

EE6550 Machine Learning, Spring 2016

Homework Assignment #4 Report

102061210 王尊玄

1. Iris_set_ver :

This data set is pretty small, so I didn't apply cross validation over the set. For both kernel, T better not to be too big, or more specifically, 1 is good enough. Also for Gaussian kernel, free parameter sigma better not be too big, approximately smaller than 1. If you follow advices above, you will simply 100% predict over testing data.

2. adult :

Since it takes very long to do cross validation over entire adult training set, I restrict the size of sample to 3000. Cross validation errors $\hat{R}_{CV}(\theta)$ are given in the following tables.

adult cross validation error: 5-fold, gaussian kernel				
sigma\T	1	2	3	5
0.1	0.229667	0.230333	0.231667	0.232333
0.4	0.232	0.228	0.23	0.227333
1.1	0.238	0.241	0.240667	0.241

adult cross validation error: 10-fold, gaussian kernel				
sigma\T	1	2	3	5
0.1	0.229333	0.226667	0.225667	0.227
0.4	0.227667	0.226667	0.227667	0.225667
1.1	0.240333	0.242	0.242	0.242333

adult cross validation error: 5-fold, linear kernel				
T	1	2	3	5
error	0.258667	0.258667	0.258667	0.258667

adult cross validation error: 10-fold, linear kernel				
T	1	2	3	5
error	0.258667	0.258667	0.258667	0.258667

First, Gaussian kernel perform much better than linear kernel, so I

choose using Gaussian kernel. Minimum cross validation $\hat{R}_{CV}(\theta)$ of 5-fold and 10-fold cross validation are found at $(T, \sigma) = (5, 0.4)$ and $(3, 0.1)$. With smaller size of training set 3000, I get performance of prediction over entire testing data as follows.

training set size = 3000	accuracy
T=5, Gaussian, sig=0.4	79.38%
T=3, Gaussian, sig=0.1	79.26%
T=1, linear	75.43%

Finally, I try using entire training data set to train my kernel perceptron.

entire training set	accuracy
T=5, Gaussian, sig=0.4	79.63%
T=3, Gaussian, sig=0.1	79.42%

The above hypotheses can be found in precomputed_h.mat and precomputed_h2.mat.