CSCI 1520: Algorithmic Aspects of Machine Learning (Spring 2025)

Written Assignment 2

Due at 11:59pm ET, Thursday, Mar 13

1. (5 points) We will explore the uniqueness of the PageRank vector and how to compute it using iterative methods. Consider an unweighted directed graph $G = (V, E)$ with $|V| = n$. We define the transition matrix $M \in \mathbb{R}^{n \times n}$ of G as

$$M_{i,j} = \begin{cases} \frac{1}{d(j)} & \text{if there is an edge from node } j \text{ to node } i, \\ 0 & \text{otherwise,} \end{cases}$$

where $d(j)$ is the outgoing degree of node j.

Let $0 < \alpha < 1$ be the teleport probability. Let $\mathbf{1} \in \mathbb{R}^n$ be the all-ones vector.

(1) Prove that $(I - (1-\alpha)M)$ is strictly column diagonally dominant.
(A matrix A is strictly column diagonally dominant iff $|A_{j,j}| > \sum_{i \neq j} |A_{i,j}|$ for all j.)

(2) Prove that there is a unique vector $r^\star \in \mathbb{R}^n$ such that $r^\star = \alpha \frac{\mathbf{1}}{n} + (1-\alpha) M r^\star$.
(You can use these facts without proving them: all strictly column diagonally dominant matrices are invertible, and the inverse of an invertible real matrix is a real matrix.)

(3) For a vector x, we define the ℓ_1-norm of x as $\|x\|_1 = \sum_i |x_i|$.
For a matrix A, we define the ℓ_1-norm of A as $\|A\|_1 = \max_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_1}$.
Prove that $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |A_{i,j}|$.

(4) The PageRank vector $r^\star \in \mathbb{R}^n$ can be approximated as follows:
- Start with any nonnegative vector $r_0 \in \mathbb{R}^n$ with $\|r_0\|_1 = 1$.
- For $i = 1, 2, \ldots, t$, iteratively compute $r_i = \alpha \frac{\mathbf{1}}{n} + (1-\alpha) M r_{i-1}$.

Prove that after t iterations for some $t = O\left(\frac{\log(1/\epsilon)}{\alpha}\right)$, we have $\|r_t - r^\star\|_1 \leq \epsilon$.
(Hint: You may find the inequality $\|Ax\|_1 \leq \|A\|_1 \|x\|_1$ useful.)

2. (5 points) We will examine some basic properties of the Singular Value Decomposition (SVD). Let I be the identity matrix of appropriate dimensions. The (compact) SVD of a rank-r matrix $M \in \mathbb{R}^{m \times n}$ is a factorization of the form $M = U \Sigma V^\top$, where $U \in \mathbb{R}^{m \times r}$, $\Sigma \in \mathbb{R}^{r \times r}$, $V \in \mathbb{R}^{n \times r}$, $U^\top U = I$, $V^\top V = I$ (i.e., U and V have orthonormal columns), and Σ is a real nonnegative diagonal matrix. The diagonal entries of Σ are the singular values of M: $\sigma_i = \Sigma_{i,i}$. Suppose $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$.

(1) Suppose $m > r$. Prove that the following statement is false: $UU^\top = I$.

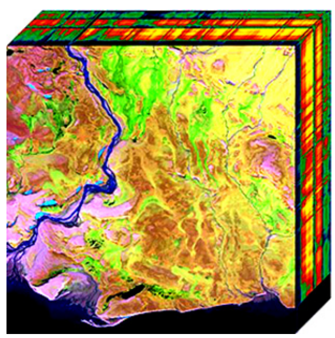
(2) A vector $z \neq 0$ is an eigenvector of a matrix A iff $Az = \lambda z$ for some $\lambda \in \mathbb{R}$.
Prove that every column of V is an eigenvector of $M^\top M$.

(3) The Frobenius norm of a matrix M is defined as $\|M\|_F = \sqrt{\sum_{i,j}(M_{i,j})^2}$.
Prove that $\|M\|_F^2 = \sum_{i=1}^r \sigma_i^2$.
(Hint: You can use the cyclic property of trace without proving it: $\mathrm{tr}(ABC) = \mathrm{tr}(BCA)$.)

(4) The spectral or operator norm of M is defined as $\|M\|_2 = \max_{x \neq 0} \frac{\|Mx\|_2}{\|x\|_2}$.
Prove that $\|M\|_2 = \sigma_1$.
(Hint: You can use $\|U\|_2 = \left\|V^\top\right\|_2 = 1$ without proving it. You may find the inequality $\|Ax\|_2 \leq \|A\|_2 \|x\|_2$ useful.)

