Problem Set 3
Deadline: April 25th (at 23:59) in Gradescope

Instructions

- You should think about each problem by yourself for at least an hour before choosing to collaborate with others.
- You are allowed to collaborate with fellow students taking the class in solving the problems (in groups of at most 3 people for each problem). In fact this is encouraged so that you interact with and learn from each other. However, you must write up your solutions on your own. If you collaborate in solving problems, you should clearly acknowledge your collaborators for each problem.
- You are not allowed to search for answers or hints on the web. You are encouraged to contact the instructors or the TA for a possible hint if you feel stuck on a problem and require some assistance.
- Solutions typeset in LATEX are preferred.
- Feel free to use the Discussion Board or email the instructor or the TA if you have any questions or would like any clarifications about the problems.
- You are urged to start work on the problem set early.

1) Recall that rank($A$) of a matrix $A$ is the number of linearly independent columns of $A$; it is also the same as the number of linearly independent rows of $A$. We say an $n \times n$ matrix $A$ is nonsingular if rank($A$) = $n$. In addition, det($A$) $\neq 0$ if and only if $A$ is nonsingular. Let $G = (X,Y,E)$ be a given bipartite graph.

Using the above terminology, we can rewrite the algorithm that tests whether $G$ has a perfect matching as follows: For each edge $x_i,y_j$ of $G$, choose $A_{i,j}$ uniformly and independently from the set $\{0,1,\ldots,n^2\}$, and let the rest of entries of $A$ be 0. Return yes if rank($A$) = $n$ and no otherwise.

a) **extra credit:** Design an $O(n^3)$ time algorithm that computes rank($A$).

b) Let $A$ be the following matrix: For each nonadjacent pair $x_i,y_j$, let $A_{i,j} = 0$; choose the rest of the entries of $A$ arbitrarily. Show that if rank($A$) = $k$, then $G$ has a matching of size at least $k$.

c) Design a randomized algorithm to compute the size of the maximum matching of $G$.

2) Let $u,v \in \mathbb{R}^d$ and $g \in \mathbb{R}^d$ be a random Gaussian vector, i.e., for each $1 \leq i \leq d$, $g_i \sim \mathcal{N}(0,1)$.

a) What is the expected value of $\langle g,u \rangle$?

b) What is the expected value of $\langle g,u \rangle \cdot \langle g,v \rangle$?

c) What is the expected value of $|\langle g,u \rangle|$? You can use that p.d.f. of a $\mathcal{N}(0,1)$ is $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

3) (Dimension reduction for SVMs with margin) Suppose we are given two sets $A,B$ of unit vectors each of size $n$ in $\mathbb{R}^d$ with the guarantee that there exists a hyperplane $\{x : \langle v,x \rangle = 0\}$ such that every point in $A$ is on one side and every point in $B$ is on the other. Furthermore, suppose the $\ell_2$ distance of each point in $A$ and $B$ to this hyperplane is at least $\epsilon$.

a) Let $a,b$ be two unit vectors in $\mathbb{R}^d$, and let $F : \mathbb{R}^d \rightarrow \mathbb{R}^k$ be a linear mapping such that $\|F(a)\|^2 = (1 + \epsilon)\|a\|^2$, $\|F(b)\|^2 = (1 + \epsilon)\|b\|^2$ and $\|F(a - b)\|^2 = (1 + \epsilon)\|a - b\|^2$, where we say $A = (1 + \epsilon)B$ if $(1 - \epsilon)A \leq B \leq (1 + \epsilon)A$. Show that

$$|\langle F(a),F(b) \rangle - \langle a,b \rangle| \leq O(\epsilon).$$

3-1
b) Show using the Johnson Lindenstrauss lemma that in a random linear mapping to $O(\log n/\epsilon^2)$ dimensions the points in $A, B$ are still separable by a hyperplane with margin $\epsilon/2$. 