1 Linear Least Squares Regression

Before getting started implement linear regression code of Diabetes dataset, we plot the pair-wise scatters graph of data and distribution graph of target as show in Fig. 1, Fig. 2, Fig. 3. For overview the relationship of data and how target distribution.The data distribution and the pair-wise scatter graph of data looked a bit strange. In Fig. 1 show y-axis of this graph that is one of the dimension of Diabetes dataset have many same data. It might relate to dimension of Diabetes dataset in x-axis but in this way didn't problem to use this dataset.

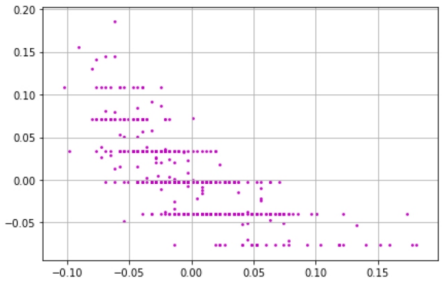

Figure 1: First pair-wise scatters plot of Diabetes Dataset

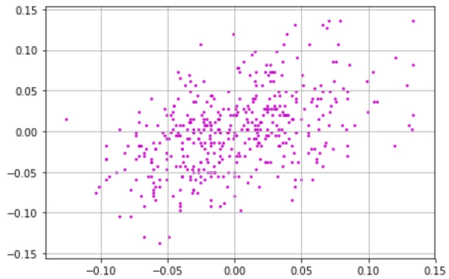

Figure 2: Second pair-wise scatters plot of Diabetes Dataset

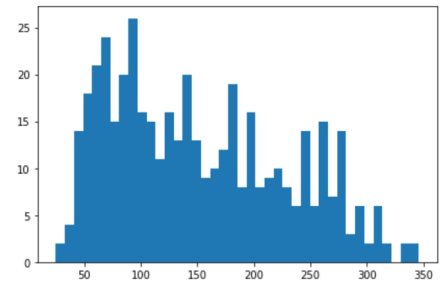

Figure 3: Distribution Histogram of target in Diabetes Dataset

When we implement linear model from sklearn and linear predictor that is solved by the pseudo-inverse method by used the same dataset we will get result as illustrate in Fig. 4 and Fig, 5. This two method were have the same result but it have the y-axis of results didn't the same, since Diabetes dataset have data in range of negative number to positive number it could have the result as the pseudo-inverse method. However, the result from sklearn was different because it have setting to be changed the intercept by default of it parameters.

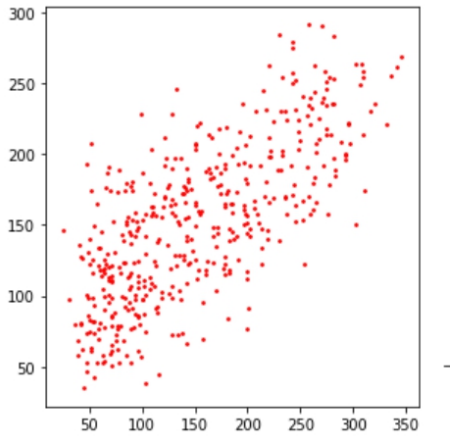

Figure 4: The result of linear model from sklearn

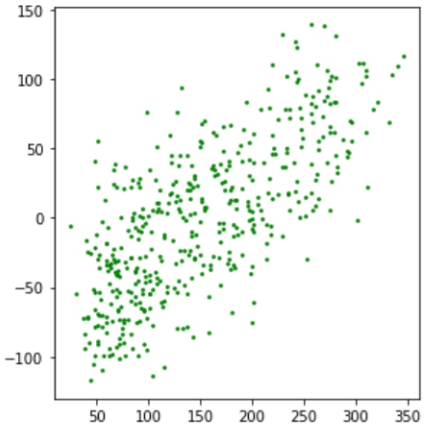

Figure 5: The result of the pseudo-inverse method

2 Regularization

When we compare the bar graph of weight side by side as in Fig. 6. It show overall weight of Tikhonov regularization less than weight of The pseudo-inverse method because it add regularization coefficient diagonal matrix into normal equations for the least squares problem.

$w = (\Phi^T\Phi)^{-1}\Phi^T t$

(Normal equations for the least squares problem)

$w = (\lambda I + \Phi^T\Phi)^{-1}\Phi^T t$

(Simple extension of the least-squares solution)

These two equation from [3].

