

**CS678A: Learning with kernels**  
**Assignment #2: SVMs and Optimization**

Due on: 11 Oct. 2015, 23.00  
MM: 400

30 Sep. 2015

Preamble:

In this assignment you are free to use any platform and/or library. Some of these are Java, Python (with numpy, scipy, scikit-learn, pyML etc.), C++, Octave, R, Scilab, Matlab. Open source platforms preferred.

In addition you will need the LIBSVM and LIBLINEAR packages available at:

<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

<https://www.csie.ntu.edu.tw/~cjlin/liblinear/>.

You will work with 3 data sets available on the LIBSVM datasets page.

<https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/>

i) Leukemia, ii) covtype.binary, iii) rcv1.binary.

They are data sets where the number of features and instances are related in different ways:  $\#features \gg \#instances$ ,  $\#features \ll \#instances$  and both  $\#features$  and  $\#instances$  is large.

1. Using any QP solver implement the chunking and SMO algorithms for SVMs. Run your implementations on the above data sets and compare the classification performance of both for each data set and the time taken. Study how the different stopping criteria behave.

(250)

2. Use LIBSVM and LIBLINEAR libraries and find the classification performance with linear, polynomial (low degree) and RBF kernels in each case. Also, compare the time taken in each case.

Remember to tune parameters. Also, try both with and without scaling data and record the difference in performance in each case.

(150)