CSIS0351/8601: Randomized Algorithms
Homeworks 3 & 4 **Due Date:** 28 Nov 2011

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).

The weight of this homework is the same as those of Homeworks 1 and 2 *together*. This homework has 260 points, of which 60 points are extra credit.

1. **(15 points) Some Technical Proofs from Johnson-Lindenstrauss Lemma**

   (a) Suppose g is a random variable with normal distribution $N(0,1)$. Prove the following.

   i. For odd $n \geq 1$, $E[g^n] = 0$.
   ii. For even $n \geq 2$, $E[g^n] \geq 1$.

   (Hint: Use induction. Let $I_n := E[g^n] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^n e^{-\frac{x^2}{2}} dx$. Use integration by parts to show that $I_{n+2} = (n+1)I_n$.)

   (b) Suppose γ_j's are independent uniform $\{-1,1\}$-random variables and g_j's are independent random variables, each having normal distribution $N(0,1)$. Suppose v_j's are real numbers, and define $X := (\sum_j \gamma_j v_j)^2$ and $\widehat{X} := (\sum_j g_j v_j)^2$. Show that for all integers $n \geq 1$, $E[X^n] \leq E[\widehat{X}^n]$.

   (c) Suppose Z is a random variable having normal distribution $N(0,\nu^2)$. Compute $E[e^{tZ^2}]$. For what values of t is your expression valid?

2. **(25 points) Can Johnson-Lindenstrauss Lemma preserve area?**

   (a) Suppose the distances between three points are preserved with multiplicative error ϵ. Is the area of the corresponding triangle also always preserved with multiplicative error $O(\epsilon)$, or even some constant multiplicative error?

   (b) Suppose u and v are mutually orthogonal unit vectors. Observe that the vectors u and v together with the origin form a right-angled isosceles triangle with area $\frac{1}{2}$. Suppose the lengths of the triangle are distorted with multiplicative error at most ϵ. What is the multiplicative error for the area of the triangle?

   (c) Suppose a set V of n points are given in Euclidean space $\mathbb{R}^n$. Let $0 < \epsilon < 1$. Give a randomized algorithm that produces a low-dimensional mapping $f : V \to \mathbb{R}^T$ such that the areas of all triangles formed from the n points are preserved with multiplicative error ϵ. What is the value of T for your mapping? Please give the exact number and do not use big O notation.

   (Hint: If two triangles lie in the same plane (a 2-dimensional affine space) in $\mathbb{R}^n$, then under a linear mapping their areas have the same multiplicative error. For every triangle, add an extra point to form a right-angled isosceles triangle in the same plane.)

3. **(20 points) VC-dimension of Axis-aligned rectangles.**

   (a) Prove that no 5 points on the plane $\mathbb{R}^2$ can be shattered by the class C of axis-aligned rectangles (that map points inside a rectangle 1 and otherwise 0).

   (b) Compute the VC-dimension of the class C_k of k-dimensional axis-aligned rectangles in $\mathbb{R}^k$. In particular, you need to find a number d such that there exist d points in $\mathbb{R}^k$ that can be shattered by the C_k, and prove that any $d+1$ points in $\mathbb{R}^k$ cannot be shattered by C_k.

4. **(10 points) Conditional Expectation.** Suppose $Y : \Omega \to \mathbb{R}$ is a random variable and $W : \Omega \to \mathcal{U}$ is a random object defined on the same probability space $(\Omega, \mathcal{F}, Pr)$. Prove that $E[Y] = E[E[Y|W]]$. You may assume that both Ω and $\mathcal{U}$ are finite.

5. **(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?

6. **(75 points) ϵ-Sample for (X, C) with VC-dimension d.** Suppose X is a set and C is a collection of boolean functions such that (X, C) has VC-dimension d. In this question, we derive a sufficient number m of independent random samples from X with distribution D such that the resulting bag S is an ϵ-sample under class C of boolean functions with probability at least $1 - \delta$.

   (a) **Introducing Extra Randomness.** (15 points) Suppose we sample $2m$ copies independently from X to form the bag W. Then, we pick m copies out of W at random to form S. In other words, W can be view as a point in X^{2m}, and we pick m distinct coordinates at random and use them to form S.

   Let A be the event that there exists some $F \in C$ such that $|E_X[F(x)] - E_S[F(x)]| > \epsilon$.

   Let B be the event that there exists some $F \in C$ such that $|E_X[F(x)] - E_S[F(x)]| > \epsilon$ and $|E_W[F(x)] - E_S[F(x)]| > \frac{\epsilon}{4}$.

Prove that $Pr[A] \leq 2Pr[B]$.