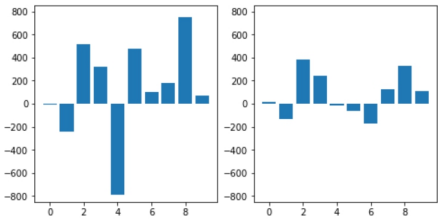

Figure 6: Bar graphs of the weights (Left: The pseudo-inverse method, Right: Tikhonov regularization)

3 Sparse Regression

Your comparisons should look similar to Fig. 1 in laboratory instruction. In each of these cases, compare the prediction errors. In the case of the sparse regression, would you say the features with nonzero weights are more meaningful (to answer, you have to find the source of the data and look at the variables)?

• Once we compare the three bar graph of weight in Fig. 7, the weight for sparse Regression have some of weight disappear. When we using sparse Regression to prediction data it might be have error more than the pseudo-inverse method and Tikhonov regularization. Moreover, the reason that some of weight in sparse regression disappear because penalty function of spare regression didn't squared, that if λ is sufficiently large it make some of the coefficients w_j are come to zero and some of weight of sparse regression that present in Fig. 7 become zero. Nonzero weight is meaningful because when we use it in equation it will not make some of the result disappear or equal to zero.

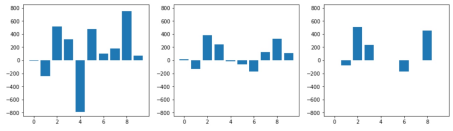

Figure 7: Bar graphs of the weights (Left: The pseudo-inverse method, Center: Tikhonov regularization, Right: Sparse Regression(lasso))

4 Regularization Path

We can give many different coefficient at the initial of the regression and regularization until it zero as in Fig. 8 to select the best fit model. In Fig. 9 present regularization path in the other way.

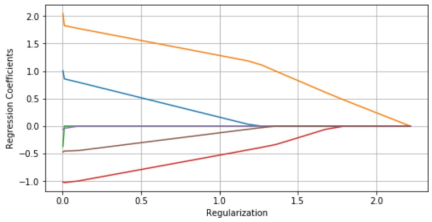

Figure 8: regularization path graph 1

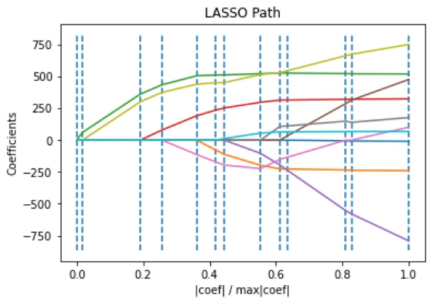

Figure 9: regularization path graph 2

5 Solubility Prediction

First we plot distribution histogram of target in Husskonen Solubility Features dataset to overview the data as in Fig. 10.

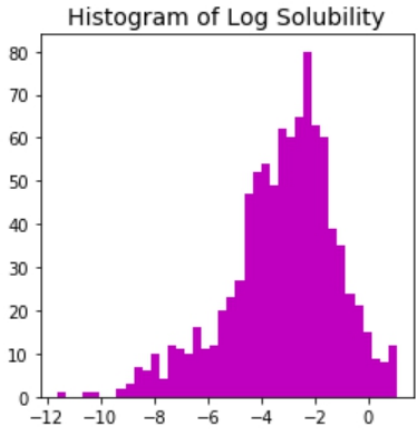

Figure 10: Distribution Histogram of target in Husskonen Solubility Features Dataset

If you were to select the top ten features to predict solubility, what would they be? How good is the prediction accuracy with these slected features when compared to using all the features and a quadratic regularizer?

• First we need test and train for top ten features and using all features to fit the model before use it with dataset, some of the result illustrate in Fig. 11 and Fig. 12 are the result of train by used all features and test by used all features, respectively.

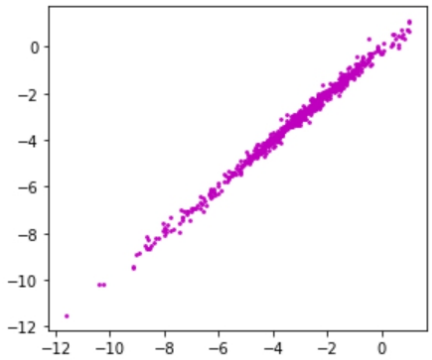

Figure 11: The result of trained by used all features(Train dataset from [1])

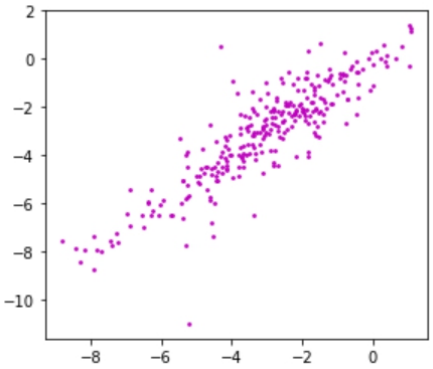

Figure 12: The result of tested by used all features(Test dataset from [1])

Then, we test top ten features and using all features, the result as show in Fig. 14 and Fig. 15. We will see the result of top ten features dataset it look not be similar to a Straight line(as in Fig. 13 than the result of all features dataset. Top ten features dataset look seem overfit than the result of using all features dataset that might because top ten features is have huge data more than dimension than have only 10 dimension.

