## EE6550 Machine Learning, Spring 2016

## Homework Assignment #4 User Manual

102061210 王尊玄

1. Main routine runs over all process including n-fold cross validation to find optimal free parameters  $\theta$  and applying kernel perceptron algorithm with  $\theta$  over entire training set, and finally predict on testing data and estimate performance of my kernel perceptron algorithm. Note that the final training over entire set is not actually entire set, because I restrict the size of sample to 3000 for not consuming too much time. You can do "actual" entire set training by following commands.

Where T and sig are free parameter and can be chosen as your will. Also, (T, sig) = (5, 0.4) or (3, 0.1) are optimal free parameter pair found by my cross validation.

- 2. Parameters (can be found in the top of main program):
  - whichData: 1 for adult, 2 for iris set ver
  - nCross: in n-fold cross validation, nCross = n, may be 5 or 10.
  - TChoice: multiples of sample size, which is T in kernel perceptron algorithm, better to be an integer. The vector is defaulted to [1 2 3 5].
  - sigChoice: sigma used in Gaussian kernel, defaulted to [0.1 0.4 1.1], better not too big.
  - kernelChoice: 1 for linear kernel, 2 for Gaussian kernel.
- 3. Instead of running overall process with cross validation, you can merely use precomputed hypotheses to obtain performance on testing data, which can be done typing following command in command window.

```
>> load adult_testing
>> x_test = adult_testing(:,1:end-1);
>> y_test = adult_testing(:,end);
```

```
>> load precomputed_h.mat
>> y_predict = kernelPerceptron_predict(x_test,x,y,kernel,bestAlpha);
>>fprintf('Accuracy=%.3f%%\n',sum(y_predict==y_test)/length(y_test)*100);
```

, where precomputed-h.mat can also be precomputed\_h2.mat.