Homework Assignment 3

March 9, 2017
Due at March 16, 2017

This homework requires you to implement multi-class SVM classification based on LibSVM package and solve a practical multi-class classification problem. The LibSVM package can be freely downloaded from `https://www.csie.ntu.edu.tw/~cjlin/libsvm/`

1 Requirement of Homwwork 3

a. Implement traditional one-versus-one and one-versus-rest task decomposition methods to solve a multi-class probelm mentioned below.

b. Implement part-versus-part task decomposition method to solve the same multi-class problem.

c. Use two different kernel functions, namely linear and RBF, in all your classifiers.

d. Compare the advantages and disadvantages of these three task decomposition methods.

2 Data for Homework 3

The training data and test data are in directory `~/HW_Public/HW3/data/` on FTP.

The datasets (train.txt, test.txt) contains protein sequences from 12 subcellular locations. 20 dimensions stand for 20 amino acid composition of the protein sequences.

- Number of proteins : 7579. 6065 training samples, 1514 test samples.
- Number of classes: 12 (label of {0,1,...,11}).

The data file are of the format below:

label	$dim1 : value$	$dim2 : value$	...	$dim20 : value$
0	1 : 0.095861	2 : 0.010893	...	20 : 0.032680