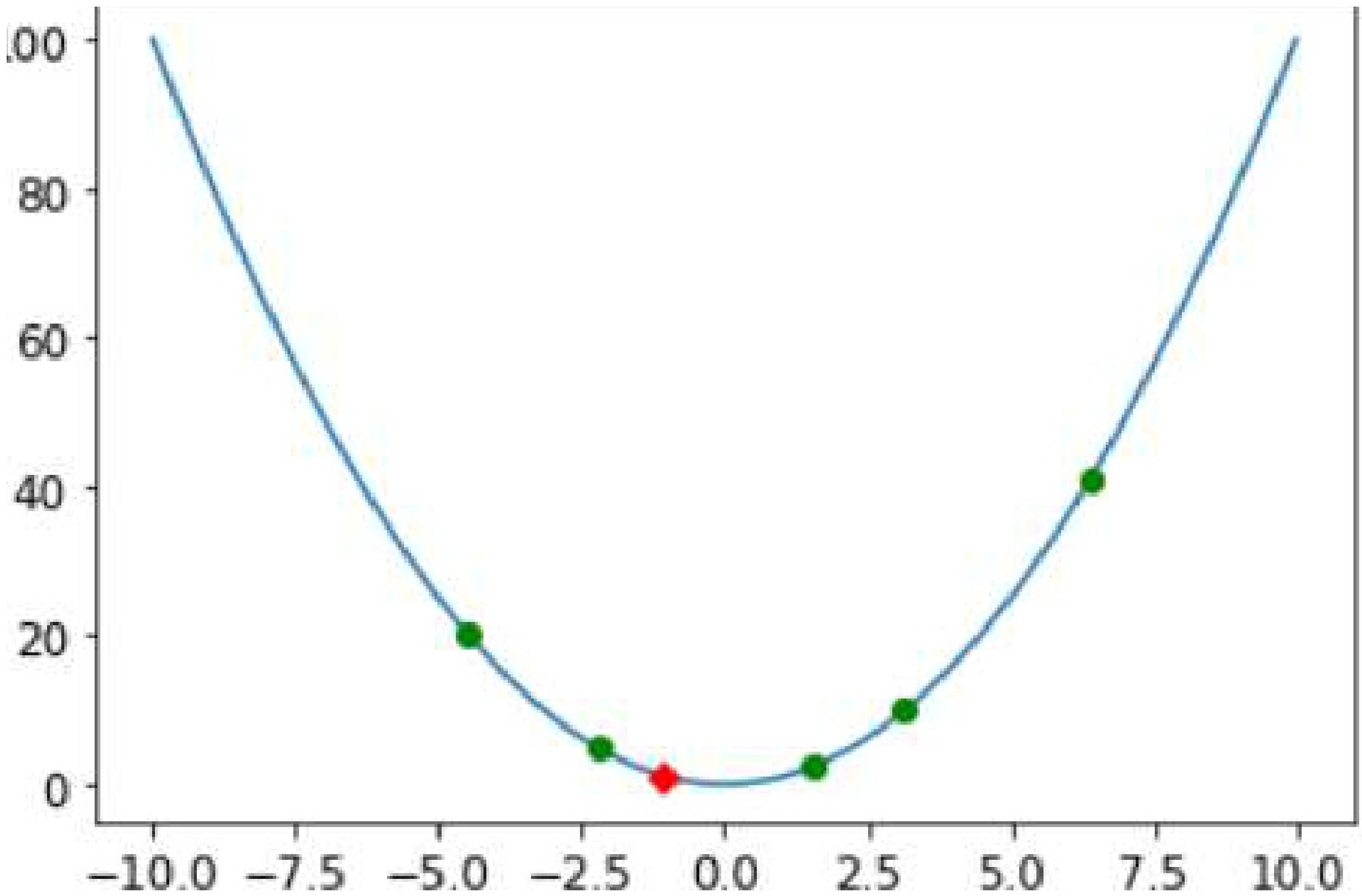
D

D

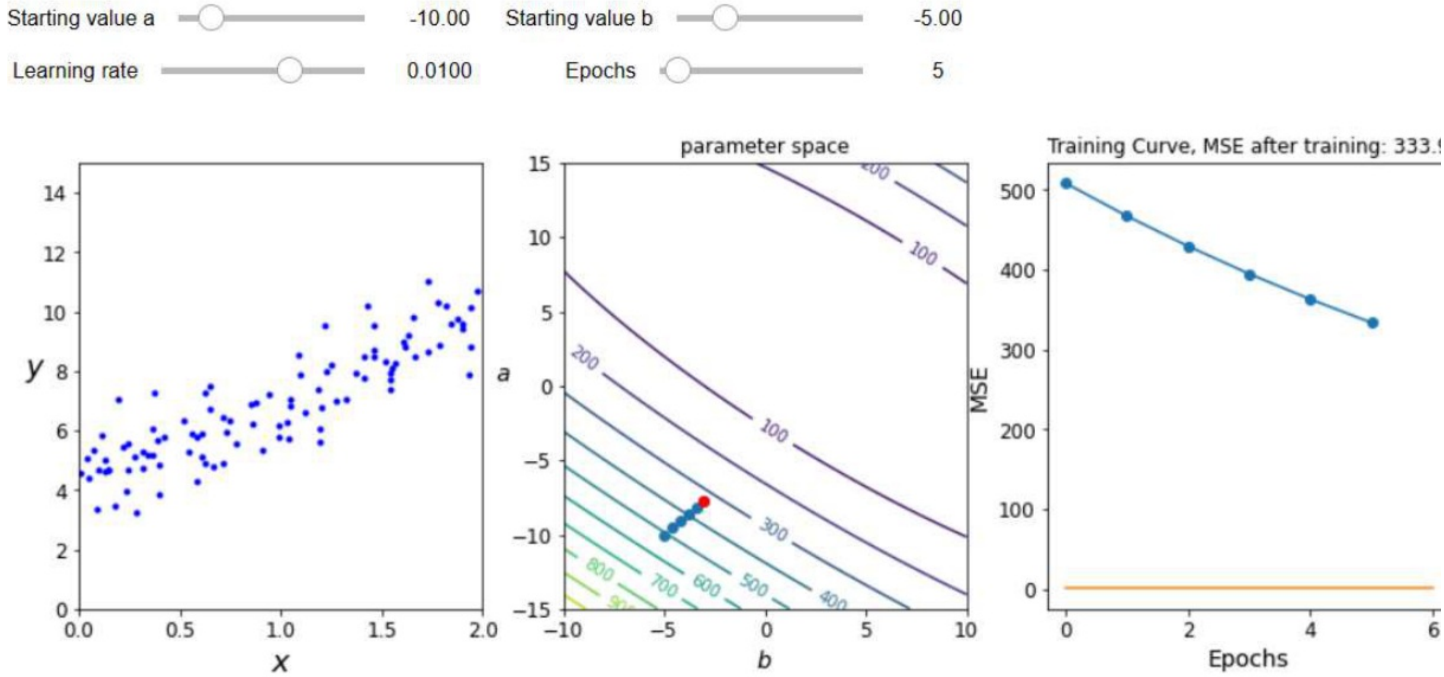


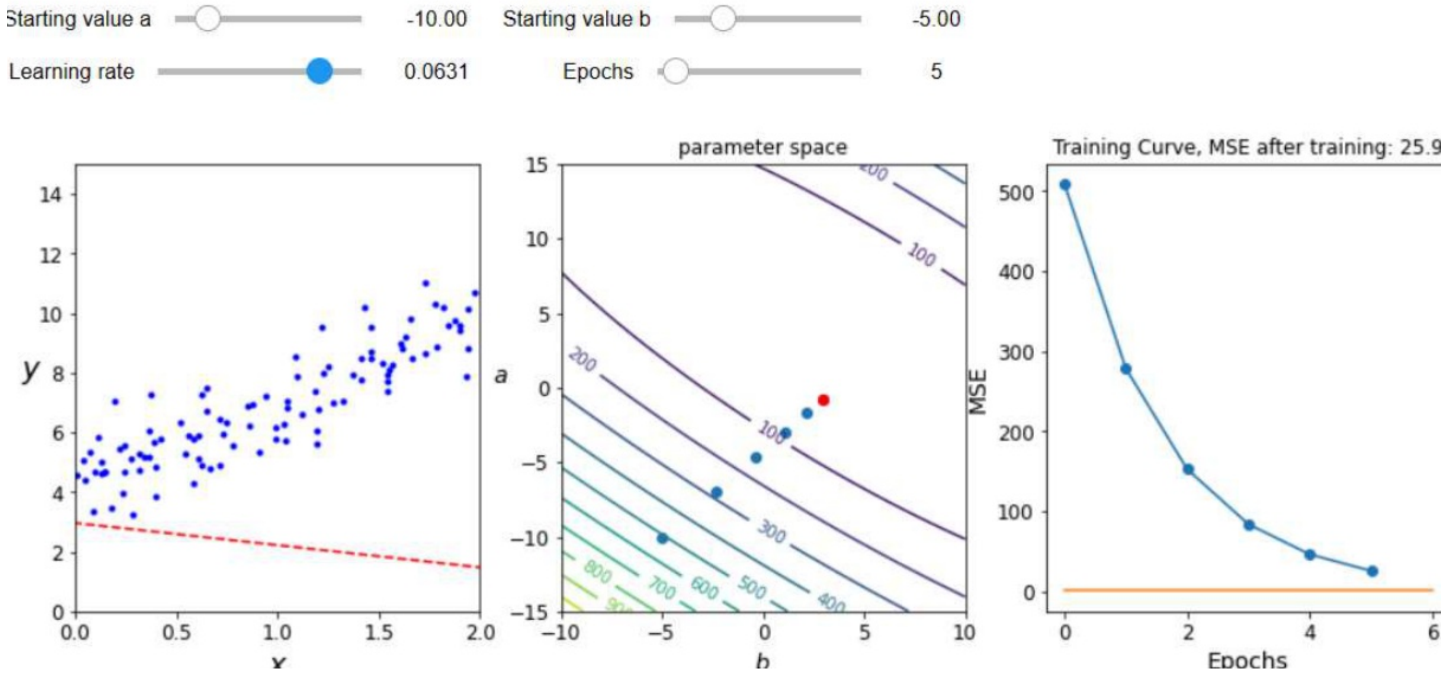
ehavior with a higher learning rate:

