Mathematics of Machine Learning
Homework 3

Instructors: Afonso S. Bandeira & Nikita Zhivotovskiy
Course Coordinator: Pedro Abdalla Teixeira

March 12, 2021

Try to solve the questions before looking to the answers. Every item must be proved rigorously. Starred problems are harder.

Problem 1

Consider a matrix $A \in \mathbb{R}^{n\times m}$ with singular values $\sigma_1(A) \geq \sigma_2(A) \ldots \geq \sigma_{\min\{m,n\}}(A)$. Prove the following results

(a) Variational characterization for the sum of singular values:

$$\sum_{i=1}^{\min\{m,n\}} \sigma_i(A) = \sup_{\sigma_1(Q)\leq 1} \langle Q, A\rangle_F.$$

Here $\langle , \rangle$ denotes the Frobenius inner product.

(b) Schatten p-norm is indeed a norm for $p = 1, 2, \infty$.

(c) A norm $\|.\|$ is said unitarily invariant if $\|UAV\| = \|A\|$ for all $U \in \mathcal{O}(m)$ and $V \in \mathcal{O}(n)$. Are Schatten p-norm unitarily invariant for $p = 1, 2, \infty$?

Problem 2

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix with positive eigenvalues $\lambda_1 > \lambda_2 \geq \ldots \lambda_n > 0$. Consider the following iteration for a given initial vector θ^0,

$$\theta^t := \frac{A\theta^{t-1}}{\|A\theta^{t-1}\|_2}.$$

(a) Prove that the iteration converges to a normalized eigenvector v_1 associated with the largest eigenvalue λ_1 if the initial vector is not orthogonal to v_1.

(b) Give examples in which the iteration above do not converge to v_1.

Problem 3

* Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix and let B be a principal submatrix of A of dimension $n - 1$ (B is obtained by deleting the same row and column of A). Prove the following facts:

(a) If $\alpha_1 \geq \ldots \geq \alpha_n$ are the eigenvalues of A and $\beta_1 \geq \ldots \beta_n$ are the eigenvalues of B, then

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \ldots \geq \alpha_n.$$

Hint: Use the Courant-Fischer principle for symmetric matrices.

(b) Let $G = (V, E)$ be a graph. If $G(S)$ is the graph induced by a subset $S \subset V$, then the average degree of $G(S)$ is at most the largest eigenvalue of the adjacency matrix A of G.

Hint: Find an upper bound for the average degree of $G(S)$ in terms of a submatrix of A.