CIS 4190/5190: Applied Machine Learning	Spring 2023
Homework 2	
Handed Out: January 25	*Due: February 8*

Name: Jingxuan Bao

PennKey: bjx

PennID: 82897132

Note: This document is a read-only file. To create an editable version click on Menu in the top left corner of your screen and choose the Copy Project option.

1 Multiple Choice & Written Questions

1. (a) decrease bias, decrease variance
   (b) increase bias, decrease variance
   (c) decrease bias, increase variance
2. A. global maximum
3. (a) $\frac{\partial R_1}{\partial \beta_j} = \lambda sgn(\beta_j)$
   $\frac{\partial R_2}{\partial \beta_j} = 2\lambda\beta_j$
   (b) L_1 does a better job. Based on the assumption, the partial derivation, the MSE is zero, and β_j is small. This means $2\lambda\beta_j$ should also is a small value. But the $sgn(\beta_j)$ is a constant value related to λ. So only if β_j is equal to zero, the higher value of derivation L_1 hold, could do a better job to push β_j into zero.
4. (a) For $x \in Uniform([0, 1])$
   True function should be $y = x$
   Suppose we train the linear regression model with zero MSE, so the linear model should be $y = x, a = 1, b = 0$
   $MSE = \frac{1}{n}\sum_{i=1}^{n}(0 - x_i)^2$
   if $n- > \infty$, so $MSE = \int_{-1}^{0} x^2 dx = \frac{1}{3}$
   (b) For $x \in Uniform([-1, 0])$
   True function should be $y = 0$
   Suppose we train the linear regression model with zero MSE, so the linear model should be $y = 0, a = 0, b = 0$
   $MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - 0)^2$
   if $n- > \infty$, so $MSE = \int_{0}^{1} x^2 dx = \frac{1}{3}$

5. $f_{\hat{\beta}} = \hat{\beta}^T x$
   $= x^T \hat{\beta}$
   $= x^T (X^T X)^{-1} X^T Y$ $(\hat{\beta} = (X^T X)^{-1} X^T Y, Y = (y_1, y_2, ..., y_n)^T)$
   Let $F = (X^T X)^{-1} X^T$
   If we have $f_{\hat{\beta}} = \sum_{i=1}^{n} k_i(x; X) y_i$
   Let $k_i = x_i f_i$, where f_i is the ith column of F
   So we find the function k_i

2 Python Programming Questions

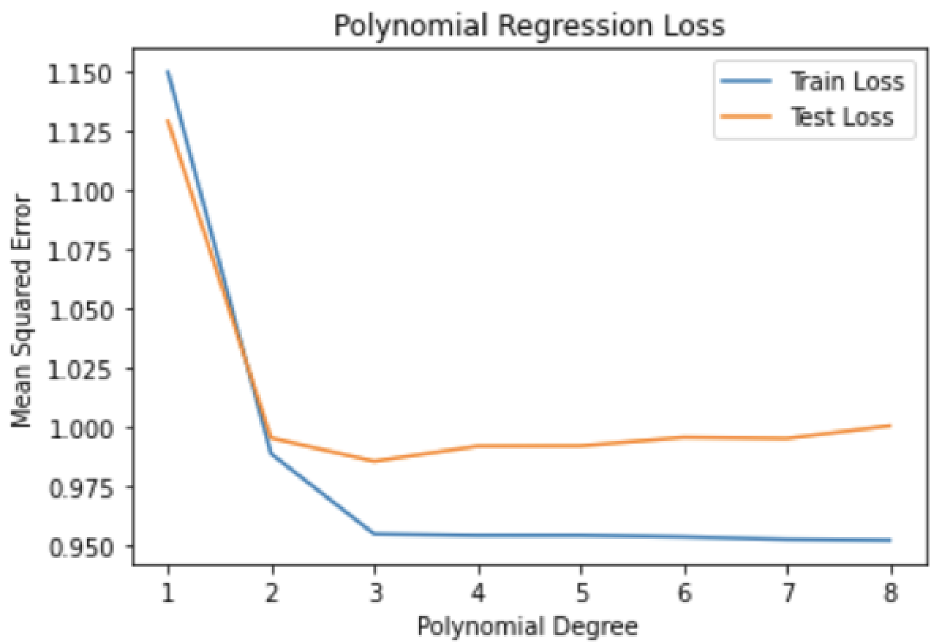

Figure 1: Figure for Q1.3

1. According to Figure 1 above, when the degree is small, the loss for both train dataset and test dataset decreases very fast if the degree increase. In this stage, the more complex model tends to bring a lower value of loss for both train dataset and test dataset. But after the degree increase over 3, the test loss tends to increase. That's because the model is overfitted. And also the decreasing trend of train loss also becomes slower after degree over 3. So based on this figure, select degrees as 3 is the best choice.