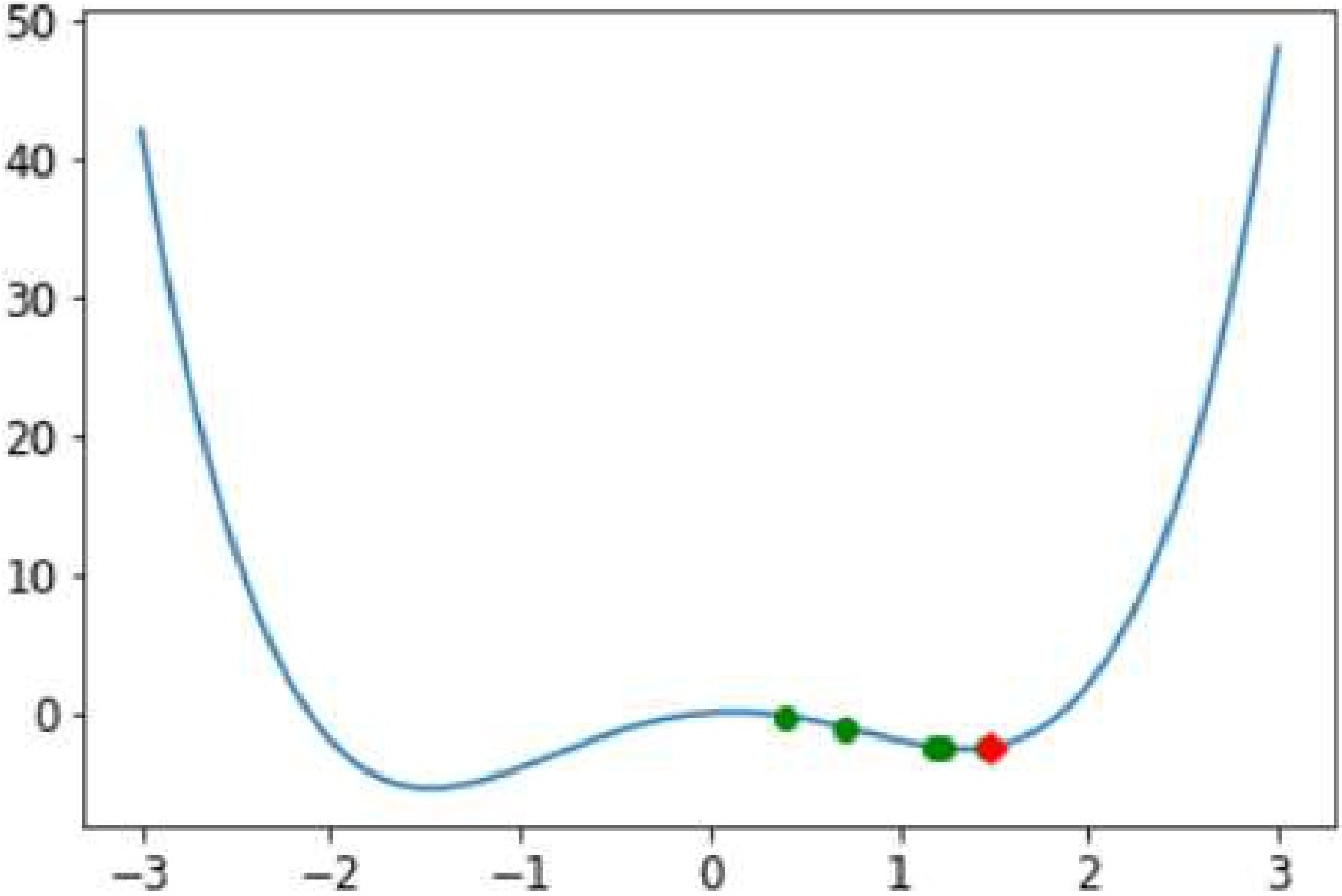


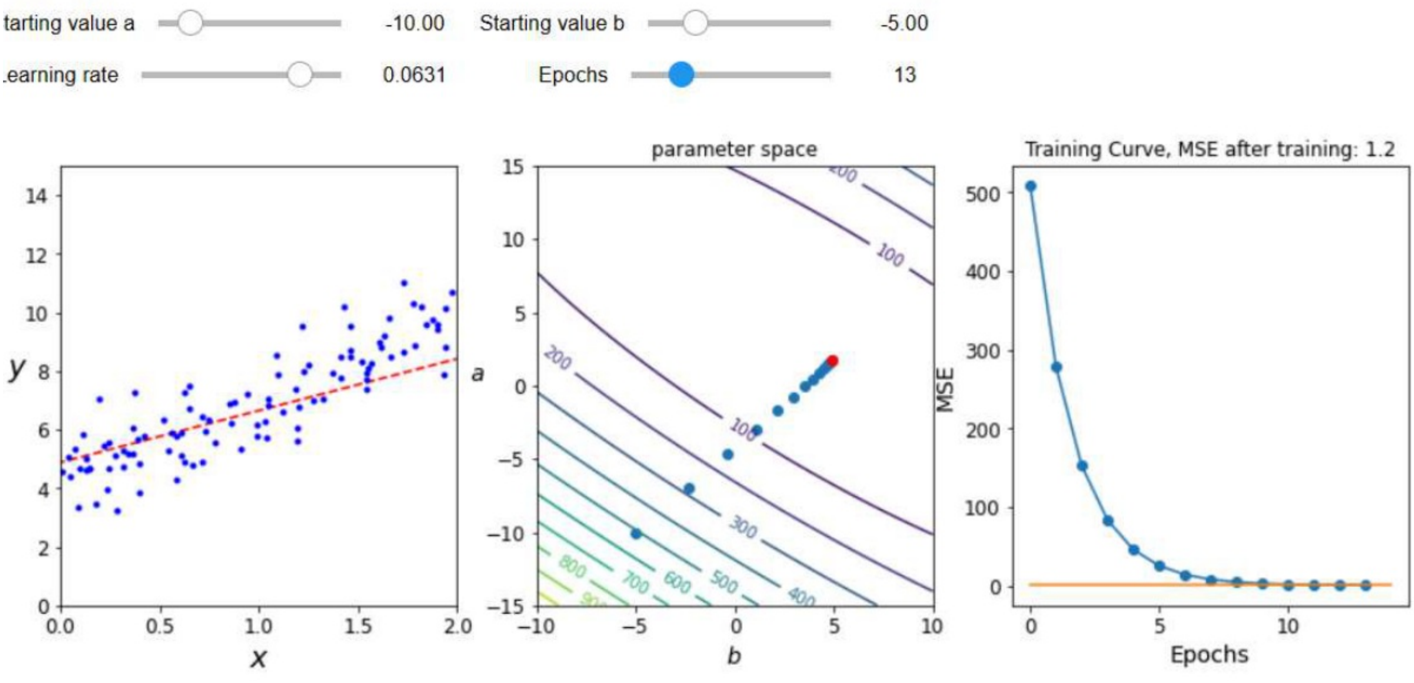


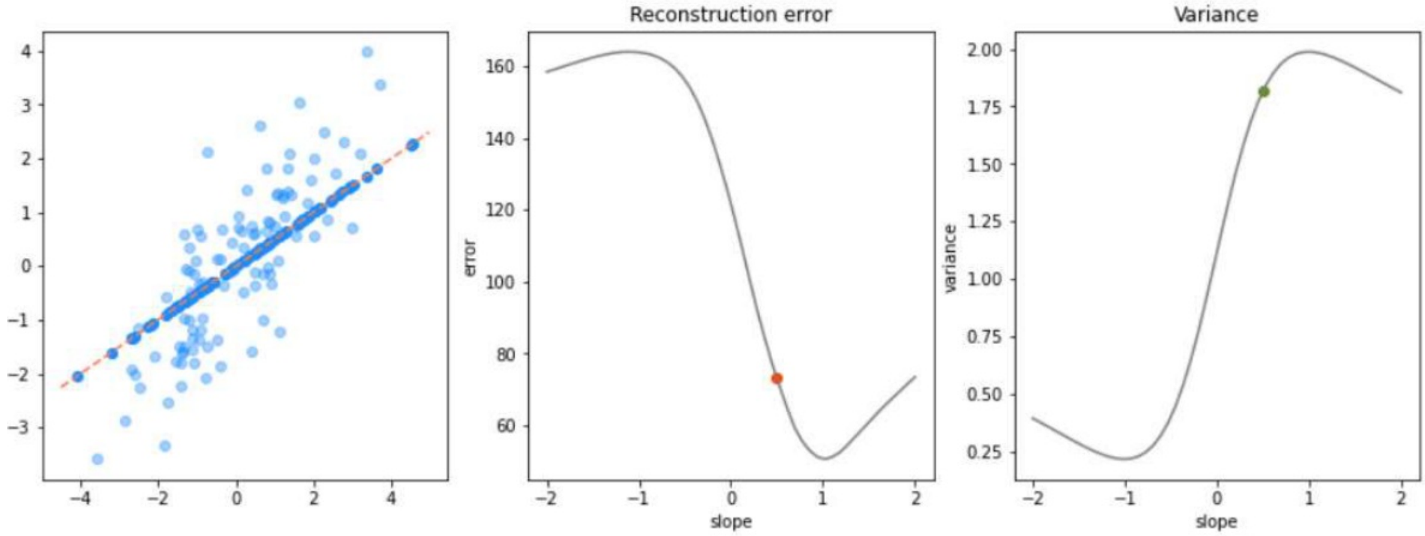
100
80
60
40
20
0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0


