NYCU Introduction to Machine Learning, Homework 1

Deadline: 23:59, Oct. 10 (Tue), 2023

Part. 1, Coding (50%):

In this coding assignment, you are required to implement linear regression using only NumPy. Please note that only NumPy can be used to implement your model. Therefore, you will get no points by simply calling sklearn.linear_model.LinearRegression.
You are required to implement linear regression by closed-form solution and gradient descent.

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

Requirements:

- Implement Linear Regression by closed-form solution.

Criteria:

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

(40%) Linear Regression Model - Gradient Descent Solution

Requirements:

- Update your weights and intercept by gradient descent (you can implement mini-batch gradient descent or stochastic gradient descent if you want).
- Use MSE (Mean Square Error) as your loss function.

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

(Note: You must use this provided function, do not multiply any constant factor such as 1/2)

- Tune the learning rate and epoch hyper-parameters (and batch size if you implement mini-batch gradient descent) to make your testing MSE loss as closed as the closed-form solution.

(next page)

Criteria:

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.
   note: error rate: (gradient_descent_loss - closed_form_loss) / closed_form_loss * 100

Points	**error rate**
20	**< 0.5%**
15	**< 1%**
10	**< 3%**
5	**< 5%**
0	**>= 5%**

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? 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.

(next page)

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 λ 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 λ 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 λ is set too large. For example, $\lambda=1000000$ or $\lambda=10^{100}$