COMP8601: Advanced Topics in Theoretical Computer Science
Homework 4 **Due Date:** 5 Dec 2013

Rules: Discussion of the problems is permitted, but writing the assignment together is not (i.e. you are not allowed to see the actual pages of another student).

This homework has 125 points, of which 25 points are extra credit.

1. **(20 points) Using ϵ-Nets for Learning.** Suppose X is a set with some underlying distribution D and C is a class of boolean functions on X, and the VC-dimension of (X, C) is d. Moreover, suppose there is some function $f_0 \in C$ that corresponds to some classifier that we wish to learn. The model we have is that we can sample a random $x \in X$ and ask for the value $f_0(x)$. After seeing m such samples S in X, we pick a function $f_1 \in C$ that agrees with f_0 on S. The hope is that f_1 and f_0 would agree on most points in X (according to distribution D).

   (a) Define another class C' of boolean functions on X such that if S is an ϵ-net under C', and $f \in C$ is a function that disagrees with f_0 on more than ϵ fraction (weighted according to D) of points in X, then there exists some $x \in S$ such that $f(x) \neq f_0(x)$. Prove the VC-dimension of (X, C') for the class C' that you have constructed.

   (b) How many samples are enough such that with probability at least $1 - \delta$ the function f_1 returned disagrees with f_0 on at most ϵ weighted fraction of points in X?

2. **(10 points) Integral for Square Root of Logarithm.** Prove that for any constant $c \geq 1$, we have $\int_0^1 \sqrt{\log \frac{c}{x}} dx = \Theta(1)$.

3. **(40 points) McDiarmid's Inequality.** Let $\mathcal{O}$ be a set of objects. Let $X_1, X_2, \ldots, X_m$ be independent random objects taken from $\mathcal{O}$. Let $h : \mathcal{O}^m \mapsto \mathbb{R}$ be a function that satisfies the "c_i-Lipschitz" property: for all $i \in [m]$ and objects $x_1, x_2, \ldots, x_m, x_i' \in \mathcal{O}$,

$$|h(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m) - h(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_m)| \leq c_i.$$

   (a) (10 points) Let Y be a real random variable with $\mathbf{E}[Y] = 0$ such that $a \leq Y \leq b$. Let $\lambda \in \mathbb{R}$. We prove that $\mathbf{E}\left[e^{\lambda Y}\right] \leq e^{\frac{\lambda^2 (b-a)^2}{8}}$.

      i. Suppose $a < b$ and $p := -\frac{a}{b-a}$. Define function $L(z) := -pz + \ln(1 - p + pe^z)$ for $z \in \mathbb{R}$. Prove that $\mathbf{E}\left[e^{\lambda Y}\right] \leq e^{L(\lambda(b-a))}$.
      (Hint: The function $e^{\lambda z}$ with respect to z is convex.)

      ii. Prove that $L(\lambda(b-a)) \leq \frac{\lambda^2 (b-a)^2}{8}$.
      (Hint: If a function L is twice differentiable in $\mathbb{R}$, then for $y \in \mathbb{R}$ we have $L(y) = L(0) + L'(0)y + \frac{L''(z)}{2} y^2$ for some z between 0 and y.)

(b) (15 points) For $i \in \{0, 1, \ldots, m\}$ define $X^{(i)} := (X_1, \ldots, X_i)$, where $X^{(0)}$ can be considered as a random object independent of $X_1, \ldots, X_m$. Let $X := X^{(m)}$. Define $Z_i := \mathbf{E}\big[h(X)|X^{(i)}\big]$. Let $t > 0$. Prove that $\mathbf{E}\big[e^{t(Z_i - Z_{i-1})}|X^{(i-1)}\big] \leq e^{\frac{t^2 c_i^2}{8}}$ for $1 \leq i \leq m$.
(Hint: Consider the random variable $(Z_i - Z_{i-1})|X^{(i-1)}$. What is $\mathbf{E}\big[Z_i - Z_{i-1}|X^{(i-1)}\big]$?
Find a and b such that $a \leq (Z_i - Z_{i-1})|X^{(i-1)} \leq b$ and that $b - a \leq c_i$. Try $a = \inf_{x \in \mathcal{O}}(\mathbf{E}\big[h(X)|X^{(i-1)}, X_i = x\big] - \mathbf{E}\big[h(X)|X^{(i-1)}\big])$ and $b = \sup_{x \in \mathcal{O}}(\mathbf{E}\big[h(X)|X^{(i-1)}, X_i = x\big] - \mathbf{E}\big[h(X)|X^{(i-1)}\big])$.)

(c) (5 points) Prove that for $1 \leq i \leq m$, we have $\mathbf{E}\big[e^{t(Z_i - Z_0)}\big] \leq e^{\frac{t^2 c_i^2}{8}} \mathbf{E}\big[e^{t(Z_{i-1} - Z_0)}\big]$.
(Hint: For arbitrary functions f and g, we have $\mathbf{E}\big[\mathbf{E}\big[f(X^{(i-1)})g(X^{(i)})|X^{(i-1)}\big]\big] = \mathbf{E}\big[f(X^{(i-1)})\mathbf{E}\big[g(X^{(i)})|X^{(i-1)}\big]\big]$.)

