Algorithmic Aspects of Machine Learning: Problem Set # 2

Instructor: Ankur Moitra

Due: November 6

You can work with other students, but you must write-up your solutions by yourself and indicate at the top who you worked with!

Problem 1

(a) Suppose we want to solve the linear system $Ax = b$ (where $A \in \mathbb{R}^{n\times n}$ is square and invertible) but we are only given access to a noisy vector $\tilde{b}$ satisfying

$$\frac{\|b - \tilde{b}\|}{\|b\|} \leq \varepsilon$$

and a noisy matrix $\tilde{A}$ satisfying $\|A - \tilde{A}\| \leq \delta$ (in operator norm). Let $\tilde{x}$ be the solution to $\tilde{A}\tilde{x} = \tilde{b}$. Show that

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \frac{\varepsilon\, \sigma_{\max}(A) + \delta}{\sigma_{\min}(A) - \delta}$$

provided $\delta < \sigma_{\min}(A)$.

(b) Now suppose we know A exactly, but A may be badly conditioned or even singular. We want to show that it may still be possible to recover a specific coordinate x_j of x. Let $\tilde{x}$ be any solution to $A\tilde{x} = \tilde{b}$ and let a_i denote column i of A. Show that

$$|x_j - \tilde{x}_j| \leq \frac{\|b - \tilde{b}\|}{C_j}$$

where C_j is the norm of the projection of a_j onto the orthogonal complement of $\text{span}(\{a_i\}_{i\neq j})$.

Problem 2

Recall that $u \odot v$ denotes the Khatri-Rao product between two vectors, and if $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ then $u \odot v \in \mathbb{R}^{mn}$ and corresponds to flattening the matrix uv^T into a vector, column by column. Also recall that the Kruskal rank k-rank of a collection of vectors $u_1, u_2, ..., u_m \in \mathbb{R}^n$ is the largest k such that *every* set of k vectors are linearly independent.

In this problem, we will explore properties of the Khatri-Rao product and use it to design algorithms for decomposing higher-order tensors.

(a) Let k_u and k_v be the k-rank of $u_1, u_2, ..., u_m$ and $v_1, v_2, ..., v_m$ respectively. Prove that the k-rank of $u_1 \odot v_1, u_2 \odot v_2, ..., u_m \odot v_m$ is at least $\min(k_u + k_v - 1, m)$.

(b) Construct an example where the k-rank of $u_1 \odot u_1, u_2 \odot u_2, ..., u_m \odot u_m$ is exactly $2k_u - 1$, and not any larger. To make this non-trivial, you must use an example where $m > 2k_u - 1$. *Hint:* You could use my favorite overcomplete dictionary, from class.

Further Clarification: Here I would like you to construct a *family* of examples, so that the inequality you proved in **(a)** is tight is infinitely often tight. Moreover all of the vectors in your example should be distinct.

(c) Given an $n \times n \times n \times n \times n$ fifth order tensor $T = \sum_{i=1}^{r} a_i^{\otimes 5}$ give an algorithm for finding its factors that works for $r = 2n - 1$, under appropriate conditions on the factors $a_1, a_2, ..., a_r$. *Hint:* Reduce to the third-order case.

In fact for random or perturbed vectors, the Khatri-Rao product has a much stronger effect of *multiplying* their Kruskal rank. These types of properties can be used to obtain algorithms for decomposing higher-order tensors in the highly overcomplete case where r is some polynomial in n.

Problem 3

In class we saw how to solve ICA using non-convex optimization. In this problem we will see how to solve ICA using tensor decomposition. Suppose we observe many samples of the form $y = Ax$ where A is an unknown non-singular square matrix and each coordinate of x is independent and satisfies $\mathbb{E}[x_j] = 0$ and $\mathbb{E}[x_j^4] \neq 3\,\mathbb{E}[x_j^2]^2$. The distribution of x_j is unknown and might not be the same for all j.

(a) Write down expressions for $\mathbb{E}[y^{\otimes 4}]$ and $(\mathbb{E}[y^{\otimes 2}])^{\otimes 2}$ in terms of A and the moments of x. (You should not have any A's inside the expectation.)

(b) Using part (a), show how to use the moments of y to produce a tensor of the form $\sum_j c_j a_j^{\otimes 4}$ where a_j denotes column j of A and the c_j are nonzero scalars.

(c) Show how to recover the columns of A (up to permutation and scalar multiple) using Jennrich's algorithm.

Problem 4

In the *planted clique* model, we start with a random graph $G(n, 1/2)$ (i.e. there are n vertices and every possible edge exists independently with probability $1/2$) and then plant a clique on a random vertex-subset $S \subseteq V$ of size $|S| = k$ (i.e. connect every vertex in S to every other vertex in S). Given the graph (and k), the goal is to exactly find the clique with high probability (i.e. probability tending to 1 as $n \to \infty$). Consider the regime where $k = k(n)$ satisfies $\sqrt{n \log n} \ll k(n) \ll n$ (where e.g. $k \ll n$ means $k = o(n)$, i.e. $k/n \to 0$ as $n \to \infty$).

(a) Give a simple algorithm (based on counting the degree of each vertex) and prove that it finds the clique with high probability. Use Hoeffding's inequality:

Theorem (Hoeffding)**.** *Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in [0,1]$ and let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Then for any $t > 0$,*

$$\mathbb{P}\left(|\bar{X} - \mathbb{E}[\bar{X}]| \geq t\right) \leq 2\exp(-2nt^2).$$

(b) Now consider the following *semirandom* model. First we generate a random graph $\tilde{G}$ with a planted clique as above. Then an adversary is allowed to delete any number of edges that are not in the clique (i.e. edges that do not have both endpoints in S), producing a graph G. Show that the algorithm from part (a) fails in the semirandom model.

(c) Consider the optimization problem from [Feige–Krauthgamer '99]:

$$\max_{d,\{u_i\}} \sum_{i \in V} \langle d, u_i \rangle^2$$

where $d \in \mathbb{R}^n$ is a unit vector and $\{u_i\}$ is an *orthogonal representation* of the graph, i.e. a unit vector $u_i \in \mathbb{R}^n$ is assigned to each vertex $i \in V$ such that if (i,j) are non-adjacent then $\langle u_i, u_j \rangle = 0$. This optimization problem can be solved efficiently using *semidefinite programming*. The intended form of the solution is $u_i = d$ for all $i \in S$ and $u_i \perp d$ for all $i \notin S$. Show that this procedure is *robust* to the semirandom model in the following sense. Suppose that $\tilde{G}$ has the property that every optimal $d, \{u_i\}$ takes the intended form. Show that G also has this property, regardless of the adversary's actions.