Starting value a
-10.00
Starting value b
-5.00
Learning rate
0.0100
Epochs
5
parameter space
Training Curve, MSE after training: 333.
y
x
a
b
MSE
Epochs

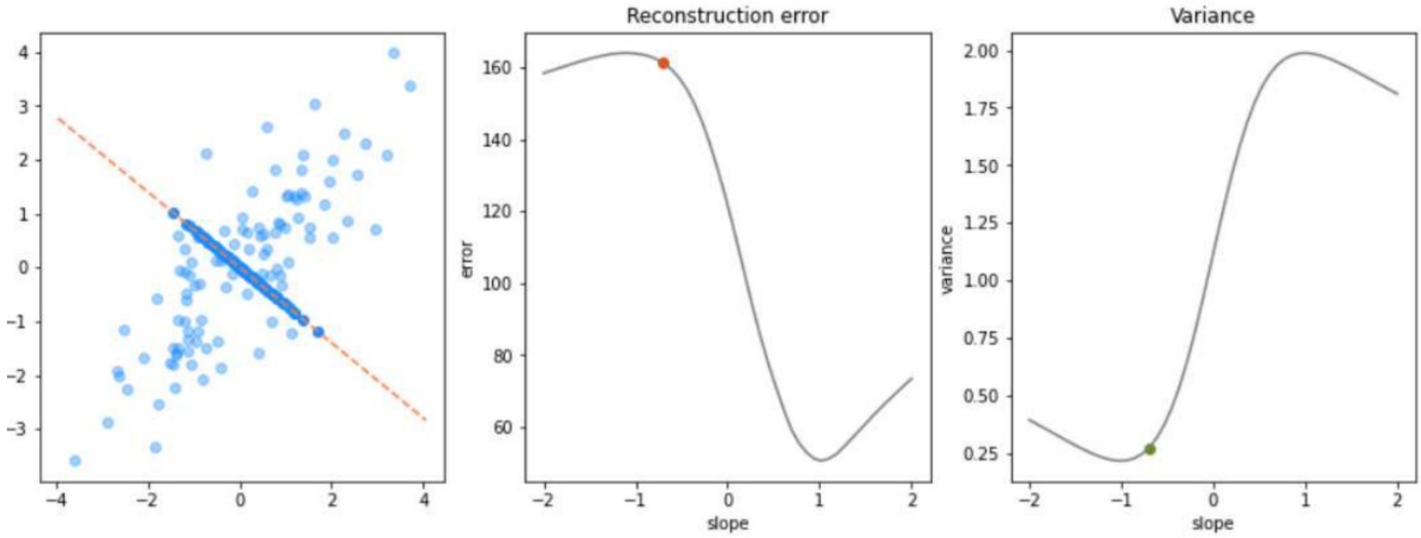

Starting value a
-10.00
Starting value b
-5.00
Learning rate
0.0631
Epochs
5
parameter space
Training Curve, MSE after training: 25.9
y
a
MSE



