STAT 535
Lecture VI
Support Vector Machines

©Marina Meilă
mmp@stat.washington.edu

Reading: Murhpy Ch 14 (14.1,14.2–14.2.4 kernels, 14.4 and equations (14.28,14.29) kernel trick, 14.5.1.–3 Support Vector Machines) Additional Reading: C. Burges - "A tutorial on SVM for pattern recognition"

These notes: Section 2,3 (convex optimization) are optional.

A simple error bound

$$L_{01}(f_N) \;\leq\; E\left[\frac{\#\text{support vectors of} f_{N+1}}{N+1}\right] \tag{1}$$

where f_N denotes the SVM trained on a sample of size N.

Proof Sample $N+1$ points from P_{XY}, use x^1 for testing and the remaining N for training a SVM denoted by f_N; L_{01} is the expected probability of error on the x^1.

Now, with the same points, train a SVM f_{N+1} on all the $N+1$ points. If x^1 is not a support vector, then the SVM will be the same as f_N , and x^1 will be correctly classified (by the definition of "not a support vector"). If x^1 is a support vector, then it may or may not be correctly classified by f_N. Thus, the probability of classifying x^1 correctly is at least as large as the probability of x^1 not being a s.v. in f_{N+1}. Since x^1 can be any of the s.v. of f_{N+1}, the result follows.