2. According to Figure 2 below, the learning rate influences the rate of model converge. In the range of learning rate from 0.001 to 0.1, a higher learning rate will lead the model converge faster. But the model will never converge when the learning rate equals 1 (purple line). That's because a high learning rate may let the β change too fast in each iteration to miss the optimal point. In the problem, the best learning rate should be 0.1.

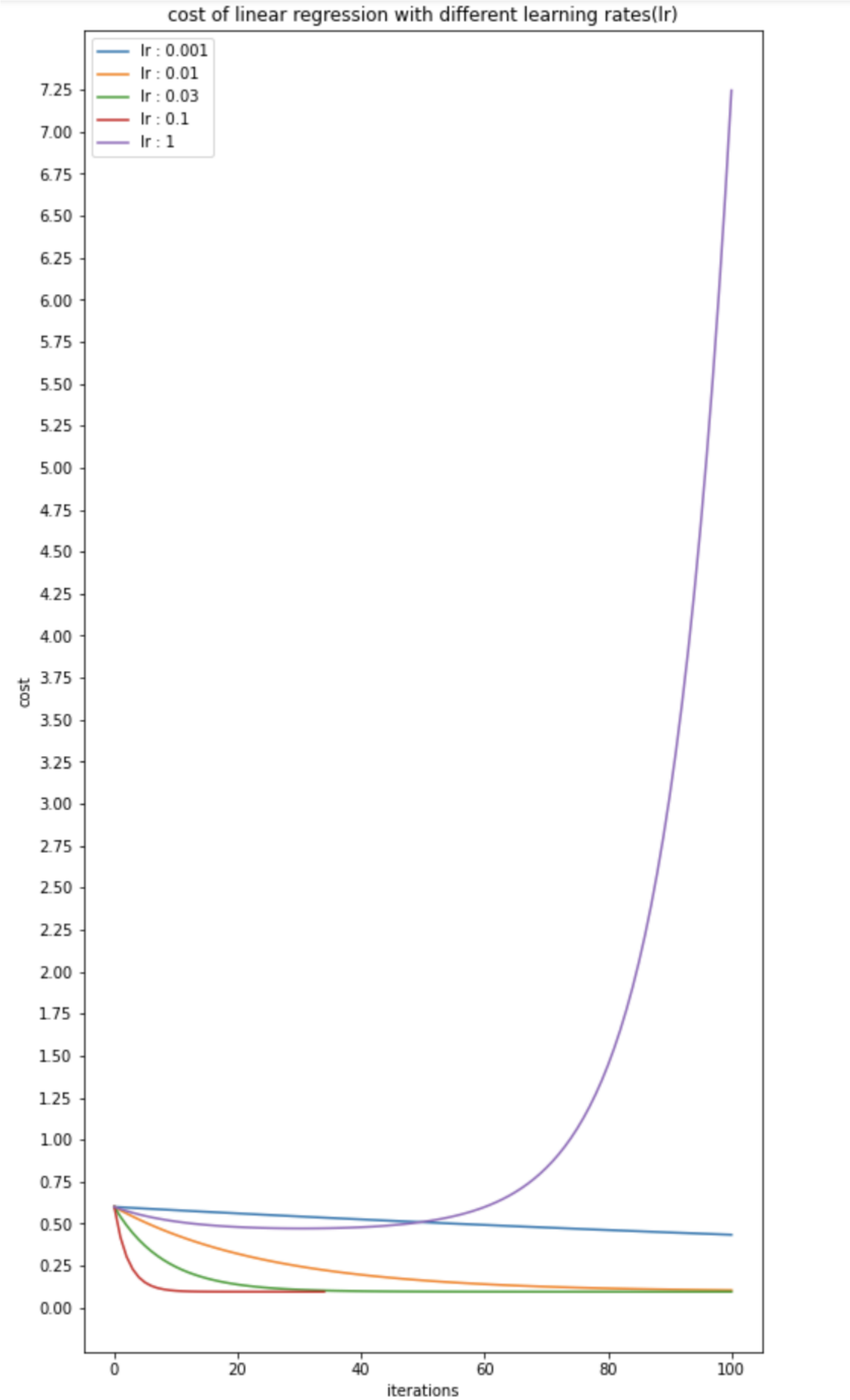

Figure 2: Figure for Q1.4