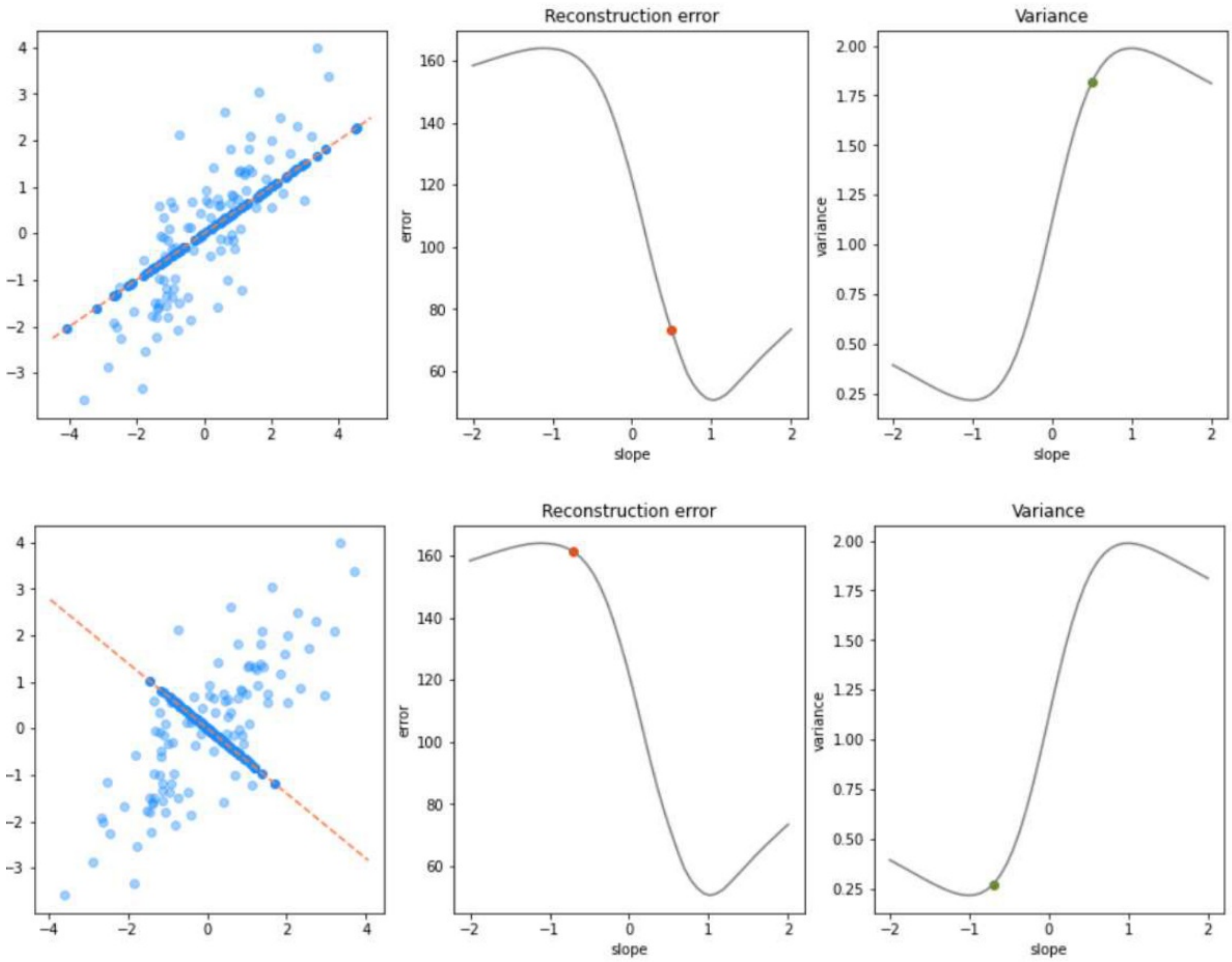

tarting value a
-10.00
Starting value b
-5.00
.earning rate
0.0631
Epochs
13
parameter space
Training Curve, MSE after training: 1.2
y
x
a
b
MSE
Epochs

