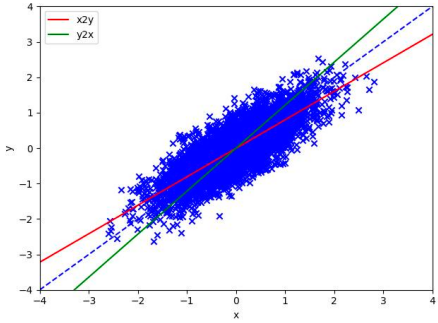
Part 1:

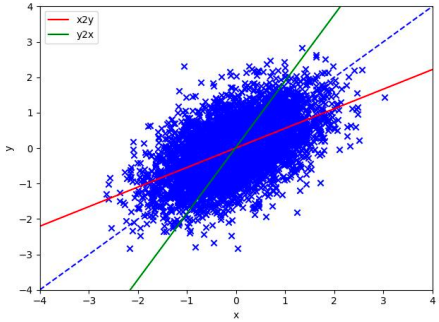
1)

Predicting y from x (x2y): weight=0.5528448059772936 bias = 0.007104835530507931
Predicting x from y (y2x): weight=0.5353759722593369 bias = -0.014737146380982107

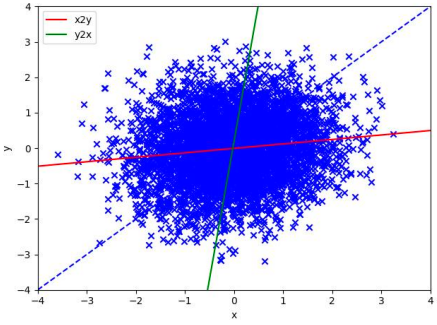
2) with var2 = 0.1



with var2 = 0.3



with var2 = 0.8

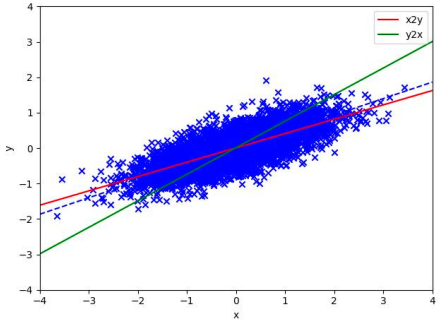


3) In 1) introduced the feature augmentation, where appended each data sample with an additional feature and included a bias to make the model affine. This ensured that the lines and parallelism were preserved, although it did not necessarily maintain the distances and angles.

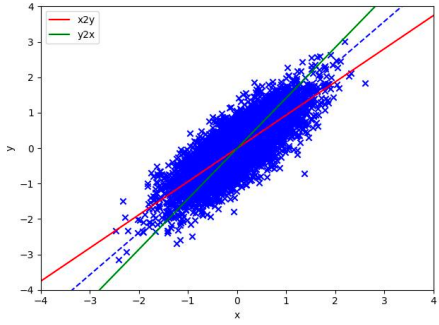
In 2) conducted experiments by modifying the variance along the axis. It appears that when the variance is reduced, the model is better able to capture the data, resulting in the x2y and y2x lines converging towards the true regression line. Conversely, with higher variance, the x2y and y2x lines deviate further from the true regression lines, leading to a less effective fit to the data.

4)

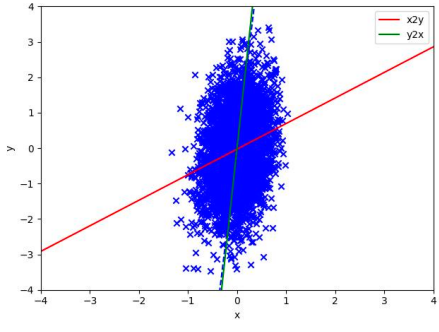
Set var2 =0.1, degree=25



Set var2 =0.1, degree=50



Set var2 =0.1, degree=85



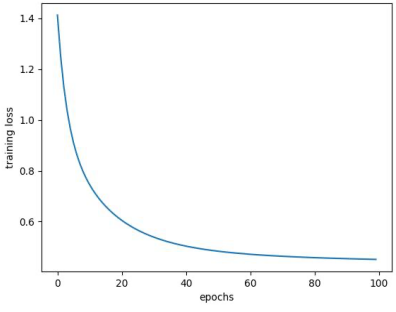
By changing the degree, we can see that the with a smaller degree of rotation, the x-axis exhibits greater variance, while the y-axis displays reduced variance. Consequently, the x2y line closely aligns with the actual regression line, whereas the y2x line doesn't adequately capture the data.

