NYCU Pattern Recognition, Homework 1

**XXXXXXXXX, OOO**

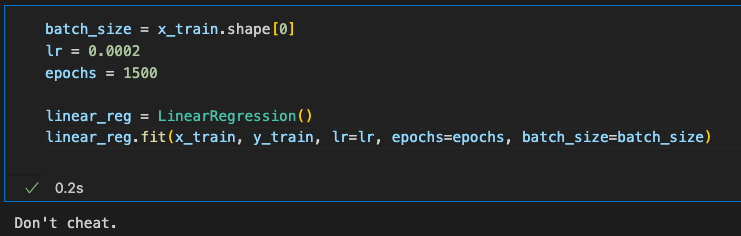
**Part. 1, Coding (70%)**:

You should type the answer and also screenshot at the same time. Otherwise, no points will be given. The screenshot and the figures we provided below are just examples. **The results below are not guaranteed to be correct.** Please convert it to a pdf file before submission. You should use English to answer the questions. After reading this paragraph, you can delete it.

1. (0%) Show the learning rate and epoch you choose

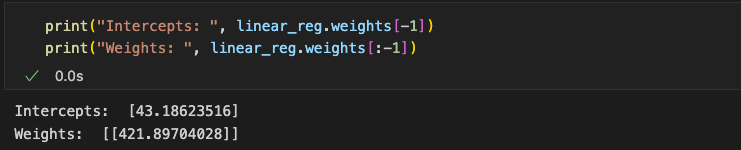
learning rate: \_\_\_

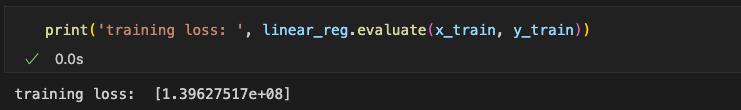
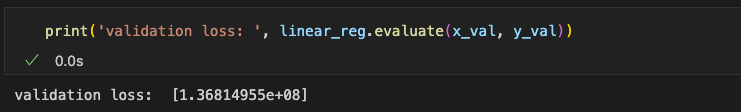
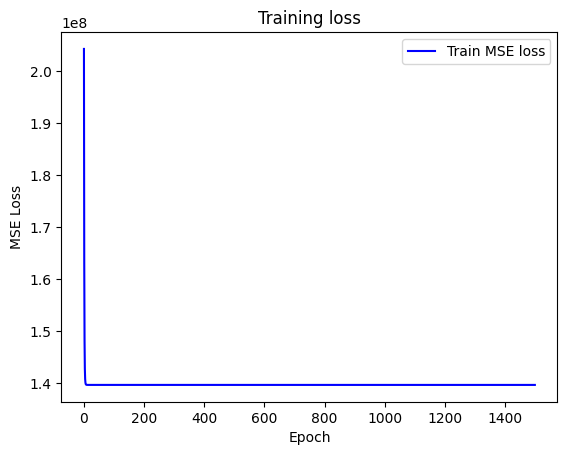
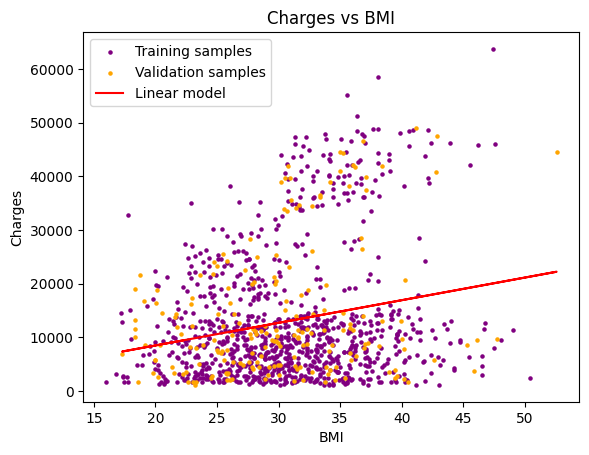
epoch: \_\_\_

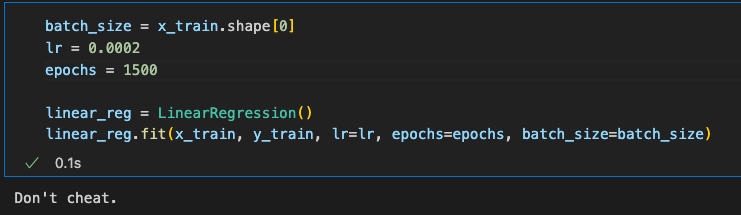


1. (5%) Show the weights and intercepts of your linear model.