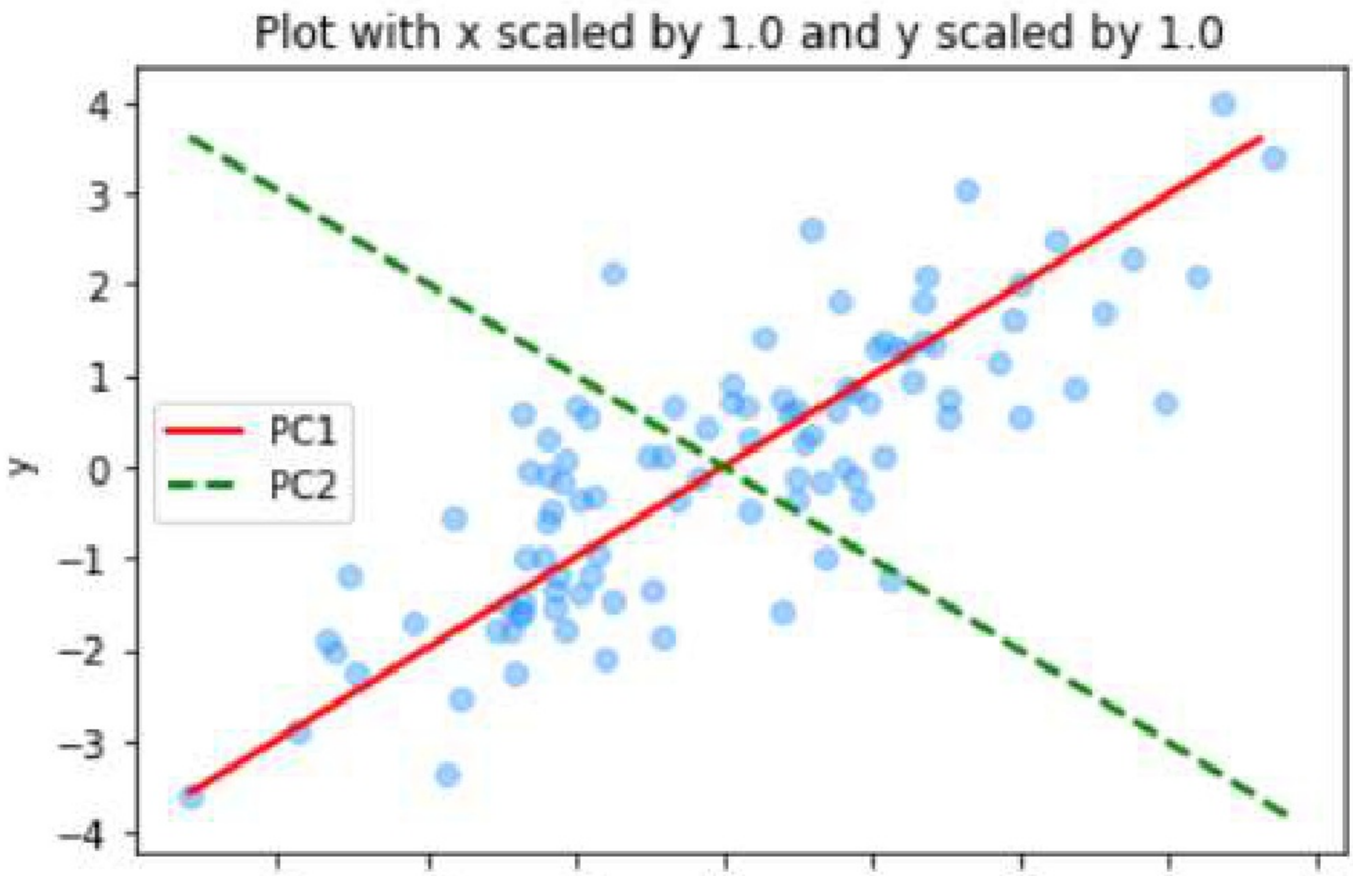


Reconstruction error
error
slope
Variance
variance
slope

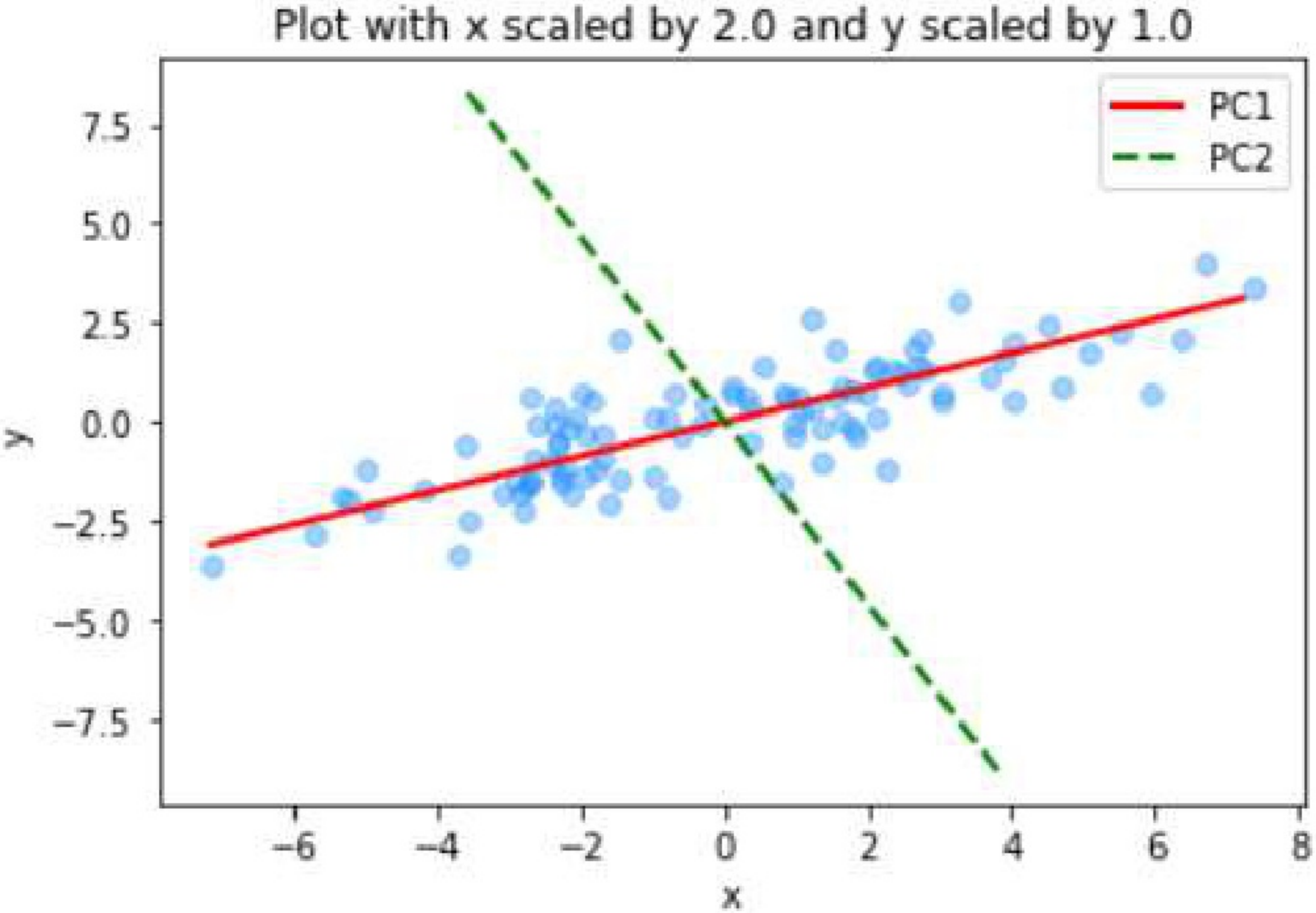


