Homework 3

CS420 Machine learning 2018 Spring*
Department of Computer Science and Engineering
Shanghai Jiao Tong University

Submission deadline: 20:00, June 7, 2018, Thursday

Submission to:

Please submit your homework in pdf format to the CS420 folder in the following FTP. File name should be like this: `015033910032_chenyajing_hw1.pdf`.

ftp://public.sjtu.edu.cn
username: cyj907
password: public

1 (30 points + 10 bonus) SVM vs. Neural Networks

(30 points) Select at least two data sets from the link (Data A) below, and then investigate classification performances of Support Vector Machine (SVM) and neural networks (e.g., MLP) on the selected data sets. You may try different experimental settings, e.g., varying the sample size of the training set, trying data sets with different dimensions, and other configurations that may affect the performance in your mind. You may also try different kernels for SVM.

(10 bonus points) Select at least one data set from the link (Data B) below, and applied SVM on it. Compare the SVM performance with the deep learning algorithm benchmarks (if any, by searching the literature on the selected data set). Discuss the strengths and weaknesses of SVM on big data sets. (Do not select the MNIST data set which has been reserved for our course project.)

Links to the data sets:

- Data A: `https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/`
- Data B: `http://deeplearning.net/datasets/`

Links to the codes for your references:

- SVM: `https://www.csie.ntu.edu.tw/~cjlin/libsvm/`
- SVM: `http://scikit-learn.org/stable/modules/svm.html`
- MLP: `http://scikit-learn.org/stable/modules/neural_networks_supervised.html`

*tushikui@sjtu.edu.cn