Iowa State University
Department of Computer Science
Machine Learning (Com S 573)
Vasant Honavar
Spring 2007
Due April 16 2007
Problem Sets 5 and 6

1. (**25 pts.**) Recall that the standard Support Vector Machine classifier finds a maximum margin separating hyperplane assuming equal misclassification cost for both positive and negative instances. Consider the problem of learning binary classifiers in a setting where the costs of misclassification of positive instances and negative instances are unequal. Precisely formulate the problem of learning Support Vector Machine Classifier to minimize the cost of misclassification when the cost of misclassification for the positive and negative instances are specified by the user and derive an algorithm for solving the resulting learning problem.

2. (**25 pts.**) Rigorously prove a mistake bound for the balanced winnow algorithm discussed in class.

3. (**25 pts.**) Consider the problem of learning a conjunctive concept in the PAC setting. We have seen a consistent learner for this task and have derived its sample complexity. The consistent learner we used does not exploit the information provided by the negative examples. Consider a two-phase algorithm for learning conjunctive concepts in the first phase, applies the algorithm described in class for learning conjunctions and in the second phase, keeps only those literals that are necessary to maintain consistency of the conjunctive concept with the negative examples. Such literals can be identified using a greedy set cover of the negative examples using a subset of the literals in the conjunctive concept produced by the first phase. Derive the sample complexity of the resulting 2-phase algorithm when learning an arbitrary target conjunctive concept c of size $size(c)$ defined over an n-variable boolean instance space and show that it is significantly better than the sample complexity of the 1-phase algorithm.

4. (**25 pts.**) Using Chernoff bounds, show that the empirical error of a hypothesis h estimated from a set S of examples (sampled within the PAC setting) is accurate to within an additive factor $\gamma/2$ of the true error of h with confidence at least $1-\delta/2k$ provided the number of examples m in S satisfies $m \geq \frac{c_0}{\gamma^2} log \frac{2k}{\delta}$ for an appropriate constant $c_0 > 0$. (This result can be used to show that a weak learner (with confidence worse than $(1-\delta)$ can be turned into a strong learner (with confidence at least $(1-\delta)$.

5. (**25 pts.**) **Rigorously prove the sample complexity of an Occam learning algorithm.**

6. (**25 pts.**) **In each base learner is trained on iid samples, and is correct with confidence $> 1/2$, how many base learners do we need to use in an ensemble to ensure that the (unweighted) majority of the classifiers is correct with confidence $1-\delta$ for some arbitrary δ such that $0 < \delta < 1/2$?**