preliminary results report

Huyunting Huang

In this stage, I use the first 2000 samples of training set to check if the function and algorithm I wrote can work. I use the kernel SVM to train the data with hyper parameter beta, and k-fold cross validation to do the cross-validation and here k=5, and beta starts with 0.02 and then goes from 0.04 to 3 with interval 0.04. The loss-function here is the 0-1 loss function

$$\text{Loss}=\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{i,true}-y_{i,pred}}{2}\right|\right)$$

where $y_{i,true}$ is the true label of the i-th sample and $y_{i,pred}$ is the predicted label of the i-th sample. I choose beta with the lowest 0-1 loss in CV as my theta and then predict the label in test set.

In training stage, The average 0-1 loss(of CV)—Beta plot is shown in Fig.1

Just as the figure shown the 0-1 loss will first go up and then slightly go down and remain stable. Which means a possible 'good' range (which can get the minimum 0-1 Loss in CV) of beta should be $\leqslant$ 0.02.

And the final result in testing is loss=0.647 which means over a half of the prediction is wrong.

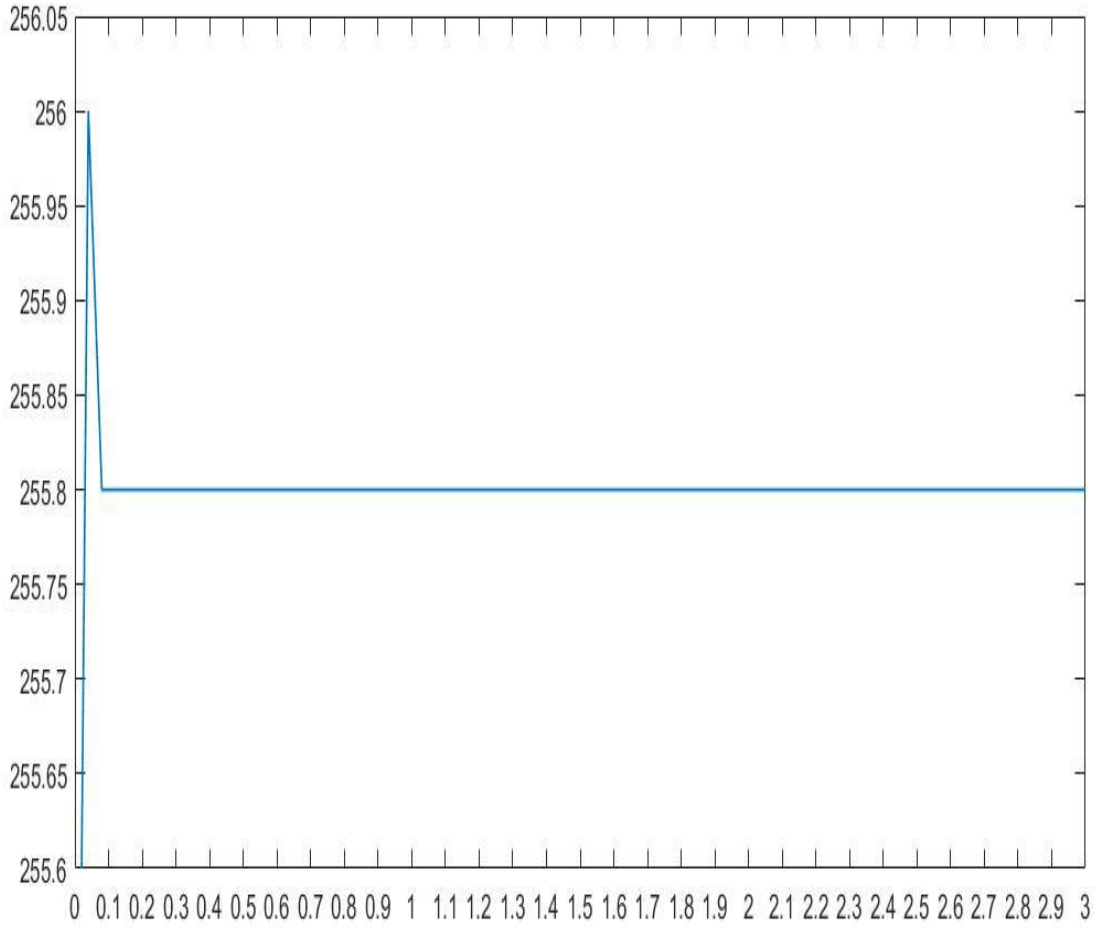

Fig.1 0-1—beta plot

Because of the result I got. Next I'll test other beta, and also try other training method such as linear SVM and I'll also increase the number of sample in training and increase k to 10. Also the loss function will changed into AMS metric shown in the original website(the URL is in project Plan)