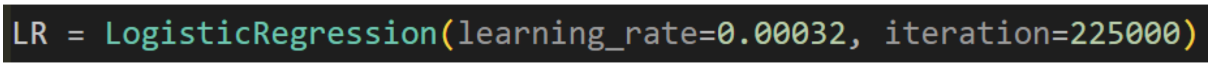
NYCU Introduction to Machine Learning, Homework 2

110550108 施柏江

Part. 1, Coding (50%):

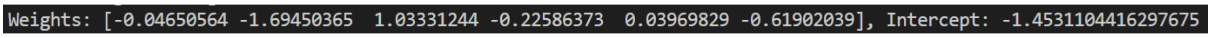
(15%) Logistic Regression

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



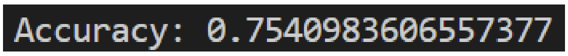
```
LR = LogisticRegression(learning_rate=0.00032, iteration=225000)
```

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



```
Weights: [-0.04650564 -1.69450365  1.03331244 -0.22586373  0.03969829 -0.61902039], Intercept: -1.4531104416297675
```

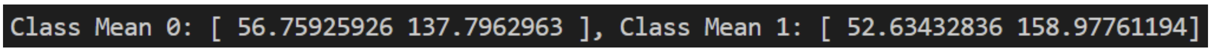
3. (10%) Show the accuracy score of your model on the testing set. The accuracy score should be greater than 0.75.



```
Accuracy: 0.7540983606557377
```

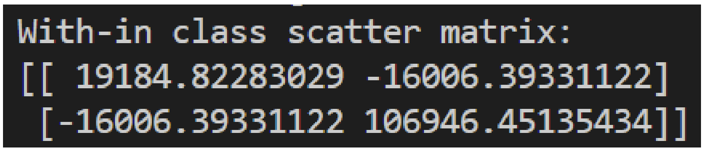
(35%) Fisher's Linear Discriminant (FLD)

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



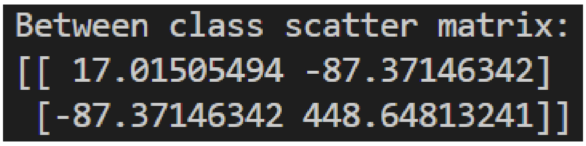
```
Class Mean 0: [ 56.75925926 137.7962963 ], Class Mean 1: [ 52.63432836 158.97761194]
```

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



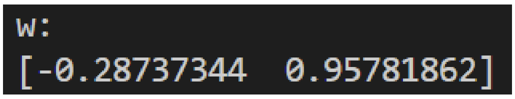
```
With-in class scatter matrix:
[[ 19184.82283029 -16006.39331122]
 [-16006.39331122 106946.45135434]]
```

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



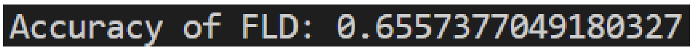
```
Between class scatter matrix:
[[ 17.01505494 -87.37146342]
 [-87.37146342 448.64813241]]
```

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



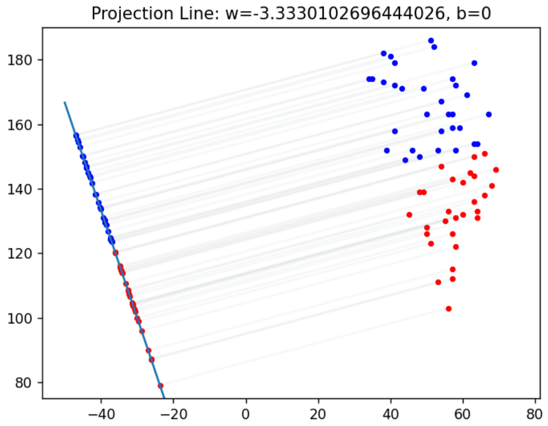
```
w:
[-0.28737344  0.95781862]
```

8. (10%) Obtain predictions for the testing set by measuring the distance between the projected outcome 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).



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.

Ans:

One key difference between them lies in their applications. The sigmoid function is typically used in binary classification tasks, aiming to predict outcomes with two possible classes. Its role is to compress the output into a range between 0 and 1, offering a probability interpretation for belonging to the positive class. In contrast, the softmax function finds its application in multi-class classification scenarios. It extends the functionality of the sigmoid to accommodate multiple classes, transforming output scores into probabilities that collectively sum to 1 across all classes. This adaptation makes softmax well-suited for situations where an input can be associated with one of several classes.

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.

