NYCU Introduction to Machine Learning, Homework 2

110550085 房天越

Part. 1, Coding (50%):

(15%) Logistic Regression

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

```
LR = LogisticRegression(learning_rate=0.0001, iteration=100000)
```

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

```
Weights: [-0.05401261 -0.57194471  0.81540993 -0.02539069  0.02665697 -0.46607183], Intercept: -0.052724935387732035
```

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 mi (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 SW 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 SB of the training set.

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

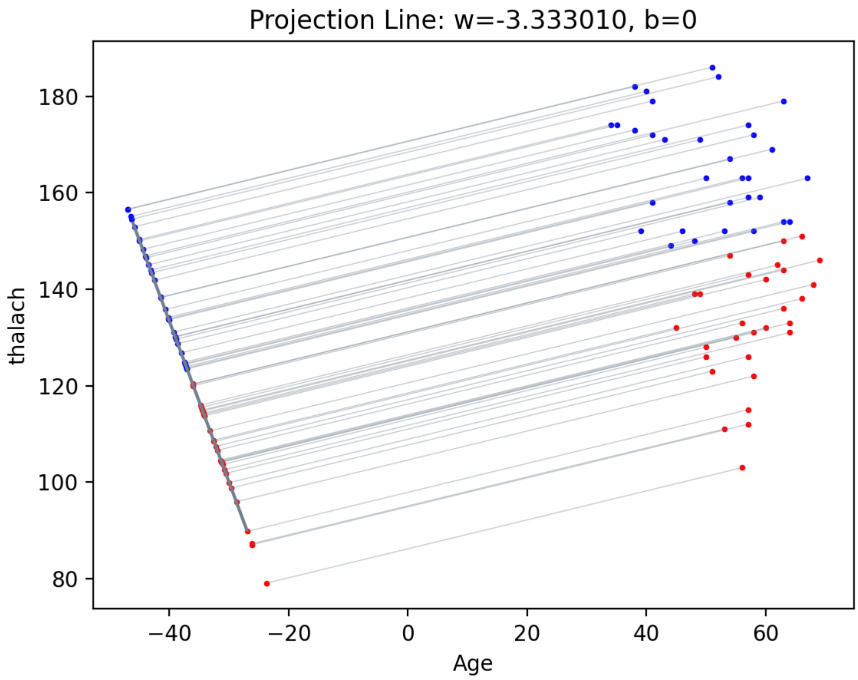
7. (5%) Show the Fisher's linear discriminant w of the training set.

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

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



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.
   I. The output of it is in the range of 0 to 1, the output can be interpreted as the probability of belonging to a particular class. While the output of the softmax function is a set of probability that sums up to 0, representing the likelihood of each class.
   II. The sigmoid function could be use in binary classification, for example, it could predict whether an email is spam or not. The softmax function could be used for multi-class classification, for example, it could classify the handwritten numbers to be 0 to 9.

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.
   MSE is generally used for problems where the output is continuous, and it could penalize large errors heavily. In this homework, the predicted

probabilities are often close to the boundaries, which are 0 or 1, MSE could lead to slow convergence and numerical instability.
Also, MSE is highly sensitive to the outliers, and in this homework, we could see a lot of data that are mislabeled or they are outliers. This could heavily impact the performance when using MSE.

3. 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?
   1) (5%) What are the metrics that are commonly used to evaluate the performance of the classifier? Please at least list three of them.
   2) (5%) Based on the previous question, how do you determine the predicted class of each sample?
   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?
   1) There are many metrics to use to evaluate the performance of the classifier. For example:
      I. Accuracy: It measures the overall correctness of the classifier by calculating the ratio of correctly predicted instances to the total instances.
      II. Precision: It assesses the accuracy of positive predictions by calculating the ratio of true positive predictions to the total positive predictions.
      III. F1 score: It combines precision and recall into a single metric, providing a balance between the two.
   2) The predicted class of each sample is the class with the highest probability (Pi) output by the logistic regression model.
   3) In such a situation, Accuracy might not be a good choice because the classifier might perform well on the majority class but perform bad on minority classes, yet still get a high accuracy.
      There are some ways to deal with such a problem, we could use Precision and F1 score like what I mentioned above. Also, we could consider sampling the majority classes more or combining the minority classes. Last but not least, we could also analyze the confusion matrix to understand where the classifier is making errors, like misclassifying monirity 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.)

4.1. (10%)

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

The partial derivative of x to this function is:

-t/σ(x)-(1-t)/(1-σ(x))

4.2. (10%)

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

First, let L be equal to this function, z=σ(x), by chain rule, we could know that:

dL/dx = dL/dz * dz/dx

Also, we can get dL/dz=-2(t-σ(x)) by applying the chain rule.

And dz/dx is σ(x)(1-σ(x))

So the final answer is -2(t-σ(x))σ(x)(1-σ(x)).