CIS 4190/5190: Applied Machine Learning | Fall 2023

Homework 2

Handed Out: September 13 | *Due: September 27*

Name: YUNJIE QU

PennKey: yunjiequ

PennID: 56923113

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) Stay the same, decrease variance
   (b) Increase bias, decrease variance
   (c) Decrease bias, increase variance
   (d) Bias and the variance keep the same.
   (e) Bias and the variance keep the same.(but inappropriate gradient descent step size might prevent model to converge to optimal point)

   Increase n, increase λ, decrease d, c and α has no impact

2. (a)

$$R_1(\beta) = \lambda \sum_{j'=1}^{d} |\beta_{j'}|$$

$$\frac{\partial}{\partial \beta} R_1(\beta) = \begin{cases} \lambda \text{sgn}(\beta_{j'}) & \text{if } \beta_{j'} \neq 0 \\ \text{undefined} & \text{if } \beta_{j'} = 0 \end{cases}$$

$$R_2(\beta) = \lambda \sum_{j'=1}^{d} \beta_{j'}^2$$

$$\frac{\partial}{\partial \beta_{j'}} R_2(\beta) = 2\lambda \beta_{j'}$$

   (b) L1 regularization

3. (a) Training on points $x \sim \text{Uniform}([0, 1])$ with $y = |x|$
   - Learned Model: The true function simplifies to $\hat{y} = x$ for $x \geq 0$.
   - Coefficients: $a = 1$, $b = 0$.

- MSE on $x \sim \text{Uniform}([-1,0])$:

$$\int_{-1}^{0} (-x - x)^2 \, dx = \frac{4}{3}$$

(b) Training on points $x \sim \text{Uniform}([-1,0])$ with $y = |x|$

- Learned Model: The true function simplifies to $\hat{y} = -x$ for $x \leq 0$.
- Coefficients: $a = -1$, $b = 0$.
- MSE on $x \sim \text{Uniform}([0,1])$:

$$\int_{0}^{1} (x - (-x))^2 \, dx = \frac{4}{3}$$

2 Python Programming Questions

TODO: Please add the resulting plots for Question 1.3 and Question 1.4.

2.1 Question 1.3 (Figure 1)

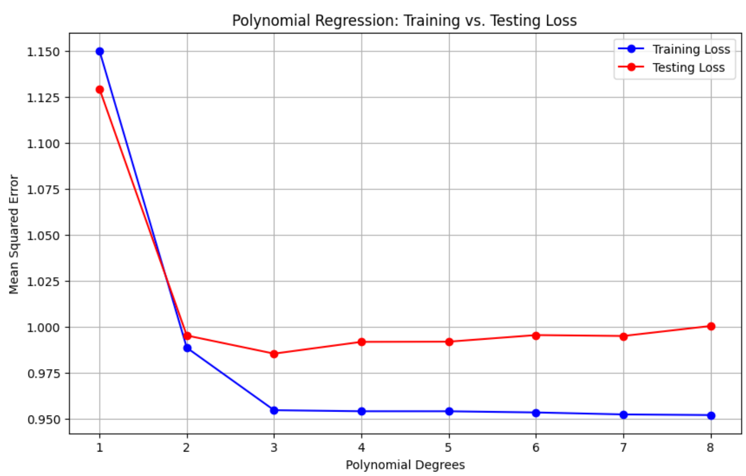

Figure 1: Figure for 1.3

- **Initial Decrease in Loss:** As polynomial degree increases from 1 to 3, both training and test losses decrease, indicating better model fit.

- **Divergence After a Point:** Beyond degree 3, while training loss keeps decreasing slightly, test loss starts increasing, signaling overfitting.

Reasoning:

- **Underfitting at Low Degrees:** At lower degrees, the model is too simplistic (high bias), resulting in high training and test losses.

- **Optimal Complexity:** At degree 3, the model complexity strikes a balance, reducing both bias and variance.
- **Overfitting at High Degrees:** Degrees beyond 3 lead to overfitting, where the model fits noise (high variance), causing increased test loss.

Conclusion:

In summary, increasing polynomial degree initially reduces bias but eventually increases variance, leading to overfitting. The optimal degree is 3, balancing pattern capture and noise avoidance. The training and test losses follow similar trends but diverge beyond degree 3, highlighting the bias-variance tradeoff.

2.2 Question 1.4 (Figure 2)

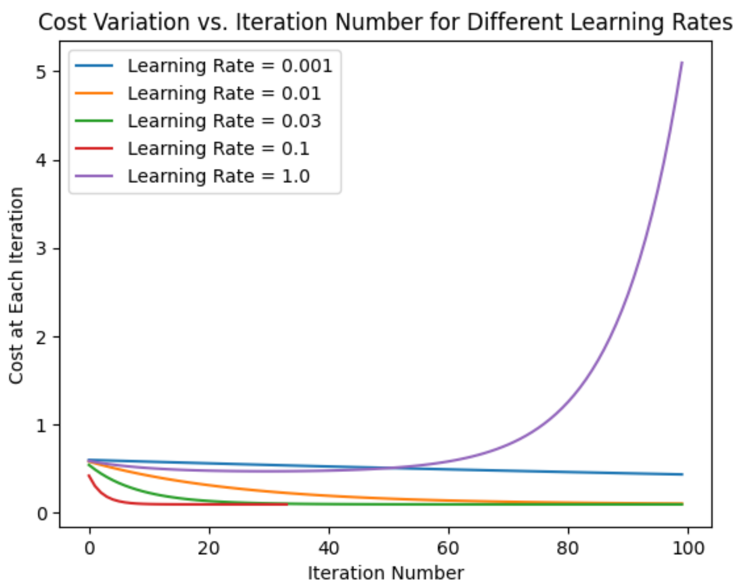

Figure 2: Figure for 1.4

- Increasing the learning rate can effectively decrease the cost at each iteration and reduce the number of iterations required for gradient descent to converge.
- However, using an excessive learning rate, such as 1.0, can lead to significantly higher costs at certain iterations, indicating divergence.
- Among the five learning rates tested (0.001, 0.01, 0.03, 0.1, and 1.0), the best learning rate appears to be 0.1 based on the plot.
- **Learning Rate 0.001:** Converges very slowly, requiring more iterations to reach a minimum.
- **Learning Rate 0.01:** Converges well but not the fastest.

- **Learning Rate 0.03:** Converges rapidly, even faster than 0.01.
- **Learning Rate 0.1:** Converges very quickly.
- **Learning Rate 1.0:** Diverges, indicating the learning rate is too high.

Effect of Increasing Learning Rate: Increasing the learning rate accelerates convergence up to a certain point, after which it can lead to divergence.

Best Learning Rate: Based on the plot and convergence behavior, a learning rate of 0.1 seems to be the most suitable among the provided options. It converges quickly without signs of divergence.