(d) (10 points) Prove that for $\epsilon > 0$, we have $\Pr(h(X) - \mathbf{E}[h(X)] > \epsilon) \leq \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^m c_i^2}\right)$.

4. **(15 points) $\frac{1}{m}$-Lipschitz Property.** In this question we give some details on the proof of Lemma 1.4 in Lecture 9. Recall that given a set X of points and a class C of boolean functions on X, we denote by $S = (x_1, \ldots, x_m)$ a bag of points sampled from X. Let $M(S) := \sup_{f \in C}(\mathsf{Avg}_S[f] - \mathbf{E}_X[f])$, where $\mathsf{Avg}_S[f]$ is the fraction of points in S that takes value 1 under function f. Also let $R_S(C) := \mathbf{E}_\sigma\left[\sup_{f \in C} \frac{1}{m} \sum_{i=1}^m \sigma_i f(x_i)\right]$, where $\sigma = (\sigma_1, \ldots, \sigma_m)$ and each σ_i is sampled independently and uniformly at random from $\{-1, 1\}$.

(a) Prove that $M(S)$ satisfies the $\frac{1}{m}$-Lipschitz property on all m coordinates.

(b) Prove that $R_S(C)$ as a function of S satisfies the $\frac{1}{m}$-Lipschitz property on all m coordinates.

5. **(15 points) Massart's Lemma.** Let $\mathcal{V}$ be a finite subset of $\mathbb{R}^S$ with $|S| = m$ where each member v of $\mathcal{V}$ is denoted by $v = (v_1, \ldots, v_m)$. Let $\sigma_1, \ldots, \sigma_m$ be random variables chosen from $\{-1, +1\}$ uniformly at random such that all σ_i's are independent.

(a) **Jensen's Inequality.** Suppose X is a random variable and $f : \mathbb{R} \mapsto \mathbb{R}$ is a differentiable convex function. Prove that $\mathbf{E}[f(X)] \geq f(\mathbf{E}[X])$.
(Hint: A differentiable function $f : \mathbb{R} \mapsto \mathbb{R}$ is convex if and only if for all $x, y \in \mathbb{R}$, it holds that $f(x) \geq f(y) + f'(y)(x - y)$.)

(b) Let $\mu := \mathbf{E}[\max_{v \in \mathcal{V}} \sum_{i=1}^m \sigma_i v_i]$. Suppose $\lambda > 0$ is some constant. Prove that $e^{\lambda \mu} \leq \sum_{v \in \mathcal{V}} \prod_{i=1}^m \mathbf{E}\big[e^{\lambda \sigma_i v_i}\big]$.
(Hint: The function $f(x) := e^{\lambda x}$ is convex.)

(c) Let $r := \max_{v \in \mathcal{V}} \sqrt{\sum_{i=1}^m v_i^2}$. Prove that $\mu \leq r\sqrt{2 \ln |\mathcal{V}|}$.
(Hint: For $x \in \mathbb{R}$, it holds that $\frac{e^x + e^{-x}}{2} \leq e^{\frac{x^2}{2}}$.)

6. **(10 points) Alternative Proof of Sauer's Lemma.** Suppose C is a class of boolean functions on X and the VC-dimension of (X, C) is at most d. Recall the following shifting procedure. We represent points in C as rows in a table, where each row corresponds to a point in C and each column corresponds to a coordinate in S. In each round, we select an arbitrary column that has not been selected, and then repeatedly change 1's into 0's if the changing does not lead to a row that is already in the table.

   Use the shifting procedure to prove Sauer's Lemma: for every subset S of X such that $|S| = m$, the cardinality of the projection $C(S)$ is at most $\binom{m}{\leq d}$.

   (Hint: You can use any results about the shifting procedure proved in class.)

7. **(15 points) Alternative Proof for the 1-inclusion Graph.** Let S be a set of size m. Let $C \subseteq \{0,1\}^S$ be a collection of points with VC-dimension d. Suppose E is the set of edges in the 1-inclusion graph of (S, C). We have proved in class that $|E| \leq d \cdot |C|$ using the shifting procedure. In this question, we prove the same result by induction on d and m, which is similar to the proof of Sauer's Lemma shown in Lecture 8.

   (a) Prove that $|E| \leq d \cdot |C|$ is true for the base cases $d = 0$ or $m = 1$.

   (b) Suppose $d \geq 1$ and $m > 1$. Let $x \in S$ and $S' := S \setminus \{x\}$. Let $C_1 := C_{|S'}$ be the projection of C on S'. Let $C_2 \subseteq C_1$ be the set of points f in C_1 such that there exist $f_1, f_2 \in C$, where f_1 and f_2 disagree on x and $f_1 \,|_{S'} = f_2 \,|_{S'} = f$. Let E_1 be the set of edges in the 1-inclusion graph G of (S', C_1), and $E_2 \subseteq E_1$ the set of edges in the induced subgraph $G[C_2]$.

      i. Give upper bounds for $|E_1|$ and $|E_2|$ in terms of d, $|C_1|$ and $|C_2|$.
      (Hint: Use the induction hypothesis.)

      ii. Prove that $|E| \leq |E_1| + |E_2| + |C_2|$.

      iii. Prove that $|E| \leq d \cdot |C|$.