NYCU Introduction to Machine Learning, Homework 1

Deadline: Oct. 25, 23:59

Part. 1, Coding (60%):

In this coding assignment, you need to implement logistic regression and linear regression by using only NumPy, then train your implemented model using **Gradient Descent** by the provided dataset and test the performance with testing data. Find the sample code and data on the GitHub page
https://github.com/NCTU-VRDL/CS_CS20024/tree/main/HW1

Please note that only NumPy can be used to implement your model. You will get no points by simply calling sklearn.linear_model.LinearRegression. Moreover, please train your linear model using Gradient Descent, not the closed-form solution.

Linear regression model

1. (10%) Plot the learning curve of the training, you should find that loss decreases after a few iterations and finally converge to zero (x-axis=iteration, y-axis=loss, Matplotlib or other plot tools is available to use)
2. (10%) What's the Mean Square Error of your prediction and ground truth?
3. (10%) What're the weights and intercepts of your linear model?

Logistic regression model

1. (10%) Plot the learning curve of the training, you should find that loss decreases after a few iterations and finally converge to zero (x-axis=iteration, y-axis=loss, Matplotlib or other plot tools is available to use)
2. (10%) What's the Cross Entropy Error of your prediction and ground truth?
3. (10%) What're the weights and intercepts of your linear model?

Print the answers from your code and paste them onto the report

Part. 2, Questions (40%):

1. What's the difference between Gradient Descent, Mini-Batch Gradient Descent, and Stochastic Gradient Descent?

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

3. Show that the logistic sigmoid function (eq. 1) satisfies the property $\sigma(-a) = 1 - \sigma(a)$ and that its inverse is given by $\sigma^{-1}(y) = \ln \{y/(1 - y)\}$.

$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

(eq. 1)

4. Show that the gradients of the cross-entropy error (eq. 2) are given by (eq. 3).

$$E(\mathbf{w}_1, \ldots, \mathbf{w}_K) = -\ln p(\mathbf{T}|\mathbf{w}_1, \ldots, \mathbf{w}_K) = -\sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln y_{nk}$$

(eq. 2)

$$\nabla_{\mathbf{w}_j} E(\mathbf{w}_1, \ldots, \mathbf{w}_K) = \sum_{n=1}^{N} (y_{nj} - t_{nj})\, \boldsymbol{\phi}_n$$

(eq. 3)

Hints:

$$a_k = \mathbf{w}_k^{\mathrm{T}} \boldsymbol{\phi}.$$

(eq. 4)

$$\frac{\partial y_k}{\partial a_j} = y_k(I_{kj} - y_j)$$

(eq. 5)