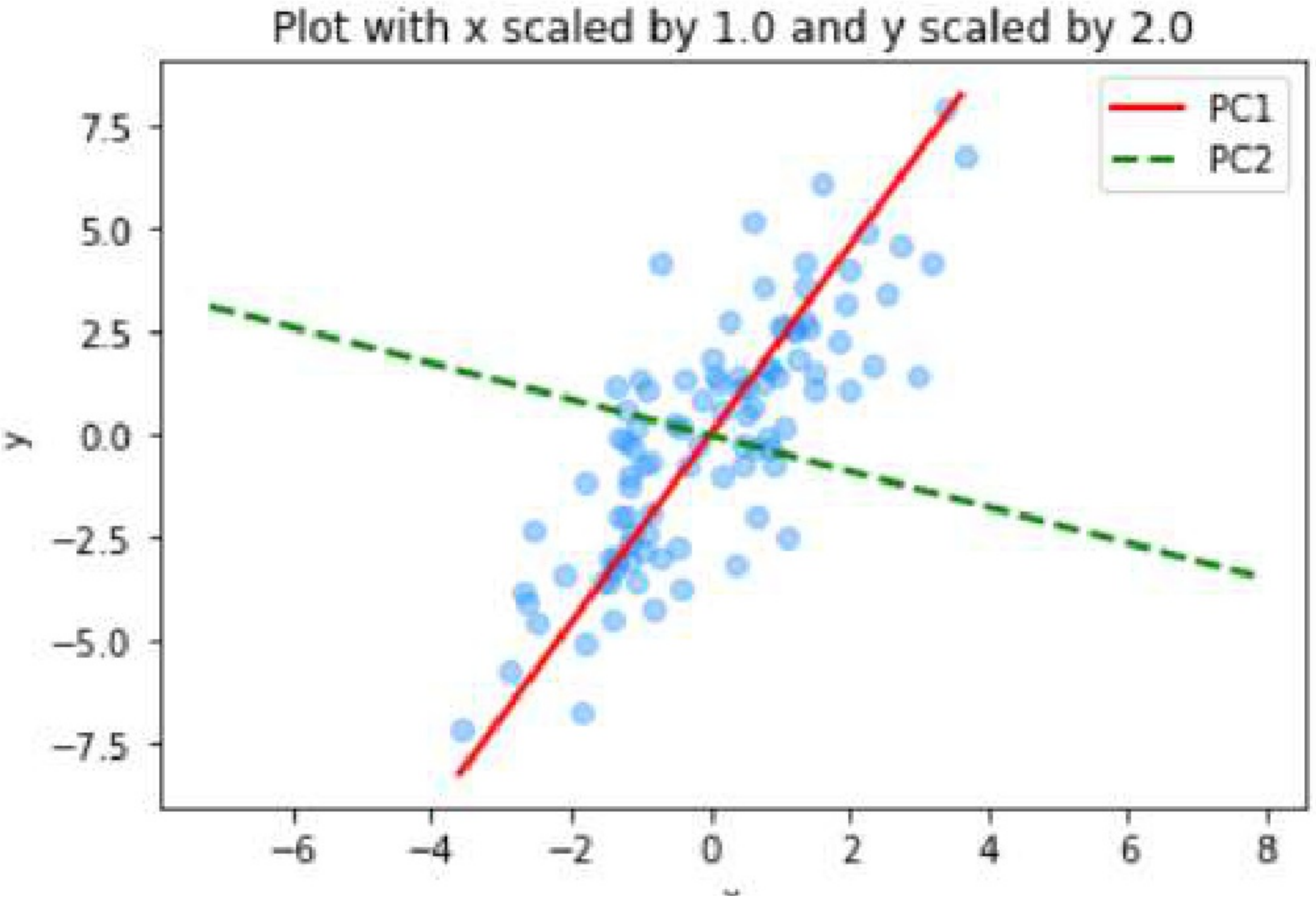
Reconstruction error
error
slope
Variance
variance
slope

