Algorithmic Aspects of Machine Learning: Problem Set # 1

Instructor: Ankur Moitra

Due: March 7th

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

Show that for a matrix M its rank and its border rank are always the same. In particular, suppose you are given a matrix M and a parameter r so that for every $\epsilon > 0$ there is a rank r matrix M_r so that M and M_r are entrywise ϵ-close. Show that M must have rank at most r. *Hint:* Use the Eckhart-Young Theorem.

Problem 2

In this problem, we will show that there are nonnegative matrices whose rank and nonnegative rank can be substantially different. Let $M \in \mathbb{R}^{n \times n}$ where $M_{i,j} = (i-j)^2$. Prove that $\text{rank}(M) = 3$ and that $\text{rank}^+(M) \geq \log_2 n$. *Hint:* To prove a lower bound on $\text{rank}^+(M)$ it suffices to consider just where it is zero and where it is non-zero.

Comment: There are examples where the separation is even more dramatic.

Problem 3

A well-known paper of Papadimitriou et al. considered the following document model: $M = AW$ and each column of W has only one non-zero and the support of each column of A is disjoint. Prove that the left singular vectors of M are the columns of A (after rescaling). You may assume that all the non-zero singular values of M are distinct. *Hint:* MM^T is block diagonal, after applying a permutation π to its rows and columns.

Problem 4

Let $M = AW$ where A is separable and the *rows* of M, A and W are normalized to sum to one. Also assume W has full row rank. Prove that Greedy Anchorwords

GREEDY ANCHORWORDS

1. Set $S = \emptyset$
2. Add the row of M with the largest ℓ_2 norm to S
3. For $i = 2$ to r
4. Project the rows of M orthogonal to the span of vectors in S
5. Add the row with the largest ℓ_2 norm to S
6. End

finds all the anchor words and nothing else. *Hint:* the ℓ_2 norm is strictly convex — i.e. for any x and y that are not multiples of each other and $t \in (0,1)$, $\|tx + (1-t)y\|_2 < t\|x\|_2 + (1-t)\|y\|_2$.

Problem 5

In this problem, we will design algorithms for decomposing higher-order tensors. Let $u \odot v$ denotee the Khatri-Rao product between two vectors, which is defined as follows: 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. Moreover 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.

(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 a family of examples 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:* One way to do this is to use a collection of orthonormal bases.

(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.

Comment: 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.