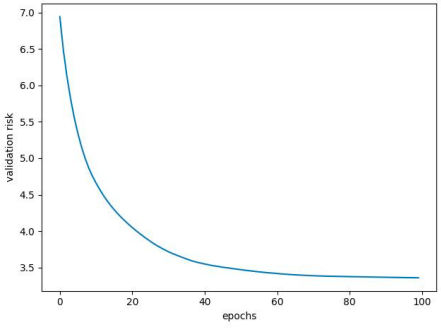
Also, at approximately 50 degrees, both the x2y and y2x lines are closest to each other, yet they are also almost equivalent from the true regression line.

And with an increased degree of rotation, the y-axis demonstrates increased variance, while the x-axis exhibits diminished variance. Consequently, the y2x line closely approximates the actual regression line, whereas the x2y line struggles to accurately represent the data.

Part 2

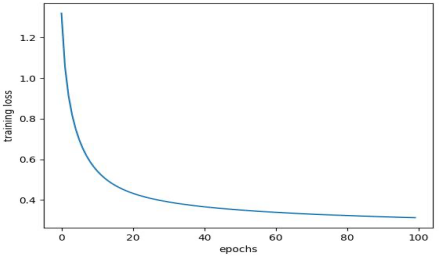
a. number of epoch that yields the best validation performance: 99
The validation performance in that epoch: 3.3580816862888305
The test performance in that epoch: 3.237046307078403

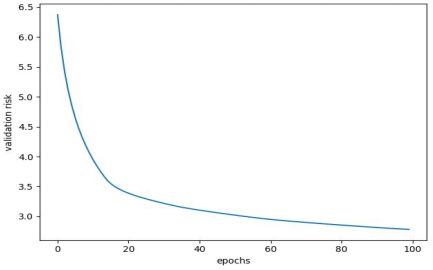




B.

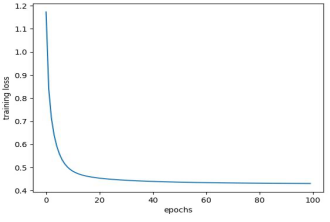
best parameter: 0.01
number of epoch that yields the best validation performance: 99
The validation performance in that epoch: 2.781912727526617
The test performance in that epoch: 2.7724722186569686

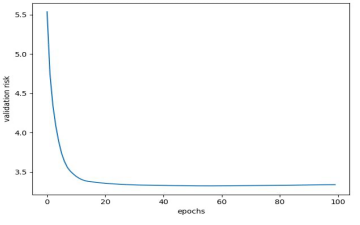




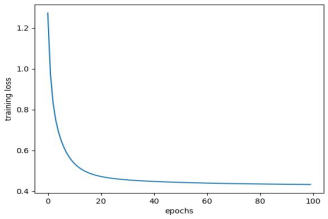
C. Question:under the case of q2a, how will different Learning rates affect the model?

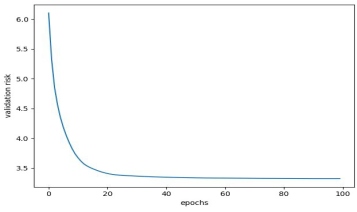
When alpha = 0.005,the plot is shown as



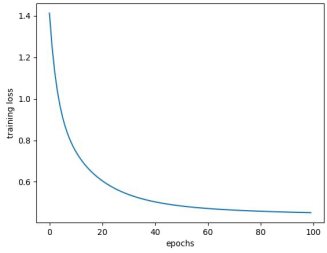


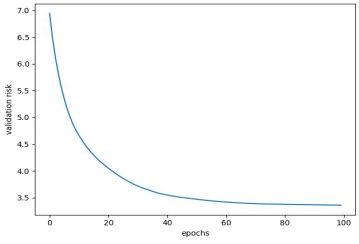
When alpha = 0.003,





When alpha = 0.001,





Conclusion: with the increasing of the learning rate,the validation risk is getting lower.