(Hint: Show that $Pr[\overline{B}|A] \leq \frac{1}{2}$.

Observe that given A, the event $\overline{B}$ implies that there is some $F_0 \in C$ such that $|E_X[F_0(x)] - E_S[F_0(x)]| > \epsilon$ and $|E_W[F_0(x)] - E_S[F_0(x)]| \leq \frac{\epsilon}{4}$. This means that $|E_X[F_0(x)] - E_{W \setminus S}[F_0(x)]| > \frac{\epsilon}{2}$.

Use Hoeffding's Inequality and you may assume $m \geq \frac{2 \ln 4}{\epsilon^2}$.)

(b) **Conditional Probability.** (10 points) For $F \in C$, define B_F to be the event that $|E_X[F(x)] - E_S[F(x)]| > \epsilon$ and $|E_W[F(x)] - E_S[F(x)]| > \frac{\epsilon}{4}$. (Hence, $B = \cup_F B_F$.)

Fix $F \in C$. Define H_F to be the event that $|E_W[F(x)] - E_S[F(x)]| > \frac{\epsilon}{4}$. Then, clearly $B_F \subseteq H_F$, and so $Pr[B_F|W] \leq Pr[H_F|W]$. We analyze $Pr[H_F|W]$.

Suppose $P_{max} := \max_{F \in C} Pr[H_F|W]$. Prove that $Pr[B] \leq (2m)^d \cdot P_{max}$.

(Hint: Recall that (X, C) has VC-dimension d. After conditioning on W which has only $2m$ points, how many boolean functions can the class C induce on W?)

(c) **Bounding P_{max}.** (40 points) This is the most technical part of the proof and this part differs the most from the proof for ϵ-net.

After W and F are fixed, we know precisely how many copies in W are marked 1 by F. Let this number be L. The only randomness left is the choice of S among W. Recall that S is formed from W by choosing m copies from the $2m$ copies in W.

We can order the objects in W in an arbitrary list, and assign one by one whether each object is in S in the following way: suppose when object a is considered, there are already x objects assigned to S and y objects assigned to $W \setminus S$. Then, object a is assigned to S with probability $\frac{m-x}{(m-x)+(m-y)}$ and to $W \setminus S$ with probability $\frac{m-y}{(m-x)+(m-y)}$.

i. Suppose the L objects marked 1 are being considered first. For $1 \leq i \leq L$, let u_i be the variable that takes value 1 if the ith object is assigned to S and -1 if it is assigned to $W \setminus S$.

   Define $U_i := \sum_{j=1}^{i} u_j$. Compute the probability that the $(i+1)$st object is assigned to S in terms of i and U_i.

   What does it mean when $U_i > 0$? When $U_i > 0$, what happens to this probability?

   Are the u_i's independent?

ii. Find an expression β in terms of ϵ and m such that $|E_W[F(x)] - E_S[F(x)]| > \frac{\epsilon}{4}$ *iff* $U_L^2 > \beta$.

   (We want to obtain an upper bound for $Pr[U_L^2 > \beta]$.)

iii. We saw that the u_i's are not independent. This makes the analysis difficult. Hence, we would like to compare the u_i's with another collection of independent random variables. For each $1 \leq i \leq L$, we define independent random variable γ_i that takes values in $\{-1, 1\}$ uniformly, i.e., each value with probability $\frac{1}{2}$. Define $Y_i := \sum_{1 \leq j \leq i} \gamma_j$.

Observe that we would like U_L^2 to be small. Can you explain intuitively why Y_L^2 is more likely to be larger than U_L^2?
Prove that $E[U_L^2] \leq E[Y_L^2]$.
(Hint: Prove by induction on i that $E[U_i^2] \leq E[Y_i^2]$. In the inductive step, you might find considering the conditional expectation $E[U_i u_{i+1} | U_i]$ useful.)
(Optional: Prove that for all non-negative integers r, $E[U_L^{2r}] \leq E[Y_L^{2r}]$. You may use this result for later parts of the question.)

iv. Let t be a positive real number. Prove that $E[\exp(tU_L^2)] \leq E[\exp(tY_L^2)]$.
(Hint: Recall the Taylor expansion $\exp(y) := \sum_{r \geq 0} \frac{y^r}{r!}$.)

v. By considering moment generating functions, prove an upper bound for $Pr[U_L^2 > \beta]$, and conclude that $P_{max} \leq 2\exp(-\frac{\epsilon^2 m}{32})$.
(Hint: Recall from the lecture on Johnson-Lindenstrauss Lemma, we have $E[\exp(tY_L^2)] \leq (1-2tL)^{-1/2}$, for $t < \frac{1}{2L}$.)