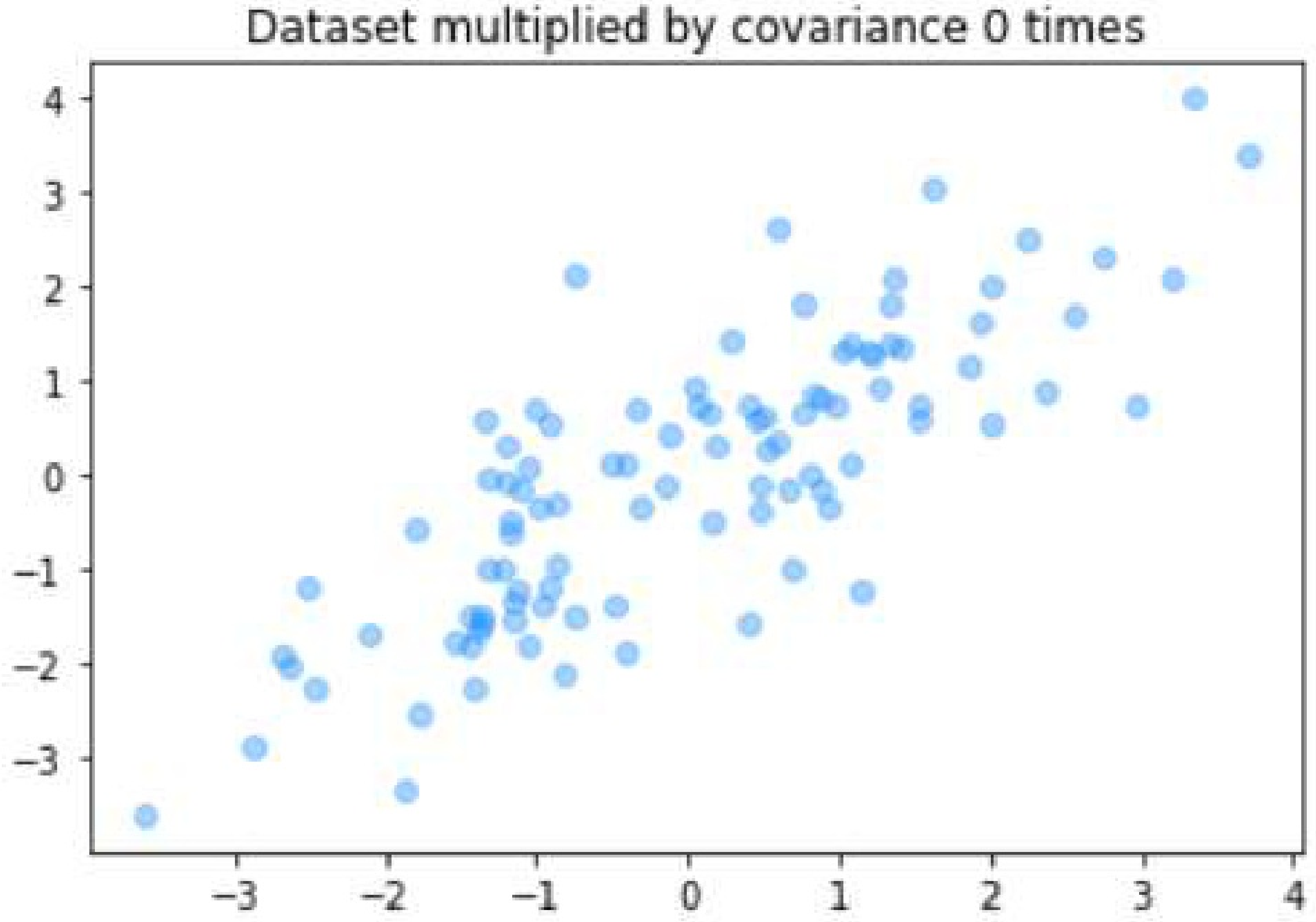

Reconstruction error
Variance
error
variance
slope
Reconstruction error
Variance
error
variance
slope

