Algorithmic Aspects of Machine Learning: Problem Set # 2

Instructor: Ankur Moitra

Due: April 23rd

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

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.

Problem 1

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.

This question concerns the stochastic block model, where we generate a graph $G = (V, E)$ as follows. Each node $u \in V$ is mapped to one of k communities according to $\pi : V \to [k]$ and the probability of an edge (u, v) is q if $\pi(u) = \pi(v)$ and otherwise is p, for $q > p$. Our goal is to recover the underlying community structure, using tensor methods.

Problem 2

Given a partition of V into A, B, C and X, for each a, b, c in A, B, C respectively, consider the statistic

$$\widetilde{T}_{a,b,c} = \frac{|\{x \in X|(x,a),(x,b),(x,c) \in E\}|}{|X|}$$

(a) Prove that $\mathbb{E}[\widetilde{T}]$ is a low rank tensor. Here the expectation is over the random choices of which edges to put in G and which to leave out, with everything else (such as the community structure π) fixed.

(b) Sketch an algorithm for recovering the community structure (i.e. π up to a relabeling of which community is which) that works provided each community makes up a large enough fraction of V, and provided that q is large enough compared to p.

Problem 3

In this question, we will explore uniqueness conditions for sparse recovery, and conditions under which ℓ_1-minimization provable works.

(a) Let $A\widehat{x} = b$ and suppose A has n columns. Further suppose $2k \leq m$. Prove that for *every* $\widehat{x}$ with $\|\widehat{x}\|_0 \leq k$, $\widehat{x}$ is the uniquely sparsest solution to the linear system if and only if the k-rank of the columns of A is at least $2k$.

(b) Let $U = \text{kernel}(A)$, and that $U \subset \mathbb{R}^n$. Suppose that for each non-zero $x \in U$, and for any set $S \subset [n]$ with $|S| \leq k$ that

$$\|x_S\|_1 < \frac{1}{2}\|x\|_1$$

where x_S denotes the restriction of x to the coordinates in S. Prove that

$$(P1) \qquad \min \|x\|_1 \text{ s.t. } Ax = b$$

recovers $x = \widehat{x}$, provided that $A\widehat{x} = b$ and $\|\widehat{x}\|_0 \leq k$.

(c) **Bonus** Can you construct a subspace $U \subset \mathbb{R}^n$ of dimension $\Omega(n)$ that has the property that every non-zero $x \in U$ has at least $\Omega(n)$ non-zero coordinates? *Hint:* Use an expander.

Problem 4

Let $\widehat{x}$ be a k-sparse vector in n-dimensions. Let ω be the nth root of unity. Suppose we are given $v_\ell = \sum_{j=1}^{n} \widehat{x}_j \omega^{\ell j}$ for $\ell = 0, 1, ..., 2k-1$. Let $A, B \in \mathbb{R}^{k \times k}$ be defined so that $A_{i,j} = v_{i+j-2}$ and $B_{i,j} = v_{i+j-1}$.

(a) Express both A and B in the form $A = VD_AV^T$ and $B = VD_BV^T$ where V is a Vandermonde matrix, and D_A, D_B are diagonal.

(b) Prove that the solutions to the generalized eigenvalue problem $Ax = \lambda Bx$ can be used to recover the locations of the non-zeros in $\widehat{x}$.

(c) Given the locations of the non-zeros in $\widehat{x}$, and $v_0, v_1, ..., v_{k-1}$, given an algorithm to recover the values of the non-zero coefficients in $\widehat{x}$.

This is called the matrix pencil method. If you squint, it looks like Prony's method (Section 4.4) and has similar guarantees. Both are (somewhat) robust to noise if and only if the Vandermonde matrix is well-conditioned, and when exactly that happens is a longer story...