Pattern Recognition HW1

M093781 趙宇涵

Part 1. Coding (60%)

Please find the following link to check my coding: <https://github.com/honey0703/CS_AT0828/blob/main/HW1/HW1.ipynb>

Here I edited the hyper parameters to leaning rate = 1e-3, iterations = 10000 to get a better training result.

1. (15%) Plot the learning curve of the training, you should find that loss decreases after a few iterations (x-axis=iteration, y-axis=loss, Matplotlib or other plot tools is available to use)

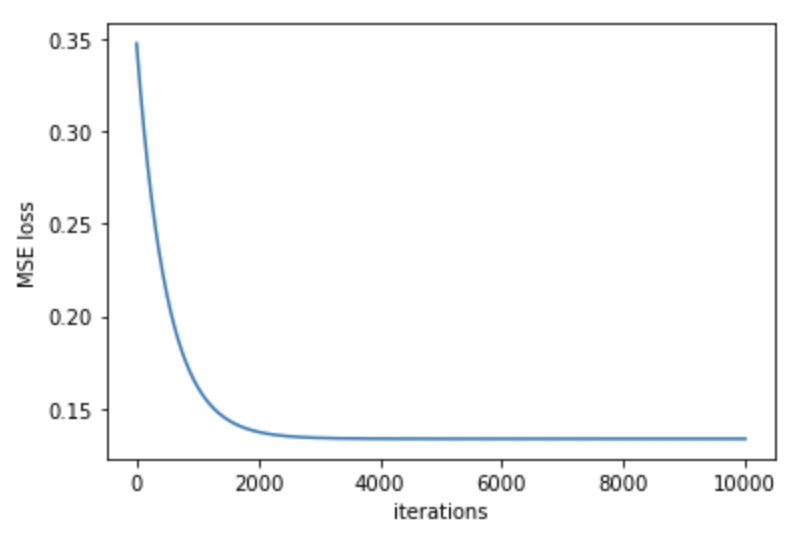


Fig.1 Learning curve of training data

1. (15%) What's the Mean Square Error of your prediction and ground truth (prediction=model(x\_test), ground truth=y\_test)

After I repeat training for more than 5 times, my MSE losses are about **0.03435**

1. (15%) What're the weights and intercepts of your linear model?

**Weight: 0.81795508**

**Intercepts: 0.7845605**

一張含有 文字 的圖片

自動產生的描述

Fig.2 weight and intercepts

1. (10%) What’s the difference between Gradient Descent, Mini-Batch Gradient

Descent, and Stochastic Gradient Descent?

The main difference between these three methods is the different size of training dataset. Here I introduce their training data and their pros and cons individually.

* Gradient Descent:

The hole training dataset is used during training process. It updates the model after calculating all loss of entire training dataset.

* + Pros: It can get a smooth curve, such as Fig.1. It can continually reach the lowest loss score.
  + Cons: The progress will take long time because all training data need to be considered during 1 update.
* Mini-Batch Gradient Descent:

It samples few examples from training dataset during 1 update. These sample data are called a mini batch.

* + Pros: It combines the pros of GD and SGD. Training process is faster than GD, and few fluctuating than SGD.
  + Cons: It combines the cons of GD and SGD. Training process is slower than SGD, and more fluctuating than GD.
* Stochastic Gradient Descent

It only uses 1 example in 1 update.

* + Pros: Training process is much faster than GD and mini-batch GD.
  + Cons: Due to updating with one example, the loss won’t always decrease. Thus, the learning curve will fluctuate seriously.