

Coding Assignment 1 - November 12, 2024

Student: Ayrton Chilibeck, achilibe@ualberta.ca

Lecturer: Lili Mou, UoA.F24.466566@gmail.com

Problem 1: Least Squares Regression

1. [20% for correct implementation of linear regression] With settings $M = 5000$, $var1 = 1$, $var2 = 0.3$, $degree = 45$, report the weight and bias for $x2y$ and $y2x$ regressions.

Note: getting the above numbers correctly is a necessary (not sufficient) condition for correct implementation.

2. [10%] Three plots of regression models in a row, each with $var2 \in \{0.1, 0.3, 0.8\}$, respectively. (Other settings remain intact: $M = 5000$, $var1 = 1$, $degree = 45$).
3. [5%] A description of the phenomena found in 1) and 2).
4. [15%] We now set $var2 = 0.1$, but experiment with different rotation degrees. The student should design a controlled experimental protocol, plot three figures in a row with different settings, and describe the findings.

1. Result for regression:

0.10650129332985178	-0.005426119116738379
0.10381919022069994	-0.0026061031501614703

2. Plots of the regression model:

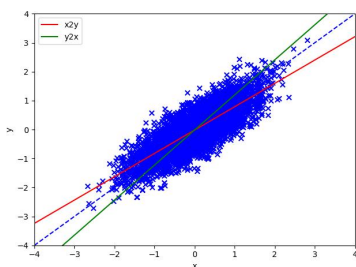


Figure 1: $var2 = 0.1$

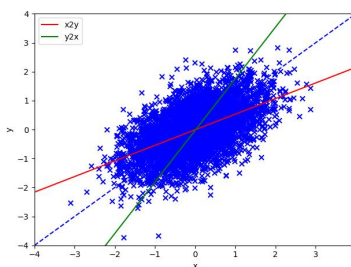


Figure 2: $var2 = 0.3$

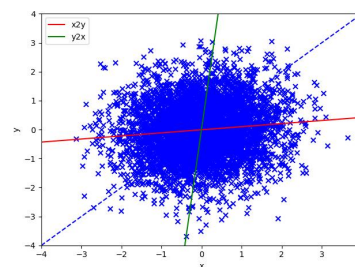


Figure 3: $var2 = 0.8$

3. As we increase $var2$ we see the regression of $x2y$ and $y2x$ diverge.
4. As we increase the rotation of the system, we see the $x2y$ regression diverge from x_{new} , y_{new} and $y2x$ converge to it as we approach a rotation of 90° .

Problem 2: Gradient Descent

In this problem, we will explore gradient descent optimization for linear regression, applied to the Boston house price prediction.

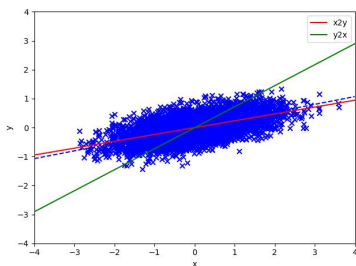


Figure 4: $\text{var2} = 0.1$,
 $\text{rot}=15$ deg

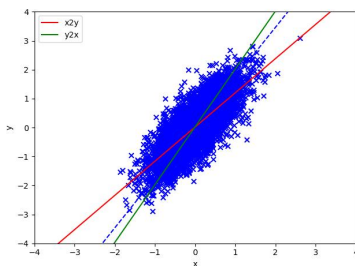


Figure 5: $\text{var2} = 0.1$,
 $\text{rot}=60$ deg

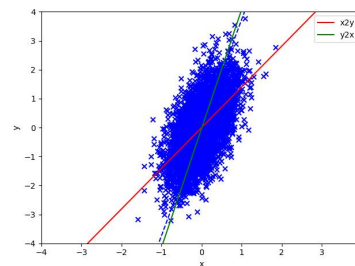


Figure 6: $\text{var2} = 0.1$,
 $\text{rot}=75$ deg

1. [30%] We implement the train-validation-test framework, where we train the model by mini-batch gradient descent, and validate model performance after each epoch. After reaching the maximum number of iterations, we pick the epoch that yields the best validation performance (the lowest risk), and test the model on the test set.