Ans:

The cross-entropy loss is well-suited for this scenario because it penalizes confident and wrong predictions more heavily. This is important because in logistic regression, we're dealing with a classification problem, and we want the model to be penalized more when it's confidently wrong about the class probabilities. On the other hand, using Mean Square Error might not be ideal for logistic regression because it assumes a continuous output. MSE tends to penalize large errors more than smaller ones, which might not align with the objectives of a classification problem.

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?

Ans:

3.1.

First, we can measures the overall correctness of the classifier by calculating the ratio of correctly predicted instances to the total instances. Moreover, we can calculate the ratio of correctly predicted positive observations to the total predicted positives. It is useful when the cost of false positives is high. Last, we can calculate the ratio of correctly predicted positive observations to all observations in the actual class. It is useful when the cost of false negatives is high.

3.2.

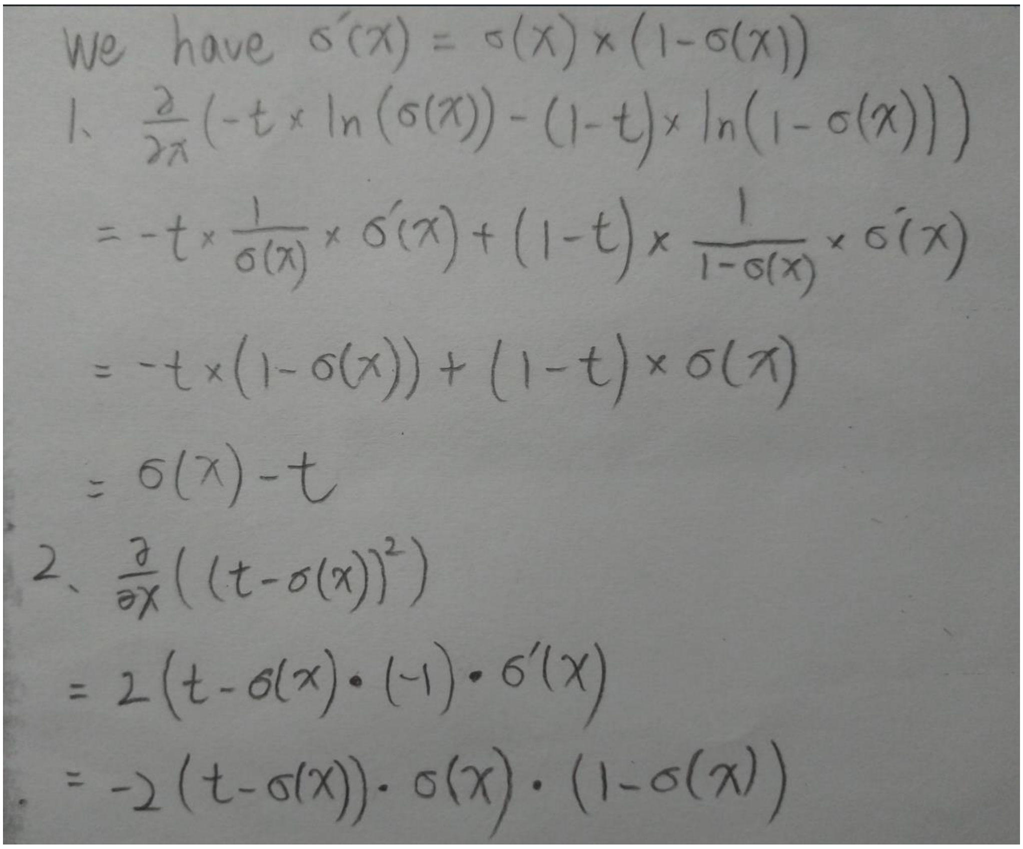
The predicted class for each sample is typically the one with the highest predicted probability. So, if P1 > P2 > ... > Pc, then the predicted class is the one corresponding to P1.

3.3.

In a class-imbalanced dataset, the model could achieve high accuracy by simply predicting the majority class. In such cases, it's essential to consider other metrics. We can use F1-Score, which combines precision and recall. Moreover, we can use confusion matrix to see a detailed breakdown of true positives, true negatives, false positives, and false negatives, helping to identify specific areas of improvement. Last, we can adjust class weights or use resampling techniques to give more importance to minority classes, ensuring a fair evaluation across all classes.

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.)

Ans:



We have $\sigma'(x) = \sigma(x) \times (1-\sigma(x))$

1、 $\frac{\partial}{\partial x}(-t \times \ln(\sigma(x)) - (1-t) \times \ln(1-\sigma(x)))$

$= -t \times \frac{1}{\sigma(x)} \times \sigma'(x) + (1-t) \times \frac{1}{1-\sigma(x)} \times \sigma'(x)$

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

$= \sigma(x) - t$

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

$= 2(t-\sigma(x) \cdot (-1) \cdot \sigma'(x)$

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