weights: \_\_\_

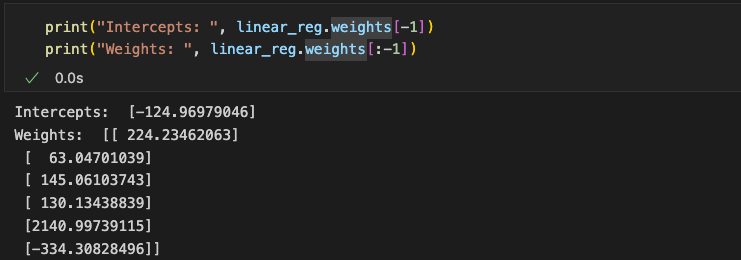
intercepts: \_\_\_

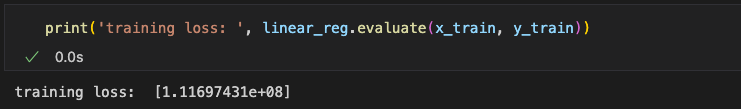
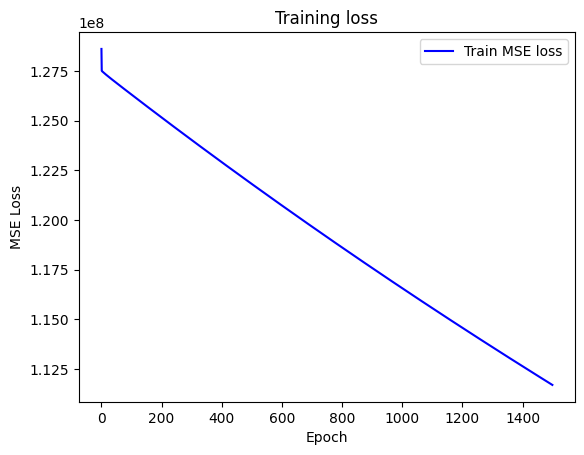
1. (5%) What’s your final training loss (MSE)?  
   training loss (MSE): \_\_\_
2. (5%) What’s the MSE of your validation prediction and validation ground truth?  
   validation loss (MSE): \_\_\_
3. (5%) Plot the training curve. (x-axis=epoch, y-axis=loss)  
   
4. (5%) Plot the line you find with the training and validation data.  
   
5. (0%) Show the learning rate and epoch you choose.  
   learning rate: \_\_\_

epoch: \_\_\_

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

weights: \_\_\_

intercepts: \_\_\_

1. (5%) What’s your final training loss?  
   training loss (MSE): \_\_\_
2. (5%) What’s the MSE of your validation prediction and validation ground truth?  
   validation loss (MSE): \_\_\_  
   
3. (5%) Plot the training curve. (x-axis=epoch, y-axis=loss)  
   
4. (20%) Train your own model and fill the testing CSV file as your final predictions.   
     
   learning rate: \_\_\_  
   epoch: \_\_\_  
   batch\_size: \_\_\_  
   Used features: \_\_\_  
     
   What data analysis have you done? Why choose the above setting? Other strategies? (please explain in detail; otherwise, no points will be given.)

**Part. 2, Questions (30%):**

(7%) 1. What’s the difference between Gradient Descent, Mini-Batch Gradient Descent, and Stochastic Gradient Descent?

Your answer here. Please answer the question in detail. You can directly type here or take a photo of your handwriting answer and paste it here. After reading this paragraph, you can delete it.

(7%) 2. Will different values of learning rate affect the convergence of optimization? Please explain in detail.

Your answer here. Please answer the question in detail. You can directly type here or take a photo of your handwriting answer and paste it here. After reading this paragraph, you can delete it.

(8%) 3. Suppose you are given a dataset with two variables, X and Y, and you want to perform linear regression to determine the relationship between these variables. You plot the data and notice that there is a strong nonlinear relationship between X and Y. Can you still use linear regression to analyze this data? Why or why not? Please explain in detail.

Your answer here. Please answer the question in detail. You can directly type here or take a photo of your handwriting answer and paste it here. After reading this paragraph, you can delete it.

(8%) 4. In the coding part of this homework, we can notice that when we use more features in the data, we can usually achieve a lower training loss. Consider two sets of features, A and B, where B is a subset of A. (1) Prove that we can achieve a non-greater training loss when we use the features of set A rather than the features of set B. (2) In what situation will the two training losses be equal?

Your answer here. Please answer the question in detail. You can directly type here or take a photo of your handwriting answer and paste it here. After reading this paragraph, you can delete it.