

# NYCU Introduction to Machine Learning, Homework 1

your student ID, your name

The screenshot and the figures we provided below are just examples. **The results below are not guaranteed to be correct.** Please make sure your answers are clear and readable, or no points will be given. Please also remember to convert it to a pdf file before submission. **You should use English to answer the questions.** After reading this paragraph, you can delete this paragraph.

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

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

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

```
Closed-form Solution
Weights: , Intercept: 
```

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

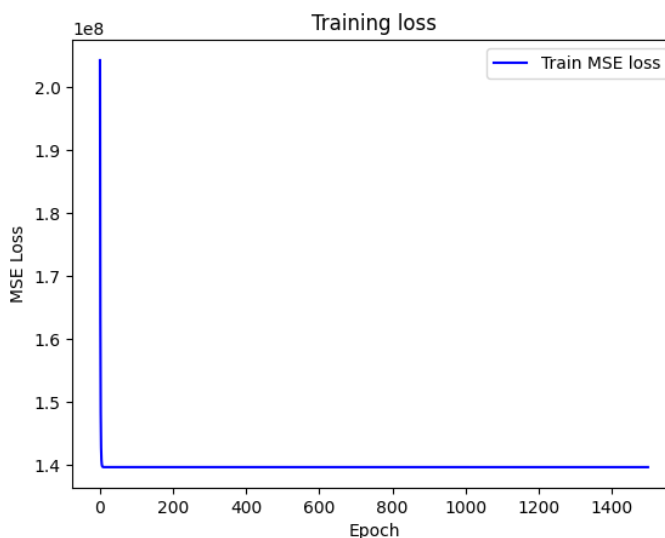
2. (0%) Show the learning rate and epoch (and batch size if you implement mini-batch gradient descent) you choose.

```
LR.gradient_descent_fit(train_x, train_y, lr=0.01, epochs=100)
```

3. (10%) Show the weights and intercepts of your linear model.

```
Gradient Descent Solution
Weights: , Intercept: 
```

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.

```
Error Rate: 0.3%
```

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

1. (10%) How does the value of learning rate impact the training process in gradient descent? Please explain in detail.

Write or type your answer here.

2. (10%) There are some cases where gradient descent may fail to converge. Please provide at least two scenarios and explain in detail.

Write or type your answer here.

3. (15%) Is mean square error (MSE) the optimal selection when modeling a simple linear regression model? 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.

Write or type your answer here.

4. (15%) In the lecture, we learned that there is a regularization method for linear regression models to boost the model's performance. (p18 in linear\_regression.pdf)

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

- 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: (both in 100 words)

- 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  $\lambda=10^{100}$

Write or type your answer here.