

Each small bar of ROC curve is rectangular or trapezoidal. Either way, we can use to calculate the area, where top is , bottom is and the height is .

But in fact, we can obtain another formula in the following way.

Consider a classification algorithm that assigns to a random observation a score (or probability) signifying membership in class 1. If the final classification between class 1 and class 0 is determined by a decision threshold , then we can denote TPR and FPR as follows:

and means x belong to class 1 and x belong to class 0, respectively. If we view as a function of , we can get

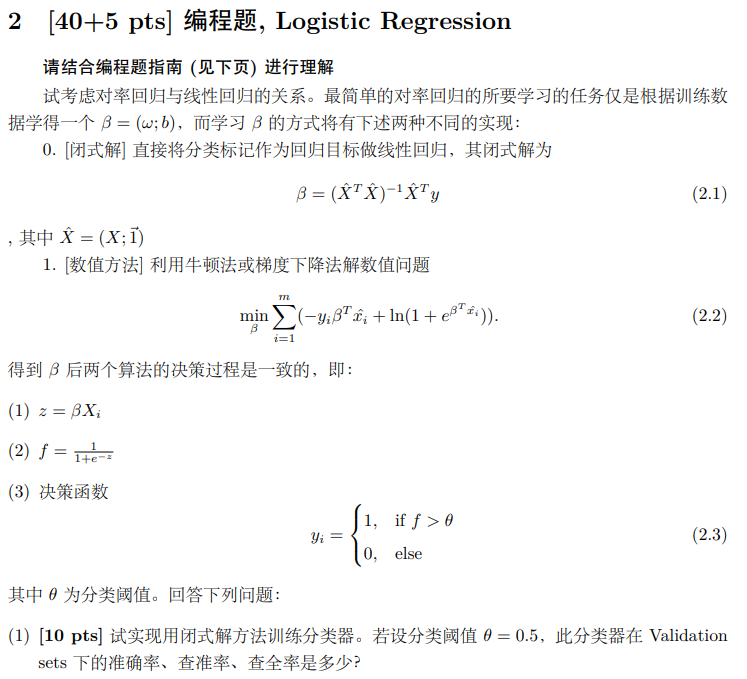
At the fourth equal sign, we used the fact that the probability density functionis the derivative with respect to of the cumulative distribution function .

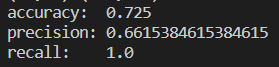
So given a randomly chosen observation belonging to class 1 , and randomly chosen observation belonging the class 0, the AUC is the probability that the evaluated classification algorithm will assign a higher score to than to .

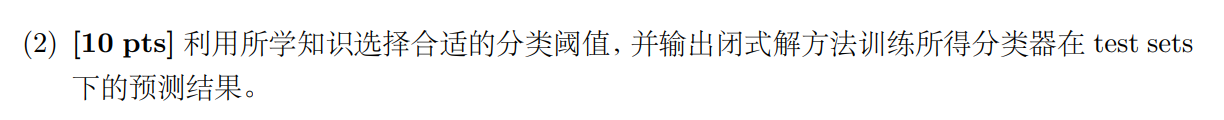
According this idea, assuming that the data set has a total of positive samples and negative samples, the predicted value is . We sort all the samples according to the predicted value in **increasing order**, and sort the numbers from to .

* For the sample with the highest probability of positive sample (assuming the sort number is), the number of negative samples with a smaller probability than it is .
* For the sample with the second highest probability of positive samples (assuming the sorting number is ), the number of negative samples with a lower probability than it is
* And so on…
* For the smallest positive sample probability, assume that the sorting number is , the number of negative samples with a smaller probability than it is

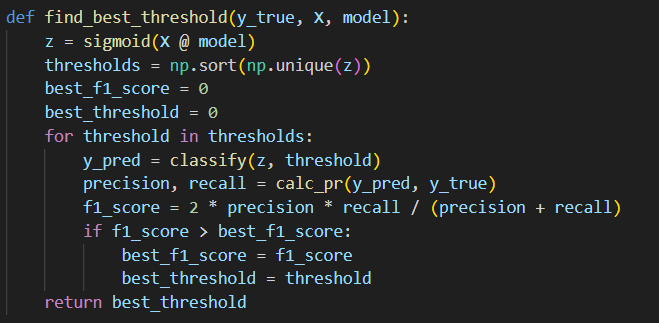
Then in all cases, the number of positive sample scores greater than negative samples is , then can be written as

. 

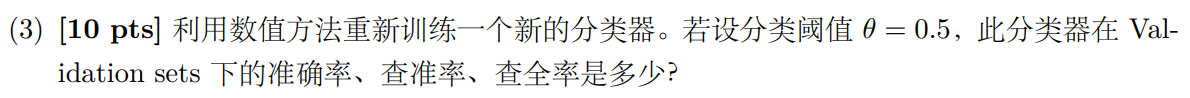




We can obtain the most suitable threshold by calculating the F1 score, the specific method is as follows:



Then, we can get the best threshold is 0.5139247728112848(In fact, 0.7192036272932429 is ok either. My point is there are very many values with the same maximum F1 score). By using this threshold, we just need to train on test set, predict and save data to a .csv file.

