NYCU Introduction to Machine Learning, Homework 2

109550085,陳欣妤

The screenshot and the figures we provided below are just examples. **The results below are not guaranteed to be correct.** Please make sure your answers are clear and readable, or no points will be given. Please also remember to convert it to a pdf file before submission. You should use English to answer the questions. After reading this paragraph, you can delete this paragraph.

Part. 1, Coding (50%):

(15%) Logistic Regression

1. (0%) Show the hyperparameters (learning rate and iteration) that you used.

```
LR = LogisticRegression(learning_rate=3e-4, iteration=300000)
```

2. (5%) Show the weights and intercept of your model.

```
Part 1: Logistic Regression
Weights: [-0.07027484 -1.83280359  1.05316516  0.07984141  0.03081616 -0.68540369], Intercept: 1.3944771425536842
```

3. (10%) Show the accuracy score of your model on the testing set. The accuracy score should be greater than 0.75.

```
Accuracy: 0.7540983606557377
```

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

4. (0%) Show the mean vectors m_i (i=0, 1) of each class of the training set.

```
Part 2: Fisher's Linear Discriminant
Class Mean 0: [ 56.75925926 137.7962963 ], Class Mean 1: [ 52.63432836 158.97761194]
```

5. (5%) Show the within-class scatter matrix S_W of the training set.

```
With-in class scatter matrix:
[[ 19184.82283029 -16006.39331122]
 [-16006.39331122 106946.45135434]]
```

6. (5%) Show the between-class scatter matrix S_B of the training set.

```
Between class scatter matrix:
[[ 17.01505494 -87.37146342]
 [-87.37146342 448.64813241]]
```

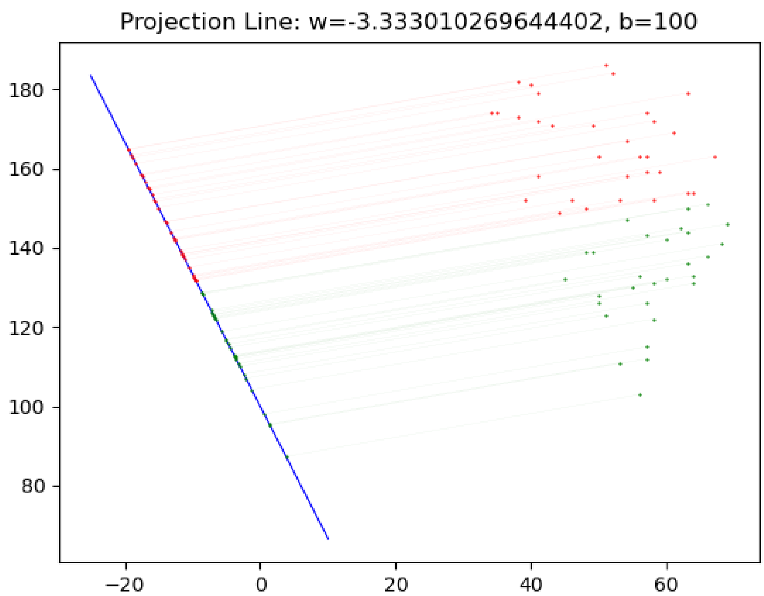
7. (5%) Show the Fisher's linear discriminant *w* of the training set.

```
W:
[[-5.68686969e-05]
 [ 1.89543951e-04]]
```

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.

```
Accuracy of FLD: 0.6557377049180327
```

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.



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.

   The sigmoid function is used for binary classification problems. However, the softmax function is used for multi-class classification problems.

   The sigmoid function is used in scenarios where there are only two possible outcomes, such as predicting whether an email is spam or not spam. The softmax function, on the other hand, is used when dealing with multi-class classification problems, such as classifying handwritten digits into one of multiple categories (0 to 9) or categorizing different types of animals.

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.

   The MSE loss function assumes Gaussian distributed errors, which is not suitable for binary classification tasks like logistic regression. Since logistic regression deals with probabilities and binary outcomes, the assumptions of Gaussian distribution may not hold. The cross-entropy loss function, on the other hand, is derived from the maximum likelihood estimation (MLE) framework, making it more appropriate for binary classification problems.

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.

     Accuracy, Precision, Recall
   - 3.2. (5%) Based on the previous question, how do you determine the predicted class of each sample?

     The predicted class for each sample is determined based on the class with the highest predicted probability. In other words, the class with the maximum probability (P1, P2, ..., Pc) is selected as the predicted class for the corresponding 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?

F1 Score metrics offer a balanced assessment of the model's performance, considering both the true positives and false positives for each class, thus providing a more reliable evaluation in the presence of class imbalance.

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%)

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

$$= -t * \frac{1}{\sigma(x)} * \sigma(x)(1 - \sigma(x)) + (1 - t) * \frac{1}{1-\sigma(x)} * (-1) * \sigma(x)(1 - \sigma(x))$$

$$= -t * (1 - \sigma(x)) + (1 - t) * (-\sigma(x))$$

$$= -t + t\sigma(x) - \sigma(x) + t\sigma(x)$$

$$= -t + 2t\sigma(x) - \sigma(x)$$

4.2. (10%)

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

$$= 2(t - \sigma(x)) * (-1) * \sigma(x)(1 - \sigma(x))$$

$$= -2(t - \sigma(x)) * \sigma(x)(1 - \sigma(x))$$

$$= -2(t\sigma(x) - \sigma(x)^2) * (1 - \sigma(x)$$