Mehryar Mohri
Foundations of Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 1
Due: February 17, 2009

A. Concentration Bounds

1. Let X be a non-negative random variable verifying $\Pr[X > t] \leq ce^{-2mt^2}$ for all $t > 0$ and some $c > 0$. Show that $\mathrm{E}[X^2] \leq \frac{\log(ce)}{2m}$ [*Hint*: to do that, use the identity $\mathrm{E}[X^2] = \int_0^\infty \Pr[X^2 > t]dt$, write $\int_0^\infty = \int_0^u + \int_u^\infty$, bound the first term by u and find the best u to minimize the upper bound].

2. Let $S = (X_1, \ldots, X_m)$ be a sample of size m. Consider the function Φ defined by $\Phi(X_1, \ldots, X_m) = \sup_{h \in H} |error(h) - \widehat{error}(h)|$. Apply McDiarmid's inequality to Φ. Does the result depend on the VC dimension of H?

B. PAC Learning

1. Show that equilateral triangles with a base parallel to the X-axis are PAC-learnable. Give an algorithm and careful justifications using the proof given in class. What is the VC dimension of this concept class?

2. Give a PAC-learning algorithm for the subsets of the real line formed by the union of k intervals. What is the VC dimension of this concept class?

C. VC dimension

1. Show that the VC dimension of a finite hypothesis set H is at most $\log_2 |H|$.

2. What is the VC-dimension of the set of subsets I_α of the real line parameterized by a single parameter α: $I_\alpha = [\alpha, \alpha + 1] \cup [\alpha + 2, +\infty[$.

3. What is the VC dimension of the set of all ellipsoids in $\mathbb{R}^n$?