(d) **Wrapping Everything Up.** (10 points) Prove that if $m \geq \max\{\frac{64}{\epsilon^2} \ln \frac{2}{\delta}, \frac{128d}{\epsilon^2} \ln \frac{32d}{\epsilon^2}\}$, then with probability at least $1 - \delta$, the bag S is an ϵ-sample for X under class C.

7. **(20 points) Properties of Symmetric Geometric Distribution.** Let $\alpha > 1$, and let γ be a random variable sampled from the symmetric geometric distribution $\mathsf{Geom}(\alpha)$, i.e., $\Pr[\gamma = k] = \frac{\alpha-1}{\alpha+1} \cdot \alpha^{-|k|}$. Prove that

   (a) (5pt) $E[\gamma] = 0$,

   (b) (5pt) $var[\gamma] = \frac{2\alpha}{(1-\alpha)^2}$,

   (c) (10pt) for any integer $z \geq 0$, $\Pr[|\gamma| > z] \leq \frac{1}{\alpha^z}$.

8. **(20 points) Achieving Differential Privacy with Geometric Distribution.** Let $f : \mathcal{D} \to \mathbb{Z}^d$ be a deterministic function, $0 < \epsilon < 1$ be the privacy parameter and $0 < \delta < 1$ be the failure probability. Let $\gamma_1, \gamma_2, \ldots, \gamma_d$ be random variables independently sampled from $\mathsf{Geom}(\exp(\frac{\epsilon}{\Delta f}))$, where $\Delta f := \max_{X \sim Y \in \mathcal{D}} ||f(X) - f(Y)||_1$ is the ℓ_1-sensitivity.

   Prove that the randomized function $\widehat{f}$ such that $\widehat{f}_i(X) := f_i(X) + \gamma_i$ for all $i \in [d]$

   (a) preserves ϵ-differential privacy,

   (b) is $(\frac{\Delta f}{\epsilon} \ln \frac{d}{\delta}, \delta)$-useful with respect to f, i.e., for all $X \in \mathcal{D}$, with probability at least $1 - \delta$, for all $i \in [d]$, $|f_i(X) - \widehat{f}_i(X)| \leq \frac{\Delta f}{\epsilon} \ln \frac{d}{\delta}$.

9. **(10 points) Deterministic Operation on Differentially Private Output Remains Differentially Private.** Let $f : \mathcal{D} \to \mathcal{O}_1$ be a ϵ-differentially private randomized algorithm, and let $g : \mathcal{O}_1 \to \mathcal{O}_2$ be a deterministic function. Prove that $g \circ f : \mathcal{D} \to \mathcal{O}_2$, whose value at $X \in \mathcal{D}$ is $g(f(X))$, preserves ϵ-differential privacy.

10. **(45 points) Sum of Independent Laplace Random Variables.** In this question, we derive a measure concentration result for independent random variables drawn from

Laplace distribution. We show that with high probability, the sum of independent Laplace random variables are concentrated around its mean, 0.

We use moment generating functions in a Chernoff-like argument. Let $b_1, b_2, \ldots, b_n > 0$ and $\gamma_1, \gamma_2, \ldots, \gamma_n$ be n independent random variables, where for each i γ_i is sampled from $\mathsf{Lap}(b_i)$.

(a) Prove that for each γ_i, the moment generating function is $E[\exp(h\gamma_i)] = \frac{1}{1-h^2 b_i^2}$, where $|h| < \frac{1}{b_i}$.

(b) Show that $E[\exp(h\gamma_i)] \leq \exp(2h^2 b_i^2)$, if $|h| < \frac{1}{\sqrt{2}b_i}$.
(Hint: for $|x| < \frac{1}{2}$, we have $\frac{1}{1-x} \leq 1 + 2x \leq \exp(2x)$.)

(c) Let $b_M := \max_{i \in [n]} b_i$. Also, let $\nu \geq \sqrt{\sum_{i=1}^n b_i^2}$ and $0 < \lambda < \frac{2\sqrt{2}\nu^2}{b_M}$. Prove that $\Pr[|Y| > \lambda] \leq 2\exp\left(-\frac{\lambda^2}{8\nu^2}\right)$

(d) Suppose $0 < \delta < 1$ and $\nu > \max\left\{\sqrt{\sum_{i=1}^n b_i^2}, b_M\sqrt{\ln\frac{2}{\delta}}\right\}$. Prove that $\Pr[|Y| > \nu\sqrt{8\ln\frac{2}{\delta}}] < \delta$.

(e) Prove that $\Pr[|Y| > \sqrt{8} \cdot \sqrt{\sum_{i=1}^n b_i^2} \cdot \ln\frac{2}{\delta}] < \delta$.