Assignment Report

Homework 1: Regression

Yuanzhe Dong & Zhuo Ouyang (Group 2)

Machine Learning: From Math to Code

July 9, 2023

PEKING UNIVERSITY 1898
北京大學
PEKING UNIVERSITY

Please note that we have submitted both .py and .m codes because the two of us choose different programming languages. The corresponding results and figures are mainly based on Python programs, which are developed independently from the example Matlab codes provided by the professor.

Problem 5.3

The data contained in the datasets has two columns representing data vector x and target vector y. According to the problem, we need to calculate ρ which depicts the correlation between x and y, and then to apply linear regression to the dataset, thus we can obtain $\hat{y}$ and $\bar{y}$ and get the TSS, ESS, RSS as well as R^2. The key is to use the pseudo-inverse of the augmented matrix $\mathbf{X}$ to determine the regression parameter w

$$w = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}y$$

As for the first task, we can use the covariance matrix to obtain the variance of x and y (which are the diagonal elements of the matrix) and the covariance of the two vectors (which are the non-diagonal elements). Then ρ can be calculated by the formula:

$$\rho = \frac{\sigma_{xy}^2}{\sqrt{\sigma_x^2\sigma_y^2}}$$

We plot the regression curve as below.

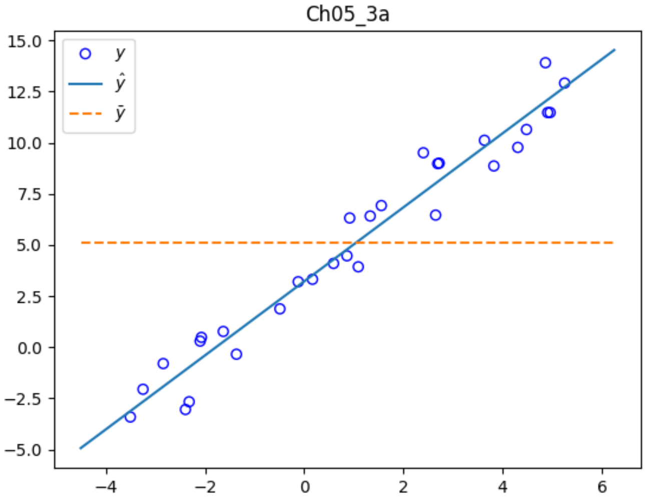

Figure 1: Linear regression for 2D dataset

The data used in the second task differs from that of the first in the dimension, which has a 3D dataset containing x_1, x_2 and the target y. But the method of solving the regression problem is the same, just more dimensions of matrix calculation. The surface derived is shown below.

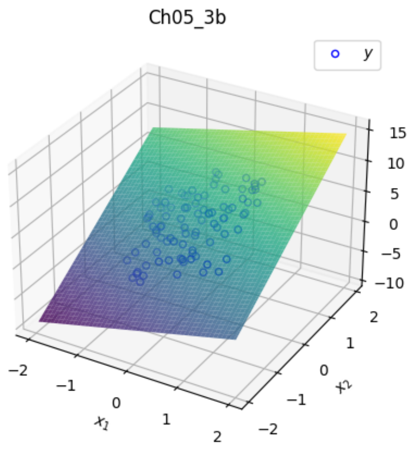

Figure 2: Linear regression for 3D dataset

Problem 5.4

In solving this problem, first we generate a set of highly correlated 2D data points, with x_2 equal to x_1 plus a certain level of noise. Thus the augmented matrix $\mathbf{X}$ is very close to singular, creating an ill-conditioned problem fit for ridge regression.

To generalize our conclusion, we choose different noise levels a, and under each level we randomly generate different x_2 based on the same x_1. In ridge regression, a modified version of the pseudo-inverse is used:

$$w = (\mathbf{X}\mathbf{X}^T + \lambda\mathbf{I})^{-1}\mathbf{X}y$$

We expect this method is more robust against noise, although with a little sacrifice of accuracy. The tradeoff largely depends on the hyperparameter λ.

For each dataset generated, we apply ridge regression with different λ, including $\lambda = 0$, which is actually the vanilla version of linear regression, and calculate the $R^2 = 1 - \mathrm{RSS}/\mathrm{TSS}$ value to assess the model. The correlation coefficient ρ between x_1 and x_2 along with the eigenvalues of $\mathbf{X}$ is also derived in order to measure how ill-behaved the data is.

noise level a	ρ	smallest eigenvalue of $\mathbf{X}$
0.01	0.999991	4.0×10^{-5}
	0.999993	3.3×10^{-5}
	0.999995	2.5×10^{-5}
0.1	0.9991	4.2×10^{-3}
	0.9993	3.0×10^{-3}
	0.9988	5.6×10^{-3}
1	0.986	0.07
	0.972	0.15
	0.966	0.18

Table 1: Parameters measuring the ill behaviour

The results under noise level $a = 0.01, 0.1, 1$ are shown as below. The left, middle and right panels correspond to the differently perturbed datasets as indicated in Table 1, and the 3 rows from top to bottom represent different models with the weight decay parameters $\lambda = 0, 0.01, 0.1$. The R^2 value under each circumstance is shown, as well as the vector $\mathbf{w}$.

