A. Concentration Bounds

1. [10 points] 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 $E[X^2] \leq \frac{\log(c) e}{2m}$.

[Hint: to do that, use the identity $E[X^2] = \int_0^\infty \Pr[X^2 > t]dt$, 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].

$$
E[X^2] = \int_0^u \Pr[X^2 > t]dt + \int_u^{\infty} \Pr[X^2 > t]dt
= \int_u^{\infty} e^{-2mt} dt
= e^{-2mu} = g(u).
$$

Function $g$ reaches its minimum for

$$
g'(u) \geq 0 \iff 1 - ce^{-2mu} \geq 0
\iff -2mu \leq -\log c
\iff u \geq u_0 = \frac{\log c}{2m},
$$

and $g(u_0) = \frac{\log(c) e}{2m}$.

2. [15 points] Let $S = (X_1, \ldots, X_m)$ be a sample of size $m$. Consider the function $\Phi$ defined by $\Phi(X_1, \ldots, X_m) = \sup_{h \in H}|error(h) - \widehat{error}(h)|$. Apply McDiarmid’s inequality to $\Phi$. Does the result depend on the VC dimension of $H$?

Changing $X_i$ into $X_i'$ affect $\Phi$ by at most $1/m$. Thus, by McDiarmid’s inequality,

$$
\Pr[\sup_{h \in H}|error(h) - \widehat{error}(h)| - E[\Phi] > \epsilon] \leq 2e^{-2m\epsilon^2}.
$$

Setting $\delta$ to match the right-hand side, this yields the following generalization bound: for any $\delta > 0$, with probability at least $1 - \delta$, for all $h \in H$,

$$
|error(h) - \widehat{error}(h)| \leq E[\Phi] + \frac{\log(2/\delta)}{2m}.
$$
The VC dimension of $H$ does not appear directly, but $E[\Phi]$ depends on the complexity of the class $H$, which can be measured in terms of the VC dimension.

B. PAC Learning

1. [25 points] 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?

One PAC algorithm consists of choosing the tightest equilateral triangle containing the positive sample points. The proof schema is the same as for the example treated in class except that here three regions can be selected along each side of the triangle. The sample complexity is then

$$m \geq \frac{3}{\epsilon} \log \frac{3}{\delta}. \quad (9)$$

It is not hard to see that the VC dimension of the class is 3. You can compare this bound with the VC bound and lower bound shown in class.

2. [25 points] 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?

This is similar to the previous problem. One PAC-learning algorithm would consist of returning the union of intervals, where each interval is defined to (just) contain a longest sequence of consecutive positive sample points. There can be at most $2^k$ such intervals. The proof is similar to the one in class except that $2^k$ regions bordering the intervals can be considered here. The sample complexity can be shown to be

$$m \geq \frac{2k}{\epsilon} \log \frac{2k}{\delta}. \quad (10)$$

The VC dimension of this class is $2k$ (see solution of HW2, 2007).

C. VC dimension

1. [5 points] Show that the VC dimension of a finite hypothesis set $H$ is at most $\log_2 |H|$. With a finite set $H$, at most $2^{|H|}$ dichotomies can be defined.
2. [20 points] What is the VC-dimension of the set of subsets $I_\alpha$ of the real line parameterized by a single parameter $\alpha$: $I_\alpha = [\alpha, \alpha+1] \cup [\alpha+2, +\infty]$. The set of three points $\{0, 3/4, 3/2\}$ can be fully shattered as follows:

$$
\begin{align*}
+ &+ + \quad \alpha = -2 \\
+ &+ - \quad \alpha = 0 \\
+ &- + \quad \alpha = -1 \\
+ &- - \quad \alpha = 3/2 - 2 + \epsilon \\
- &+ + \quad \alpha = 3/4 - 2 \\
- &- + \quad \alpha = \epsilon \\
- &- - \quad \alpha = 3/2 \\
- &- - \quad \alpha = 3/2 + \epsilon,
\end{align*}
$$

where $\epsilon$ is a small number, e.g., $\epsilon = .1$. No set of four points $x_1 < x_2 < x_3 < x_4$ can be labeled by $+ --$. This is because the three leftmost labels $+ --$ imply that $\alpha + 2 \leq x_3$ and thus also $\alpha + 2 < x_4$. Thus, the VC dimension of the set of subsets $I_\alpha$ is 3. Note that this does not coincide with the number of parameters used to describe the class.

3. [20 (bonus) points] What is the VC dimension of the set of all ellipsoids in $\mathbb{R}^n$?

The general equation of ellipsoids in $\mathbb{R}^n$ is

$$
(X - X_0)^\top A(X - X_0) \leq 1, \quad (11)
$$

where $X, X_0 \in \mathbb{R}^n$ and $A = (a_{ij}) \in S^n_+$ is a positive semi-definite symmetric matrix. This can be rewritten as

$$
X^\top AX - 2X^\top AX_0 + X_0^\top AX_0 \leq 1, \quad (12)
$$

or $\sum_{i,j=1}^n 2a_{ij}(x_i x_j + x_j x_i) - \sum_{i=1}^n a_{ii} x_i + (X_0^\top AX_0 - 1) \leq 0$ using the fact that $A$ is symmetric. Let $a_i = -2(AX_0)_i$ for $i \in [1, n]$ and let $b = X_0^\top AX_0 - 1$. Then this can be viewed in terms of the following equations of hyperplanes in $\mathbb{R}^{n(n+1)/2 + n}$

$$
W^\top Z + b \leq 0, \quad (13)
$$

with

$$
W = \begin{bmatrix}
a_1 \\
\vdots \\
a_n \\
a_{11} \\
\vdots \\
a_{ij} \\
\vdots \\
a_{nn}
\end{bmatrix}, \quad Z = \begin{bmatrix}
x_1 \\
\vdots \\
x_n \\
x_1 x_1 + x_1 x_1 \\
\vdots \\
x_1 x_j + x_j x_i \\
\vdots \\
x_n x_n + x_n x_n
\end{bmatrix} \quad n(n+1)/2 + n
$$
Since the VC dimension of hyperplanes in $\mathbb{R}^{n(n+1)/2+n}$ is $n(n + 1)/2 + n + 1 = (n + 1)(n/2 + 1)$, the VC dimension of ellipsoids in $\mathbb{R}^n$ is bounded by $(n + 1)(n + 2)/2$. 