NYCU Introduction to Machine Learning, Homework 1

110550014, 吳權祐

Part. 1, Coding (50%):

(15%) Logistic Regression

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

```
LR = LogisticRegression(learning_rate=0.00015, iteration=70000)
```

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

```
Weights: [-0.05398208 -0.59325989  0.82846562 -0.02775859  0.02684816 -0.47701721], Intercept: -0.05438151210740732б
```

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)

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

```
Class Mean 0: [ 56.75925926 137.7962963 ], Class Mean 1: [ 52.63432836 158.97761194]
```

2. (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]]
```

3. (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]]
```

4. (5%) Show the Fisher's linear discriminant w of the training set.

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

5. (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.

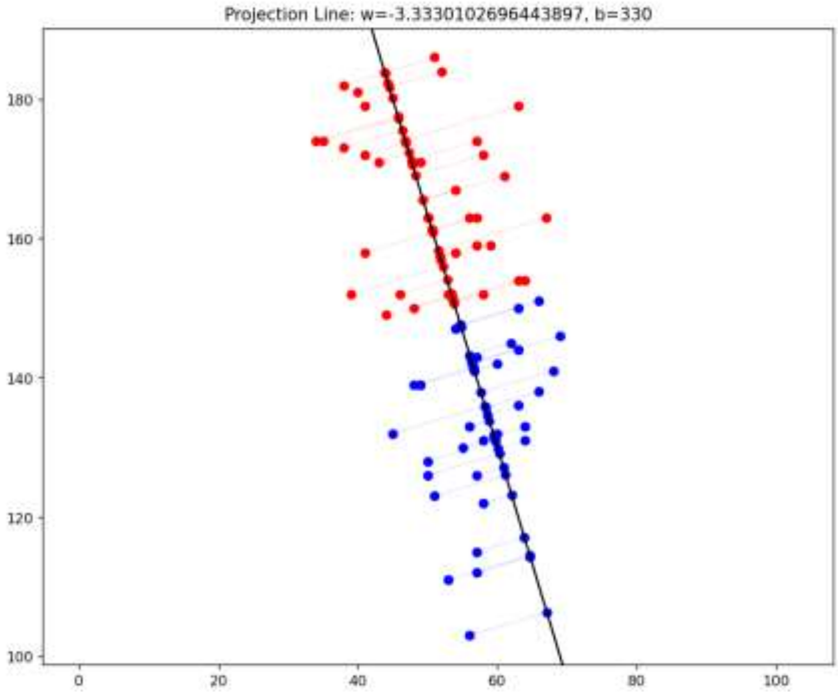
```
Accuracy of FLD: 0.6557377049180327
```

6. (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.

   Sigmoid function mapped a single value X into a probability between 0 and 1.
   Softmax function mapped a vector $X \in R^n$ into a probability distribution of n classes, and the sum of the vector is 1.
   The main difference between two functions is the mapping method, which cause to the different usage.
   Sigmoid function is usually used in the two-class classification problem, and softmax is usually used in multi-class classification.

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.

   Because if we use MSE as loss function, when predictions are close to the ground truthes, MSE are very small and this is very bad for gradient descent. We use CEE could provide larger convergence volicity since the error would not become too small.

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 cla ss?

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:

$$\frac{\{Numer\ of\ correct\ prediction\}}{\{Number\ of\ samples\}}$$

Recall (for each classes):

$$\frac{\{Numer\ of\ correct\ prediction\}}{\{Number\ of\ samples\ actual\ in\ the\ class\}}$$

Precision (for each classes):

$$\frac{\{Numer\ of\ correct\ prediction\}}{\{Number\ of\ samples\ predicted\ in\ the\ class\}}$$

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

Pick the maximum probability in P1, P2, …Pc and the corresponding class of the picked probability is the predicted class of the sample.

3.3. (5%) In a class imbalance dataset (say 90% of class-1, 9% of class-2, and 1% of cl ass-3), is there any problem with using the metrics you mentioned above and how t o evaluate the model prediction performance in a fair manner?

It not fair if we use accuracy as metrix, since if there is a model always give the prediction as class-1, it would get 90% accuracy, which is a high score but not reasonable evaluation.

So use Recall and or Precision in each classes is a more fair metrics, or we can use F1 score which combine these two metrics as $2 * \frac{Precision*Recall}{Precision+Recall}$, if the score in each classes are all the highest, the model is the best.

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.)

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}\left(1 + e^{-x}\right)^{-1} = e^{-x}\left(1 + e^{-x}\right)^{-2} = \sigma(x)\left(1 - \sigma(x)\right)$$

4.1. (10%)

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

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

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

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

4.2. (10%)

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

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

$$= -2\left(t\sigma(x) - \left(1 + t\right)\sigma(x)^2 + \sigma(x)^3\right)$$