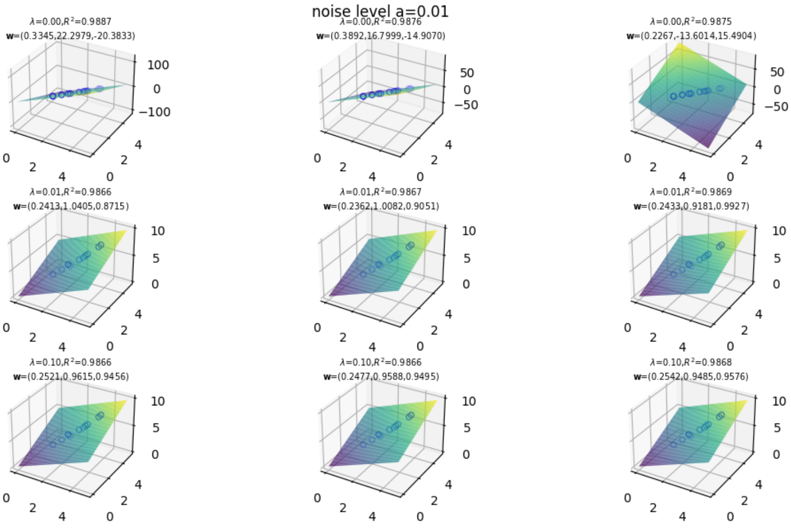

Figure 3: Ridge regression with noise level $a = 0.01$

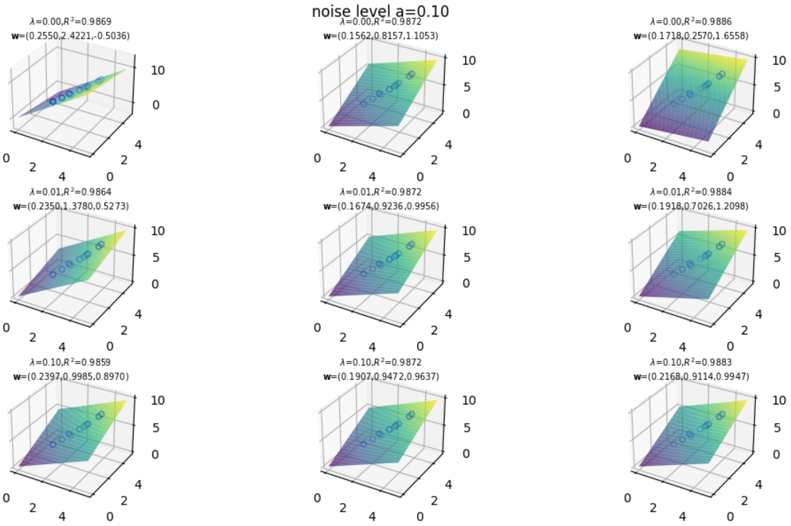

Figure 4: Ridge regression with noise level $a = 0.1$

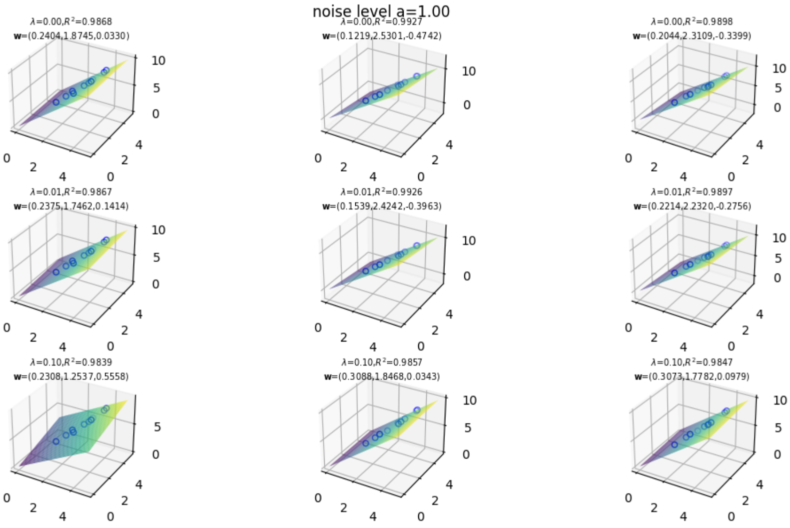

Figure 5: Ridge regression with noise level $a = 1$

We can see from the figures that in such a ill-conditioned problem, linear regression (top row) becomes prone to perturbation. With λ introduced, similar regression results are given each time, meaning the robustness is improved, but the R^2 value is slightly lower as the accuracy drops. Choosing the proper value of λ might be tricky, but it can make the model more suitable.

Note that the conclusion TSS=ESS+RSS does not hold for ridge regression, and we believe that it is appropriate to calculate R^2 as $1 - \text{RSS}/\text{TSS}$.