# Assignment-1 (CS-729)

1-Mar-2013

Pick any one problem from the following and answer it. You may answer more than one. You may refer any book/notes/web-resource etc. But you should NOT discuss with your classmate. You may contact me for clarifications. You are free to use any optimization toolbox[1] and/or any SVM toolbox[2]

1. Recall that we motivated the SVM through learning theory. In particular, we showed a hierarchy of function classes parameterized by $W$ (margin) and then derived a bound on the true risk (that holds with high probability) involving the radius-margin bound on the Rademacher average. This bound motivated the SVM formulation for ERM and choice of $W$ is to be decided by SRM. Given this background, consider formulation (5) in [1][3]. Now, provide an analogous motivation for this formulation i.e., provide the corresponding hierarchy of function class and the corresponding bound on Rademacher average etc. for this case. After doing this, generate a synthetic "ellipsoidal" binary dataset and then compare the performance of SRM employing i) SVM and ii) formulation (5).

2. In lectures we saw four types of linear function classes: i) one with bounded 2-norm ii) $\infty$-norm iii) 1-norm iv) no bound. By generating diff. synthetic datasets with various dimensions and employing the above four types of functions show the following:

[1] e.g., `cvx` available at `http://cvxr.com/cvx`

[2] e.g., `libsvm` available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

[3] Available at `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.214.3106`

(a) i) and iii) have no curse of dimensionality

(b) ii) and iv) may have curse of dimensionality

3. In the context of the above problem empirically demonstrate:

   (a) ERM candidate with iv) is very sensitive to perturbations in training data; while that with i) is not.

   (b) ERM candidate with ii) tends to have equal entries for all dim.; whereas that with iii) tends to have many entries as zero.

4. Run simulations on synthetic data of fixed dimension to compare SRM with i) and ERM with iv) in the above problem. After this, compare SRM with i) and SRM with iv). Repeat the analysis with synthetic data that includes few random noisy features.

5. Using synthetic datasets compare the performance of the following two model selection methods: i) SRM using the radius-margin bound ii) SRM with leave-one-out cross-validation[4]. In both cases, use SVM for ERM. Can you intuitively explain why one is expected to perform better than the other? Also, demonstrate the two kinds of overfittings in the leave-one-out case mentioned in lectures: overfitting at ERM level and overfitting at validation level.

# References

[1] Pannagadatta K. Shivaswamy and Tony Jebara. Ellipsoidal Kernel Machines. In Proceedings of Artificial Intelligence and Statistics, 2007.

[4] Derive a corresponding expression for guaranteed risk in this case.