NYCU Introduction to Machine Learning, Homework 2

109700046, 侯均頲

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%) Linear Classification Model - 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.

```
LR = LogisticRegression(learning_rate = 0.000047, iteration = 300000)
```

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

```
Weights: [-0.05382903 -0.73463643  0.896919   -0.0452799   0.02810365 -0.54045897],
 Intercept: -0.0651222946858563
```

3. (10%) Show the accuracy score of your model on the testing set.

```
Accuracy: 0.7540983606557377
```

(35%) Linear Classification Model - 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 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:
[[  4117.6432951  -21143.89414881]
 [-21143.89414881 108572.84804343]]
```

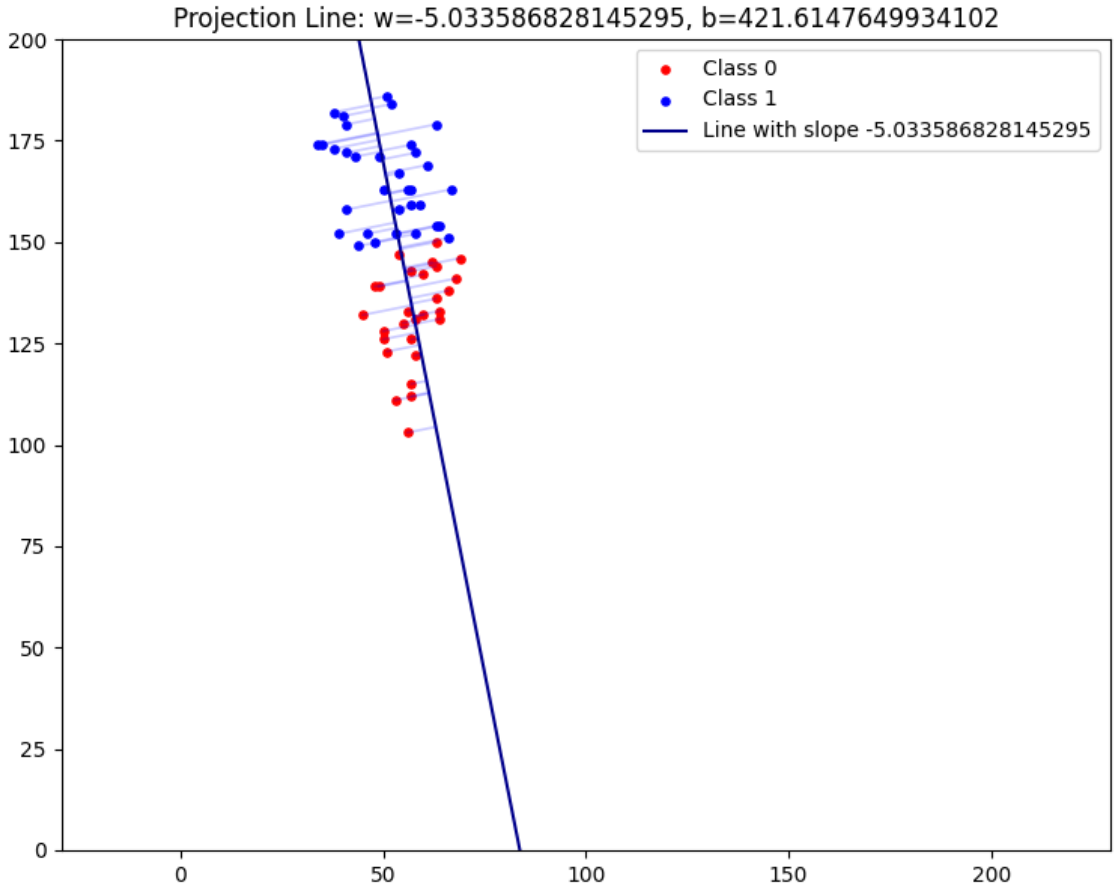
7. (5%) Show the Fisher's linear discriminant w of the training set.

