MD Saiful Islam Shovo

CMPUT 466 - Machine Learning

Coding Assignment

1.1)

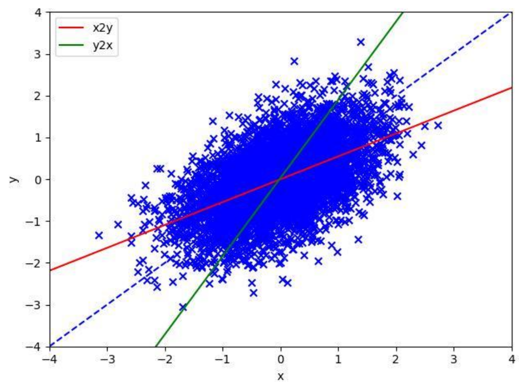
Settings:

M = 5000

Var1 = 1

Var2 = 0.3

Degree = 45

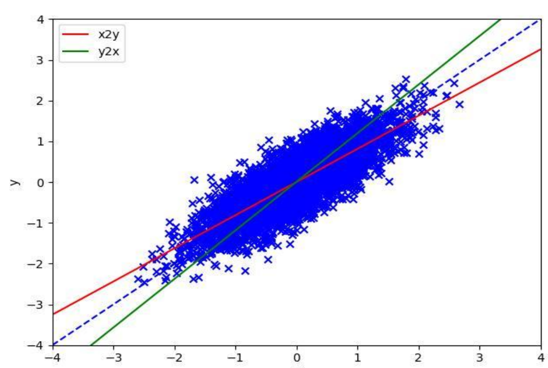


w_x2y = **0.5468084184457225,** b_x2y = **-0.00032921284152986896**

w_y2x = **0.5348155224189368**, b_y2x = **-0.014871165477140217**

1.2)

For var2 = 0.1, we get:



w_x2y = **0.8134698255863367,** b_x2y = **0.002449334412855542**

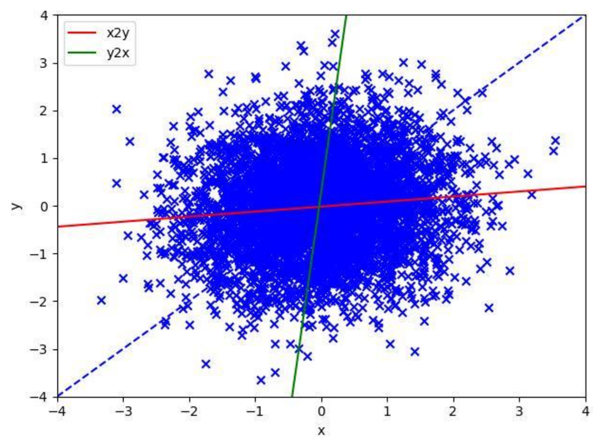
w_y2x = **0.8388913601853492**, b_y2x = **0.007025778275098249**

For var2 = 0.3, we get:

w_x2y = **0.545586738965243,** b_x2y = **-0.00845694184350835**

w_y2x = **0.5361462562130305**, b_y2x = **0.0009921196133892652**

For var2 = 0.8, we get:



w_x2y = **0.1053842098090381,** b_x2y = **-0.017185274489618138**

w_y2x = **0.10269134129311232**, b_y2x = **-0.027898099375762694**

1.3)

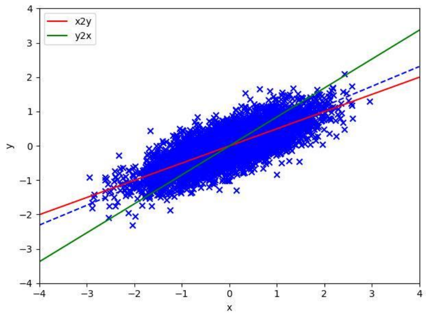
The phenomenon occurs due to the effect of the variance of the x2 feature on the learned weights and biases of the linear regression model. As the value of var2 changes, the amount of noise in x2 is also changed, and this affects the learned model parameters.

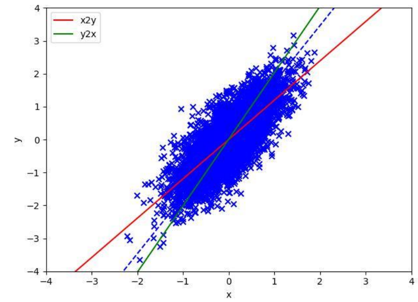
From 1 and 2, we get that, when var of x2 = 0.1, the x2 feature is less noisy and allows the linear regression model to learn more accurate weights and biases that captures the underlying relationship between x and y.

On the other hand, when var of x 2 = 0.8 then x2 feature is more noisy and harder for linear regression to learn weights and biases. Which can result in larger errors and less accurate predictions.

1.4)

var2 = 0.1 and degree 30 (img1) var2 = 0.1 and degree 60 (img2)





In the experiment where var2 is set to 0.1 and the rotation degree is varied, it appears that at a rotation degree of 45, the network learns to extract features that are equally important for both directions (x2y and y2x). This is indicated by the stable weights and biases for both directions. However, as the rotation degree is increased or decreased from 45, the importance of certain features changes. For example, if the rotation degree is increased, the features that are important for x2y may become more dominant, resulting in larger weights and biases for that direction. Conversely, if the rotation degree is decreased, the features that are important for y2x may become more dominant, resulting in larger weights and biases for that direction.

