NYCU Introduction to Machine Learning, Homework 2

Deadline: Nov. 14, 23:59

Part. 1, Coding (50%):

In this coding assignment, you are requested to implement Logistic Regression and Fisher's Linear Discriminant by using only Numpy. After that, train your model on the provided dataset and evaluate the performance on the testing data.

(15%) Logistic Regression

Requirements:

- Use Gradient Descent to update your model
- Use CE (Cross-Entropy) as your loss function.

Criteria:

1. (0%) Show the hyperparameters (learning rate and iteration) that you used.
2. (5%) Show the weights and intercept of your model.
3. (10%) Show the accuracy score of your model on the testing set. The accuracy score should be greater than 0.75.

(35%) Fisher's Linear Discriminant (FLD)

Requirements:

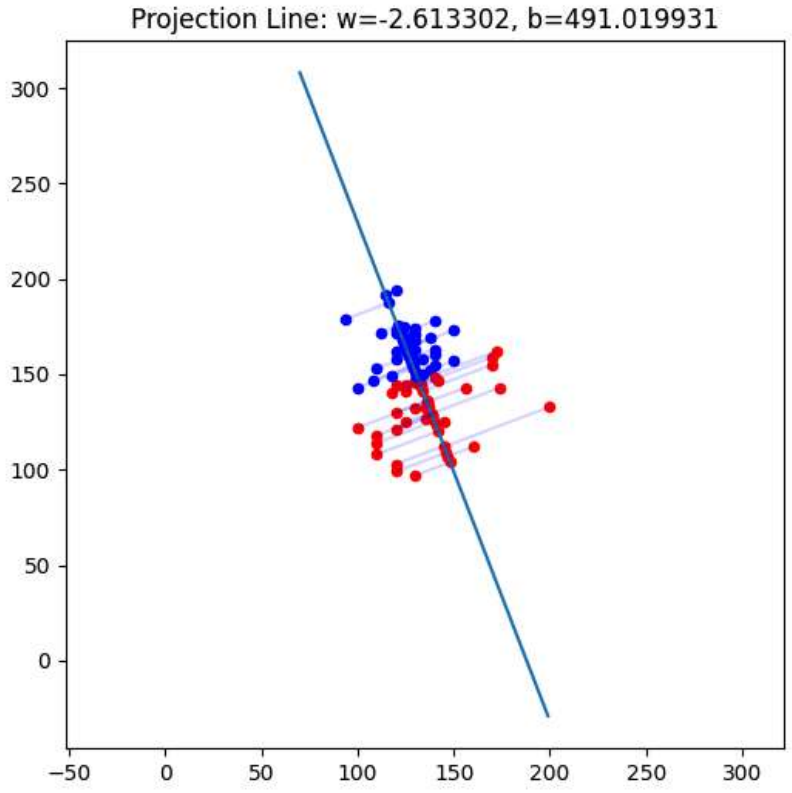
- Implement FLD to reduce the dimension of the data from 2-dimensional to 1-dimensional.

Criteria:

4. (0%) Show the mean vectors m_i (i=0, 1) of each class of the training set.
5. (5%) Show the within-class scatter matrix S_W of the training set.
6. (5%) Show the between-class scatter matrix S_B of the training set.
7. (5%) Show the Fisher's linear discriminant w of the training set.
8. (10%) Obtain predictions for the testing set by measuring the distance between the projected value of the testing data and the projected means of the training data for the two classes. Show the accuracy score on the testing set. The accuracy score should be greater than 0.65.

(next page)

9. (10%) Plot the projection line (x-axis: age, y-axis: thalach).
   1) Plot the projection line trained on the training set and show the slope and intercept on the title (you can choose any value of intercept for better visualization).
   2) Obtain the prediction of the testing set, plot and colorize them based on the prediction.
   3) Project all testing data points on your projection line. Your result should look like the below image.



(This image is a reference, not the answer)

Part. 2, Questions (50%):

1. (5%) What's the difference between the sigmoid function and the softmax function? In what scenarios will the two functions be used? Please at least provide one difference for the first question and answer the second question respectively.

2. (10%) In this homework, we use the cross-entropy function as the loss function for Logistic Regression. Why can't we use Mean Square Error (MSE) instead? Please explain in detail.

(next page)

3. (15%) In a multi-class classification problem, assume you have already trained a classifier using a logistic regression model, which the outputs are P1, P2, ... Pc, how do you evaluate the overall performance of this classifier with respect to its ability to predict the correct class?
   3.1. (5%) What are the metrics that are commonly used to evaluate the performance of the classifier? Please at least list three of them.
   3.2. (5%) Based on the previous question, how do you determine the predicted class of each sample?
   3.3. (5%) In a class imbalance dataset (say 90% of class-1, 9% of class-2, and 1% of class-3), is there any problem with using the metrics you mentioned above and how to evaluate the model prediction performance in a fair manner?

4. (20%) Calculate the results of the partial derivatives for the following equations. (The first one is binary cross-entropy loss, and the second one is mean square error loss followed by a sigmoid function. σ is the sigmoid function.)
   4.1. (10%)

$$\frac{\partial}{\partial x}\left(-t * \ln\left(\sigma(x)\right) - (1-t) * \ln\left(1-\sigma(x)\right)\right)$$

   4.2. (10%)

$$\frac{\partial}{\partial x}\left(\left(t-\sigma(x)\right)^2\right)$$