```
w:
[-0.19485739  0.98083158]
```

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. `Accuracy of FLD: 0.672131147540983`

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.



10.

Part. 2, Questions (50%):

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

   In short, sigmoid function gives out a number between 0 and 1, softmax function gives out a smooth numbers distribution (looks like probability distribution) summing up equals 1, e ach sigmoid output is independent, but numbers in softmax output are not. For scenarios cl asses are mutual exclusive to each other we can use softmax (multiclass), or we use sigmoi d if there can be multi labels.

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

   This is a linear classification problem. MSE calculate distance between prediction value an d ground truth value, cross-entropy loss calculates how well the predicted probabilities of a model match the true probabilities of the actual distribution. Since we use sigmoid functio n, the gradient become closer and closer to zero as inputs go positive infinite and negative infinite, it means MSE will be meaningless as input of sigmoid function goes far away fro m 0 since it will be very small no matter if we are close to ground truth or not. Cross-entro py loss is more sensitive to predictions that are confidently wrong, and more reactive in ab

ove case, so it is more suitable than MSE. And Cross-entropy loss is just design for probab ility distribution related error measure usage which matches the output format of sigmoid f unction.

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 ov erall 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.

   Accuracy, Recall (Sensitivity), F1 score, Precision

2. (5%) Based on the previous question, how do you determine the predicted class of each sample?

   We pass each sample into our model and get P1, P2, ... Pc, than we predict the sample as class t where Pt is the highest value of all probabilities.

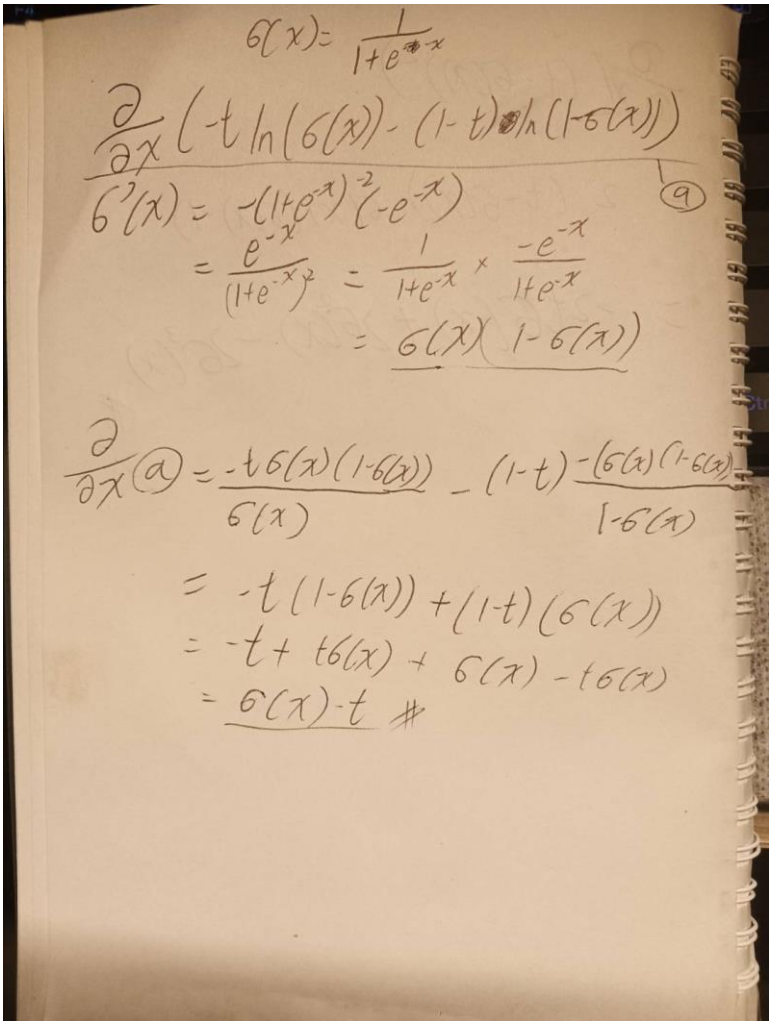
3. (5%) In a class imbalance dataset (say 90% of class-1, 9% of class-2, and 1% of cl ass-3), is there any problem with using the metrics you mentioned above and how t o evaluate the model prediction performance in a fair manner?

   All metrics I mentioned above could be biased toward majority class, since we just predict that all classes are class-1 than we get 90% accuracy, we ca n use method like balanced accuracy, which will take the sensitivity of eac h class into account and finally divided by number of 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 sigmo id function. σ is the sigmoid function.)

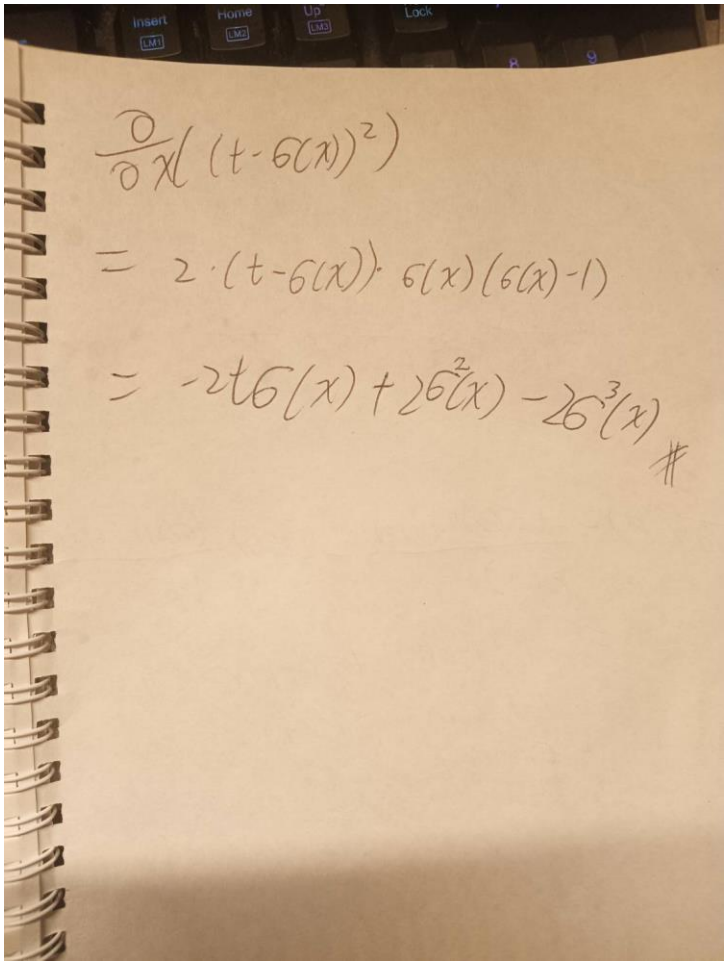
1. (10%)

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



2. (10%)

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



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

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

$$= -2t\sigma(x)+2\sigma^2(x)-2\sigma^3(x)$$ #