EE6550 Machine Learning, Spring 2016

Homework Assignment #2 User manual

102061210 王尊玄

1. Program main.m is to run cross validation and SVM training algorithm over training sample. There are several parameters that can be adjusted, which are all at the top of my code.
   - data: choose which data to deal with, 1 for adult, 2 for iris_set_ver.
   - mode: choose which kernel to be used, 1 for inner product kernel, 2 for Gaussain kernel.
   - tol: tolerance over SVM training algorithm (used in checking KKT).
   - C_test: a vector containing different slack penalty coefficients used in cross validation.
   - Sig_test: a vector containing different variance of Gaussian kernel used in cross validation.
   - miniSize: size of each mini-set.
   - numMiniSet: number of mini-sets; miniSize×numMiniSet=size of training data in cross validation.
   - maxIter: maximum iteration as a crude terminate condition to make sure my SVM training algorithm terminate in certain period of time.
2. After running main.m, which gives us best choice of free parameters C and sigma. We then can run finalTrain.m to train SVM classifier over entire training sample (adult_training) with determined C and sigma. After probably a extremely long period of time, we get Lagrange multipliers λ and shifting value b, and we can predict over testing data and estimate correctness of our SVM classifier.
3. A set of parameter trained with 10000 samples can be obtained in file best.m. Following instructions produce correctness of a SVM classifier over testing data.

```
>> load best
>> load adult_testing
>> x_test = adult_testing(:,1:end-1);
>> y_test = adult_testing(:,end);
>> y_predict = svmPredict(x,y,x_test,alpha,b,sig);
>> sum(y_predict==y_test)./length(y_test)
```