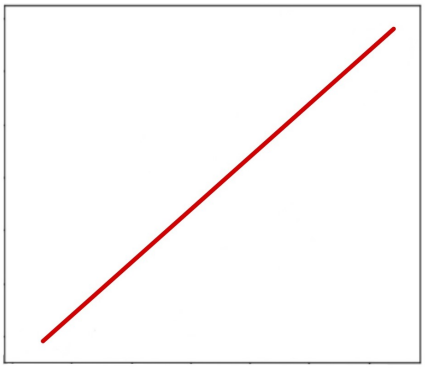

Figure 13: Straight line

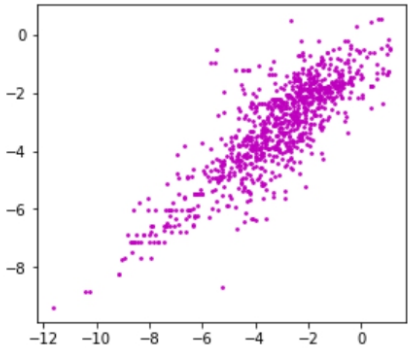

Figure 14: The result of tested by used 10 features(Dataset from [1])

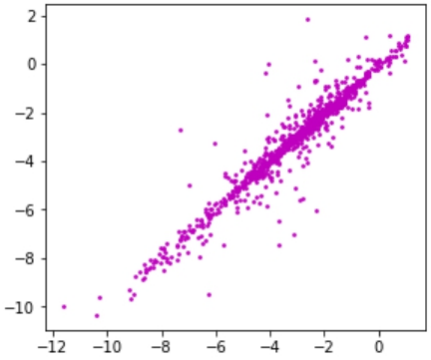

Figure 15: The result of tested by used all features(Dataset from [1])

Are you able to make any comment comparing your results to those claimed in [1] or [2]?

• When we observe Fig. 14-16, we can see the result from [2] look overfit more than the result from [1] because it have data more than [1] that make the point in the scatter plot more density.

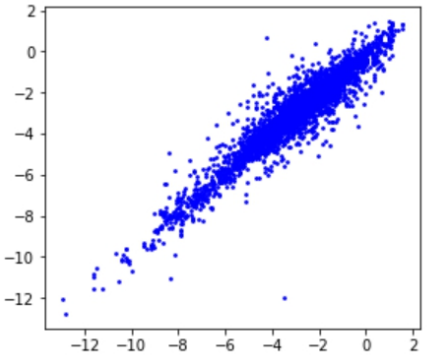

Figure 16: The result of tested by used 10 features(Dataset from [2])

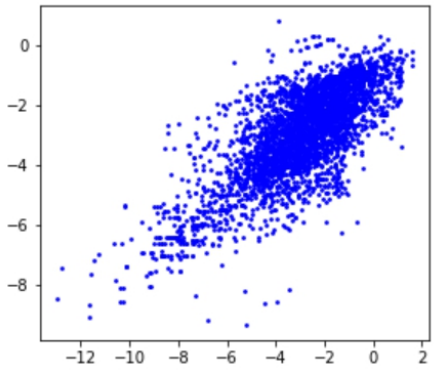

Figure 17: The result of tested by used all features(Dataset from [2])

Reference

[1] J. Huuskonen, M. Salo, and J. Taskinen, "Aqueous solubility prediction of drugs based on molecular topology and neural network modeling," Journal of Chemical Information and Computer Sciences, vol. 38, no. 3, pp. 450-456, 1998.
[2] M. Pirashvili, L. Steinberg, B. G. F., M. Niranjan, J. G. Frey, and J. Brodzki, "Im- proved understanding of aqueous solubility modeling through topological data analysis," Journal of Cheminformatics, vol. 10, no. 1, p. 54, 2018.
[3] Bishop, C.M. (2016). Pattern Recognition And Machine Learning. Springe. pp. 137-146.