

# NYCU Introduction to Machine Learning, Homework 1

110550088 李杰穎

## 1 Part. 1, Coding (50%)

### 1.1 (10%) Linear Regression Model - Closed-form Solution

1. (10%) Show the weights and intercepts of your linear model.

### 1.2 (40%) Linear Regression Model - Gradient Descent Solution

2. (0%) Show the learning rate and epoch (and batch size if you implement mini-batch gradient descent) you choose.
3. (10%) Show the weights and intercepts of your linear model.
4. (10%) Plot the learning curve. (x-axis=epoch, y-axis=training loss)
5. (20%) Show your error rate between your closed-form solution and the gradient descent solution.

Describe why MSE is effective for resolving most linear regression problems and list scenarios where MSE may be inappropriate for data modeling, proposing alternative loss functions suitable for linear regression modeling in those cases.

4. (15%) In the lecture, we learned that there is a regularization method for linear regression models to boost the model's performance.

$$E_D(\mathbf{w}) + \lambda E_W(\mathbf{w})$$

- 4.1. (5%) Will the use of the regularization term always enhance the model's performance? Choose one of the following options: "Yes, it will always improve," "No, it will always worsen," or "Not necessarily always better or worse."
- 4.2. We know that  $\lambda$  is a parameter that should be carefully tuned. Discuss the following situations:
  - 4.2.1. (5%) Discuss how the model's performance may be affected when  $\lambda$  is set too small. For example,  $\lambda = 10^{-100}$  or  $\lambda = 0$
  - 4.2.2. (5%) Discuss how the model's performance may be affected when  $\lambda$  is set too large. For example,  $\lambda = 1000000$  or  $\lambda = 10^{100}$

## 2 Part. 2, Questions (50%)

1. (10%) How does the value of learning rate impact the training process in gradient descent? Please explain in detail.
2. (10%) There are some cases where gradient descent may fail to converge. Please provide at least two scenarios and explain in detail.
3. (15%) Is mean square error (MSE) the optimal selection when modeling a simple linear regression model?