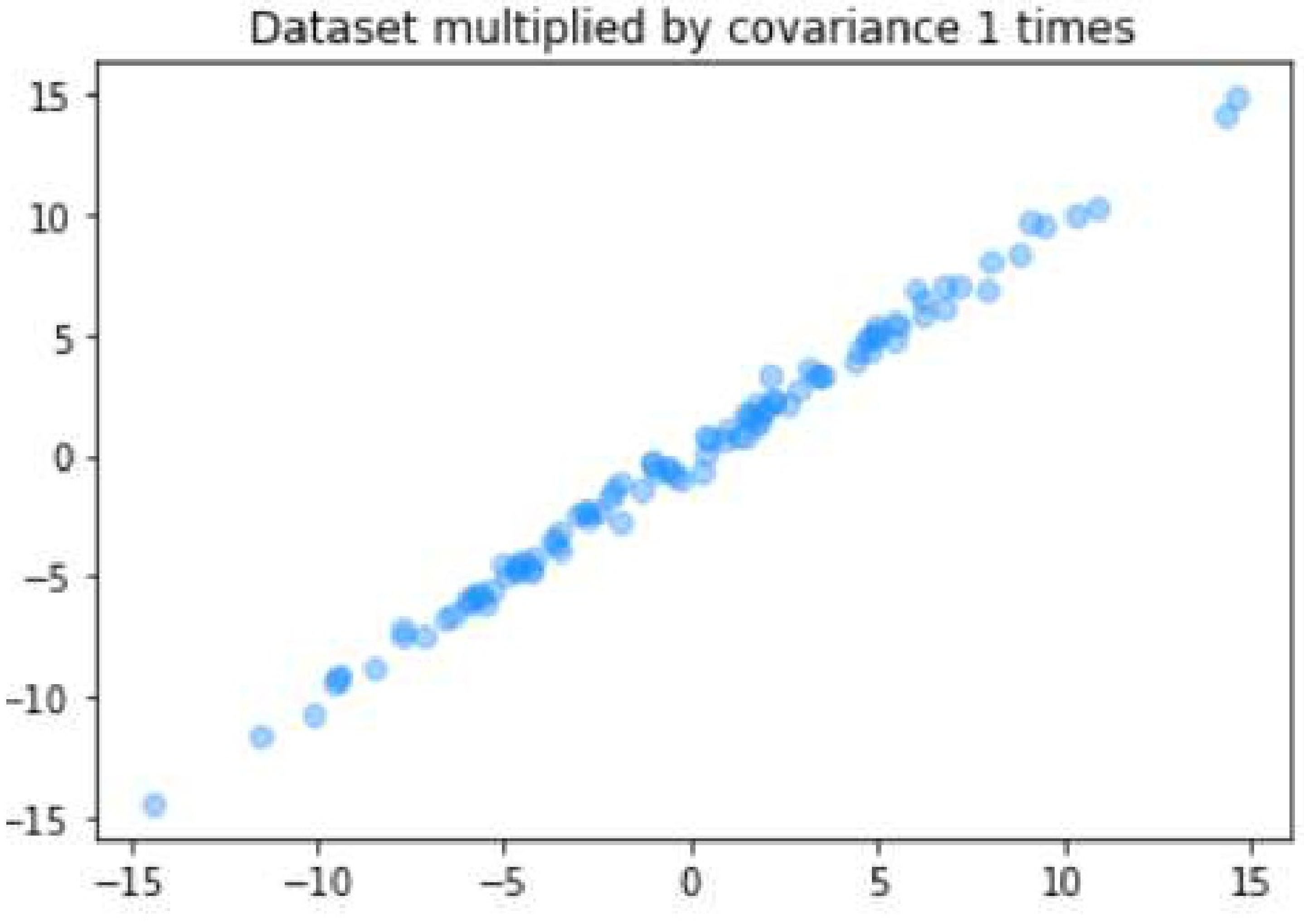

Plot with x scaled by 1.0 and y scaled by 1.0
y
PC1
PC2

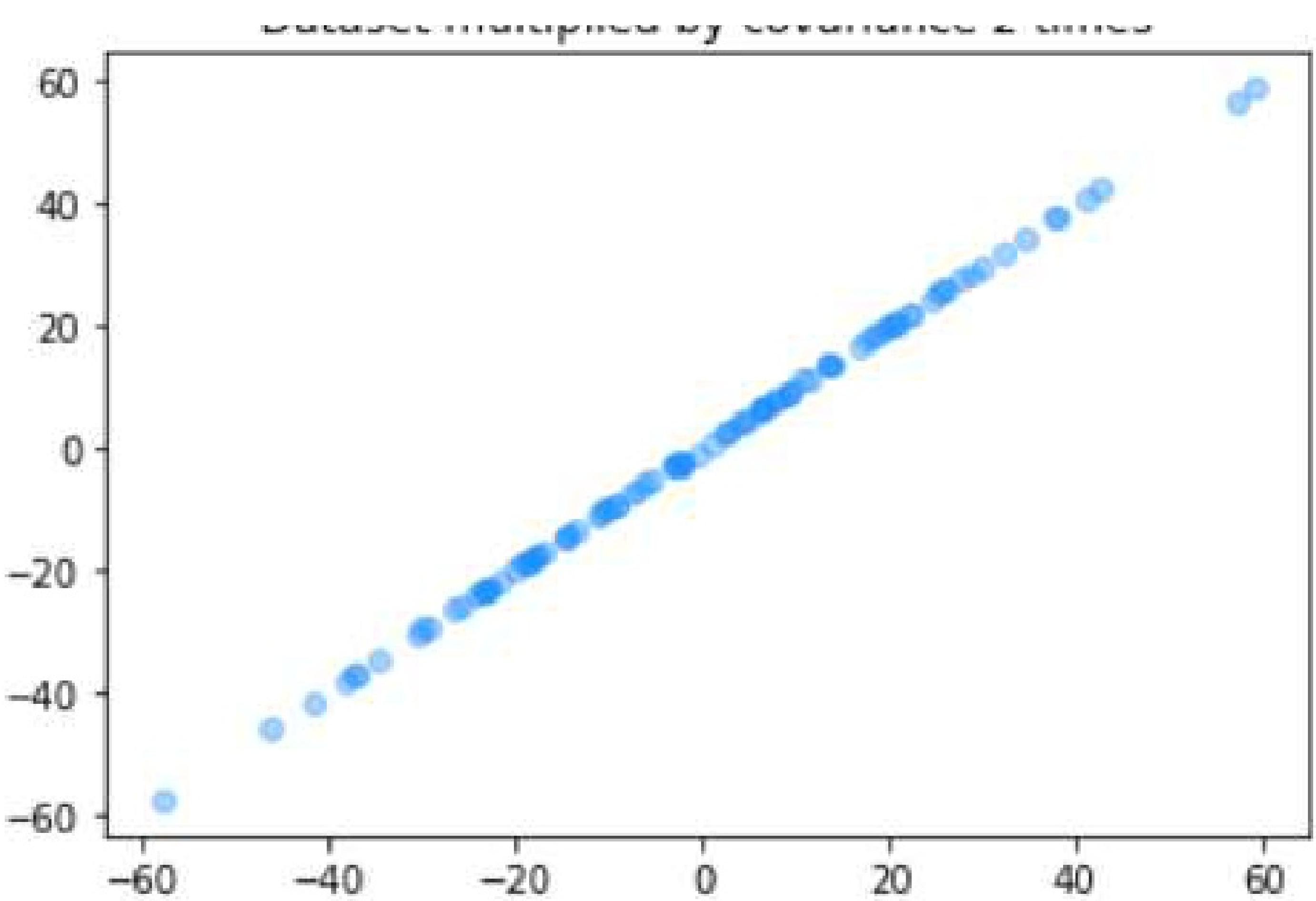

−3 −2 −1 0 1 2 3 4
x


Plot with x scaled by 2.0 and y scaled by 1.0
PC1
PC2
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
y
−6 −4 −2 0 2 4 6 8
x


Plot with x scaled by 1.0 and y scaled by 2.0
PC1
PC2
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
y
−6 −4 −2 0 2 4 6 8


Dataset multiplied by covariance 0 times


Dataset multiplied by covariance 1 times



60
40
20
0
−20
−40
−60
−60
−40
−20
0
20
40
60