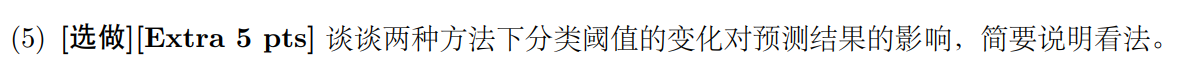


We use as the objective function of the optimization problem:

Then take the derivative to get the recursive formula of gradient decent method:

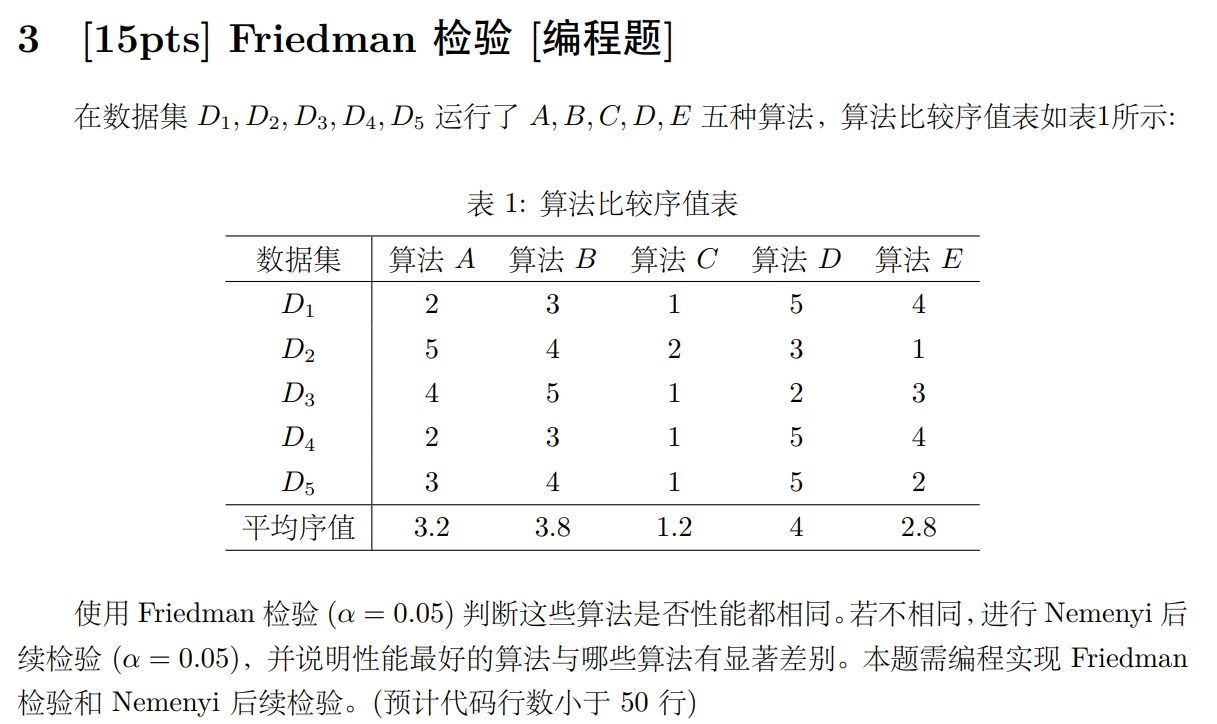
For the learning rate, we can simply set it to 0.05. The result is shown as follows:





The meaning of the closed-form solution is to predict the mark as a regression value, and the value of is approximated, not the value of . Therefore, after obtaining the regression value, any activation function can be used to complete the decision, such as tanh, sigmoid . Since the value of z is more inclined to be between [0,1] after the mark is used as the regression value, and the sigmoid function is used (note: the sigmoid function is not a gain operation in this problem), the gradient is large near z=0, That is, the disturbance caused by small errors will have a greater impact on the sigmoid output, so this method requires a more accurate threshold.

The numerical method approximates the rate value , and the range of z is not only between [0,1], but the gradient of the sigmoid function is small when z is far away from 0, that is, the small error has little influence on the output of the sigmoid. Therefore a very precise threshold is not required.

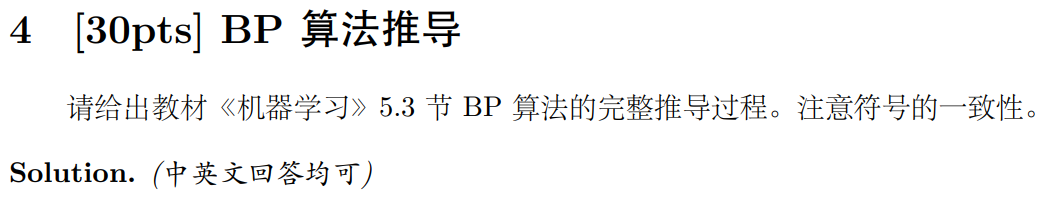


For Friedman test, we should calculate , and

where are the number of data sets, the number of algorithms and the mean rank of algorithm, respectively.

We can obtain , which means we need to turn down “all algorithms have the same performance”. Therefore, we have to do Nemenyi test next.

From and , we get . And , so there is a significant difference between the performance of algorithms C and D, but there is no significant difference between C and other algorithms.



Prove that

For formula (1), we have

Then,

So

For formula (2), we have

Then,

So

For formula (3), we have

Then,

So

For formula (4), we have

Then,

So