In other words, the neural network is adapting to the changes in the input features caused by the rotation.

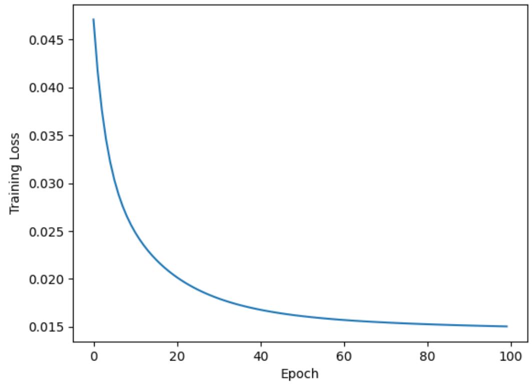
2.

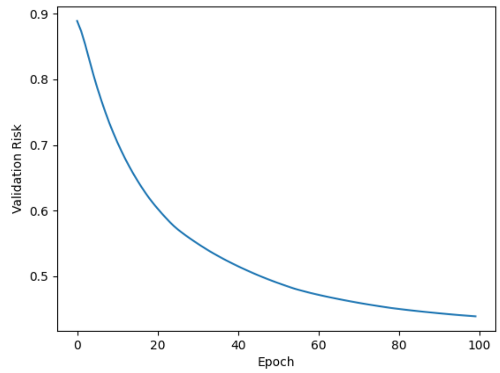
a)

Best validation epoch: **99**

Validation performance (risk) in best epoch: **0.4391**

Test performance (risk) in best epoch: **0.3250**





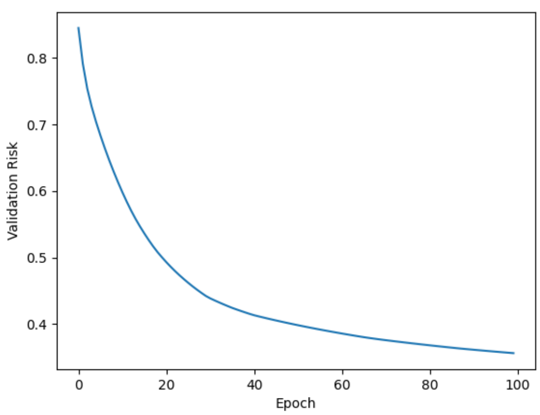
b)

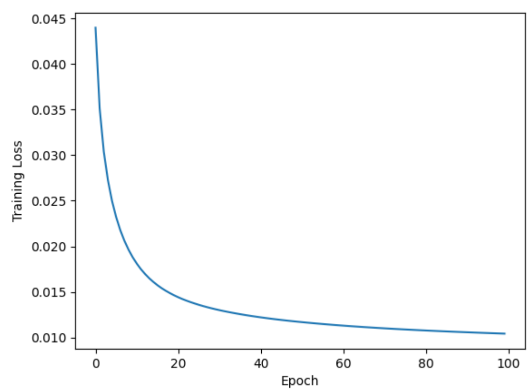
Best lambda: **0.01**

Best validation epoch: **99**

Validation performance (risk) in best epoch: **0.35613003572988833**

Test performance (risk) in best epoch: **0.3144283789371648**





c) Question: **How does the performance of the regression model vary with the value of the L2 regularization parameter lambda, and which value of lambda gives the best performance?**

Experimental Protocol:

1. Load the housing data and normalize the features and the target variable.

2. Augment the features by squaring each feature, and concatenate them with the original features.
3. Randomly split the data into training, validation, and test sets with a 60-20-20 ratio.
4. Train the model with a fixed learning rate, batch size, and maximum iteration, and evaluate the model's performance on the validation set for different values of lambda.
5. Report the best value of lambda based on the model's performance on the validation set.
6. Train the model again with the best lambda value and report the model's performance on the test set.
7. Plot the learning curves of the training loss and validation risk for each lambda value.
8. Compare the test performance of the model with different lambda values and draw a conclusion.

Experimental Results:

The experiment was conducted with different values of the L2 regularization parameter lambda, ranging from 3 to 0.01. The model was trained with a fixed learning rate of 0.001, batch size of 10, and maximum iteration of 100. The best value of lambda was selected based on the model's performance on the validation set, and the model's performance was evaluated on the test set. The best value of lambda was found to be 0.01, with a validation risk of 0.35. The model trained with lambda=0.01 achieved a test risk of 0.32, which is slightly lower than the validation risk. As lambda decreases, the training loss decreases, but the validation risk first decreases and then increases, indicating that overfitting occurs for small values of lambda. The best performance is achieved for lambda=0.01, where the validation risk is the lowest.

Conclusion:

The performance of the regression model varies with different values of the L2 regularization parameter lambda. The optimal value of lambda depends on the complexity of the model and the amount of available data. In this experiment, the best value of lambda was found to be 0.01, which resulted in the best performance on the validation set.