Foundations of Machine Learning
Department of Computer Science, NYU
Homework assignment 2
Due: March 1, 2005

1. Shattering coefficients of intervals in $\mathbb{R}$.

   Let C be the set of intervals in $\mathbb{R}$. The VC dimension of H is 2 (shown in class). Compute its shattering coefficient $\Pi_C(m)$, $m \geq 0$. Compare your result with the general bound for shattering coefficients.

2. Bound on shattering coefficients.

   Prove that the general bound given for shattering coefficients is tight: for any set X of $m > d$ elements, there exists a concept class C of VC dimension d such that $\Pi_C(m) = \sum_{i=0}^{d} \binom{m}{i}$.

3. VC dimension of a vector space of real functions.

   Let F be a finite-dimensional vector space of real functions on $\mathbb{R}^n$, $dim(F) = r < \infty$. Let H be the set of hypotheses:

   $$H = \{\{x : f(x) \geq 0\} : f \in F\}.$$

   Show that d, the VC dimension of H, is finite and that $d \leq r$ [*Hint*: select an arbitrary set of $m = r+1$ points and consider linear mapping $u : F \to \mathbb{R}^m$ defined by: $u(f) = (f(x_1), \ldots, f(x_m))$].

4. Soft margin hyperplanes

   The function of the slack variables used in the optimization problem for soft margin hyperplanes had the form: $\xi \mapsto \sum_{i=1}^{m} \xi_i$. Instead, we could use $\xi \mapsto (\sum_{i=1}^{m} \xi_i)^k$, with $k > 1$. Show that the solution can still be defined in terms of *support vectors*. Give the dual formulation of the problem.

5. SVM classification

   - Randomly generate 5,000 examples from each of two Gaussian distributions with the same diagonal covariance matrix αId_{50}, $\alpha = 50$, centered one $(-1, \ldots, -1)$ and $(1, \ldots, 1)$ in $\mathbb{R}^{50}$. Examples from the first distribution are labeled with -1, others with $+1$. Split the data into two equal sets of size 5,000, training and test [*hint*: the examples have been already generated for you. Download the files "negative.dat" and "positive.dat" from the usual place].

- Download and install the libsvm software library from:

  `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

- Select a family of kernels, polynomial or Gaussian. Use SVM classification for training set sizes 1,000, 2,000, 3,000, 4,000 and 5,000, and find the kernel parameters that minimize the test error. Plot the test error as a function of the parameters and mark the minimum.
- For a fixed parameter setting, plot the training and test errors as a function of the training set size and estimate the asymptotic test error.
- Compute the *Bayes error*, i.e., the error of the best possible classifier by using the density functions of the two classes. Compare with the results obtained using SVMs.