Algorithmic Aspects of Machine Learning: Problem Set # 1

Instructor: Ankur Moitra

Due: March 5th

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 $\text{rank}^+(M)$ is the smallest r such that there are entry-wise nonnegative matrices A and W with inner-dimension r, satisfying $M = AW$.

Problem 1

Which of the following are equivalent definitions of nonnegative rank? For each, give a proof or a counter-example.

(a) the smallest r such that M can be written as the sum of r rank one, nonnegative matrices

(b) the smallest r such that there are r nonnegative vectors $v_1, v_2, ..., v_r$ such that the cone generated by them contains all the columns of M

(c) the largest r such that there are r columns of M, $M_1, M_2, ..., M_r$ such that no column in set is contained in the cone generated by the remaining $r-1$ columns

Problem 2

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.

Problem 3

[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.

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

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 finds all the anchor words and nothing else. *Hint:* the ℓ_2 norm is strictly convex — i.e. for any $x \neq y$ and $t \in (0, 1)$, $\|tx + (1-t)y\|_2 < t\|x\|_2 + (1-t)\|y\|_2$.