Advanced Topics in Artificial Intelligence: EE6180
Indian Institute of Technology Madras — August 27, 2019
Instructor: Abhishek Sinha

Problem Set 3

- This problem set is due on **September 11, 2019** in the class.
- Each problem carries 10 points.
- You may work on the problems in groups of size at most **two**. However, **each student must write their own solution**. If you collaborate on the problems, clearly mention the name of your collaborator.

---

1. **(Zero Expectation)** Let X be a random variable such that for some $\lambda_0 > 0$:

$$\mathbb{E}(\exp(\lambda X)) \leq \exp\big(\frac{\lambda^2}{2}\big), \quad \forall |\lambda| < \lambda_0.$$

   Show that $\mathbb{E}(X) = 0$.

2. **(Moments vs. Chernoff Bounds)** In this problem we show that the moment bounds for tail probabilities are always better than Chernoff bounds. More precisely, let X be a nonnegative random variable and let $t > 0$. The best moment bound for the tail probability $\mathbb{P}(X \geq t)$ is $\min_q \mathbb{E}(X^q)t^{-q}$ where the minimum is taken over all positive integers. The best Chernoff bound is $\inf_{s>0} \mathbb{E}\big(\exp(s(X-t))\big)$. Prove that

$$\min_q \mathbb{E}(X^q)t^{-q} \leq \inf_{s>0} \mathbb{E}\big(\exp(s(X-t))\big).$$

3. **(Area-preserving Johnson-Lindenstrauss Embedding)** In this problem, you will show how to extend the Johnson-Lindenstrauss lemma to preserve the areas of all the triangles obtained from the point set X.

   (a) Show that preserving all the pairwise within a factor of $1+\epsilon$ does not necessarily imply preserving the areas of triangles within $1+\epsilon$ factor.

   (b) Let ABC be a right-angled triangle, and let f be a non-expansive embedding of its vertices into Euclidean space, with distortion $1+\epsilon, 0 < \epsilon \leq 1/8$. Bound the distortion of the area of the triangle $f(A)f(B)f(C)$.

   (c) Show that for all $X \in \mathbb{R}^d, |X| = n$, there exists an embedding of X into $\mathbb{R}^m, m = O(\frac{\log n}{\epsilon^2})$, such that for every 3 points in X the area of the triangle they form is distorted by at most $1+\epsilon$.
   Hint: Add additional points to X, and show that preserving distances in this bigger set with distortion $1+\epsilon'$ implies that all areas of triangles in X are preserved with distortion $1+\epsilon$. Note that, if f is a linear embedding and the points $\boldsymbol{x}, \boldsymbol{y}$ are collinear, then the points $f(\boldsymbol{x})$ and $f(\boldsymbol{y})$ are also collinear.

4. **(Independent Sets in Erdős-Rényi Ensemble)** Given a graph G with the node set V and edge set E, a set of nodes $I \subseteq V$ is called an *independent set* if there is no edge between any two nodes in I. Let Z be the total number of independent sets in G (which is at least one since we assume that the empty set is an independent set and at most $2^{|V|}$).

   Let $G = G(n, p)$ be a random Erdős-Rényi graph, which, as discussed in the class, is obtained by connecting two vertices i and j independently with probability p and leaving them unconnected otherwise. Let Z_n be the corresponding (random) number of independent sets in G. Establish the following concentration inequality for $\log Z_n$ around its expectation:

$$\mathbb{P}\big(\log Z_n \geq \mathbb{E}[\log Z_n] + t\big) \leq \exp\big(-\frac{t^2}{Cn}\big),$$

   for some constant C which does not depend on n.

5. **(Concentration for Spin Glasses)** Consider a set of $d \geq 2$ binary random variables $\{X_1, X_2, \ldots, X_d\}$, each taking value either $+1$ (*up-spin*) or -1 (*down-spin*), arranged in the form of a complete graph. The weight of the edge (i, j) is denoted by the real number $\theta_{ij}, \forall i < j$. The joint probability distribution of the random variables is given by the following probability mass function

$$\mathbb{P}_\theta(x_1, x_2, \ldots, x_d) = \exp\left(\frac{1}{\sqrt{d}}\big(\sum_{i<j} \theta_{ij} x_i x_j\big) - F_d(\theta)\right), \tag{1}$$

   where the function $F_d : \mathbb{R}^{\binom{d}{2}} \to \mathbb{R}$, known as the *free energy*, is given by

$$F_d(\boldsymbol{\theta}) = \log\left(\sum_{\boldsymbol{x} \in \{\pm 1\}^d} \exp\big(\frac{1}{\sqrt{d}} \sum_{i<j} \theta_{ij} x_i x_j\big)\right).$$

   serves to normalize the distribution. The probability distribution (1) was originally used to describe the behaviour of magnets in statistical physics, in which context it is known as the *Ising model*. In Machine Learning, this model is extensively used, *e.g.*, to study theoretical properties of Neural networks. In this problem, we will study concentration of the free energy function $F_d(\boldsymbol{\theta})$.

   (a) Show that $F_d(\cdot)$ is a convex function.

   (b) For any two vectors $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^{\binom{d}{2}}$, show that $|F_d(\boldsymbol{\theta}) - F_d(\boldsymbol{\theta}')| \leq \sqrt{d/2}||\boldsymbol{\theta} - \boldsymbol{\theta}'||_2$.

   (c) Now suppose that the weights $\boldsymbol{\Theta}$ are chosen as i.i.d. random variables, so that equation (1) now describes a random family of probability distributions, known as the *Sherrington-Kirkpatrick* (SK) model in statistical physics. As a special case, suppose that the weights are chosen in an i.i.d. manner as $\Theta_{ij} \sim \mathcal{N}(0, \sigma^2)$ for each $i \neq j$. Show that

$$\mathbb{P}\left[\frac{F_d(\boldsymbol{\Theta})}{d} \geq \log 2 + \frac{\sigma^2}{4} + t\right] \leq \exp(-dt^2/\sigma^2), \quad \forall t > 0.$$