Without changing default hyperparameters, we report three numbers

- (a) The number of epoch that yields the best validation performance,
- (b) The validation performance (risk) in that epoch, and
- (c) The test performance (risk) in that epoch.

and two plots:

- (a) The learning curve of the training loss, and
- (b) The learning curve of the validation risk.
- (c) where x-axis is the number of epochs, and y-axis is training loss and validation risk, respectively.

2. [10%] We now explore non-linear features in the linear regression model. In particular, we adopt point-wise quadratic features. Suppose the original features are x . We now extend it as $[x, x^2]$.

At the same time, we tune λ -penalty to prevent overfitting by

The hyperparameter should be tuned from the set $\{3, 1, 0.3, 0.1, 0.03, 0.01\}$.

Report the best hyperparameter λ , i.e., the one yields the best performance, and under this hyperparameter, the three numbers and two plots required in Problem 2(a).

3. [10%] Ask a meaningful scientific question on this task by yourself, design an experimental protocol, report experimental results, and draw a conclusion.

1. Linear regression by Gradient Descent:

- (a) Best epoch: 99
- (b) Best epoch validation risk: 0.36548514
- (c) Test risk: 0.35231195

2. Linear regression with quadratic parameters, we found that the best values were

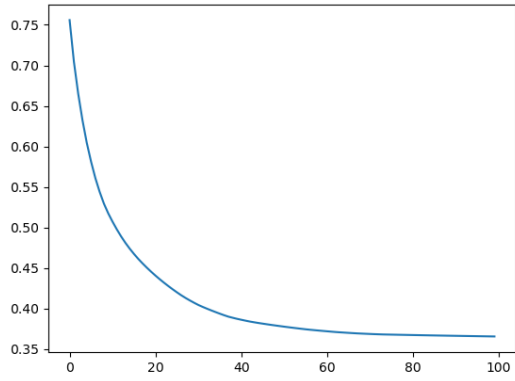


Figure 7: Validation loss

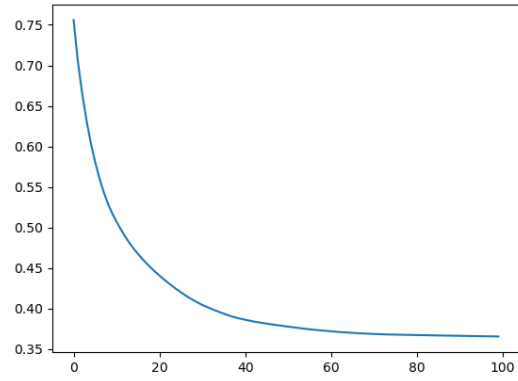


Figure 8: Validation risk

- (a) Best epoch: 99
- (b) Best epoch validation risk: 0.3358987
- (c) Test risk: 0.38372331

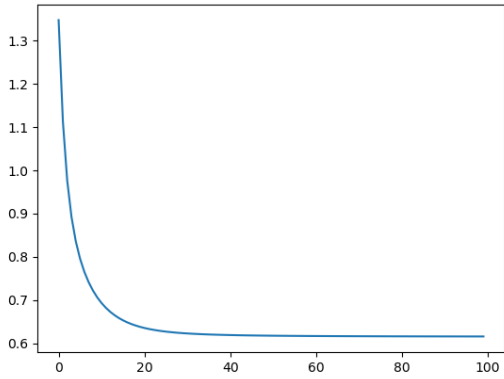


Figure 9: Validation loss

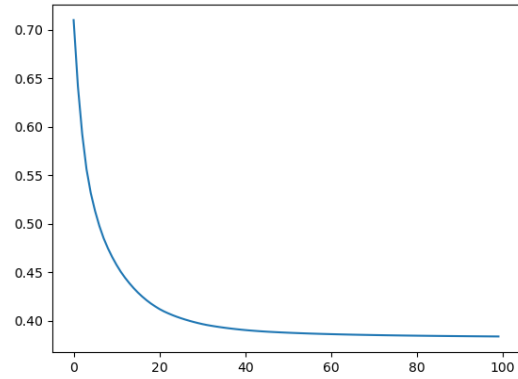


Figure 10: Validation risk

3. My research question would be does this method converge with the quadratic parameters, and does the test loss increase relative to the simple linear regression with more epochs. I found that the system continues to converge past the 100 epochs, changing up to 1000 epochs and has lower test accuracy than the linear regression. This means our quadratic augmentation overfits the data, so our simple linear regression is the more descriptive model.