(5) Let $y \in \mathbb{R}^m$ be a unit vector, i.e., $\|y\|_2 = 1$. Prove that $\left\|M^\top y\right\|_2$ is maximized when $y = u_1$, where u_1 is the first column of U.

3. (4 points) Hyperspectral imaging can capture the spectrum of light for each pixel in an image of a scene, with the goal of finding objects, identifying materials, or detecting processes. Unlike the human eye, which mostly perceives colors at three wavelengths (red, green, and blue), hyperspectral imaging can collect information across the electromagnetic spectrum and divide the spectrum into many more bands. [1]



Consider the hyperspectral unmixing problem, where the input is a hyperspectral image with $n \times n$ pixels. For each pixel, the input data provides intensity information at d different wavelengths (represented as a vector in $\mathbb{R}^d$). Each pixel contains a mixture of materials (e.g., $\frac{4}{7}$ soil, $\frac{2}{7}$ vegetation, and $\frac{1}{7}$ water). Each material has a unique signature spectrum (a vector in $\mathbb{R}^d$) that describes how it reflects light at different wavelengths. The spectrum of a pixel is a convex combination of the signature spectra of the materials in that pixel. The goal of hyperspectral unmixing is to find the signature spectrum of each material, and how much of each material is present in every pixel.

(1) Explain how to formulate hyperspectral unmixing as a Nonnegative Matrix Factorization (NMF) problem $M = AW$. What do the matrices M, A, and W represent in this context? If there are k different materials, what are the dimensions of M, A, and W?

[1]Explaination and figure are taken from Wikipedia.

(2) Translate the separability assumption into the context of hyperspectral unmixing.
(Hint: There are two possible translations because the matrices can be transposed. However, note that usually all entries of the signature spectra vectors are positive.)

4. (1 bonus point) We will examine normalization in Nonnegative Matrix Factorization (NMF). Let $R_{\geq 0}$ denote the set of nonnegative real numbers.

   Suppose we are given an input matrix $M \in \mathbb{R}_{\geq 0}^{m \times n}$. We want to compute an NMF, $M = AW$, where every column of A sums to 1.

   (1) Let $D \in \mathbb{R}^{m \times m}$ be a diagonal matrix with $D_{i,i} = \sum_{j=1}^{n} M_{i,j}$. Suppose $D_{i,i} > 0$ for all i. Let $\widetilde{M} = D^{-1}M$. Prove that every row of $\widetilde{M}$ sums to 1.

   (2) Suppose $\widetilde{M} = \widetilde{A}\widetilde{W}$, where $\widetilde{A} \in \mathbb{R}_{\geq 0}^{m \times r}$ and $\widetilde{W} \in \mathbb{R}_{\geq 0}^{r \times n}$. Let $B \in \mathbb{R}^{r \times r}$ be a diagonal matrix with $B_{j,j} = \sum_{i=1}^{m} (D\widetilde{A})_{i,j}$. Suppose $B_{j,j} > 0$ for all j.
   Let $A = D\widetilde{A}B^{-1}$. Prove that every column of A sums to 1.

   (3) Let $W = B\widetilde{W}$. Prove that (A, W) is an NMF of M.

5. (1 bonus point) We are interested in computing the largest eigenvector of a PSD matrix $M \in \mathbb{R}^{n \times n}$. Let $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$ be the eigenvalues of M, and let $v_1, \ldots, v_n$ be orthonormal eigenvectors such that $Mv_i = \lambda_i v_i$.

   The Rayleigh quotient for a matrix M and a non-zero vector x is defined as $\frac{x^\top Mx}{x^\top x}$. It can be shown that $\max_{x \in \mathbb{R}^n} \frac{x^\top Mx}{x^\top x} = \lambda_1$. Consider the following algorithm:

   Algorithm 1: The power method.

   Input : a PSD matrix $M \in \mathbb{R}^{n \times n}$ and a parameter $\epsilon > 0$.
   Output: a vector $v \in \mathbb{R}^n$.
   Pick $x_0 \sim \{-1, 1\}^n$ uniformly at random.
   Let $k = O(\frac{1}{\epsilon} \log \frac{n}{\epsilon})$.
   for $i = 1$ **to** k **do**
   $\quad x_i \leftarrow Mx_{i-1}$
   return x_k

   One can show that $\Pr\left[|\langle x, v\rangle| \geq \frac{1}{2}\right] \geq \frac{3}{16}$ for any unit vector v when x is sampled uniformly from $\{-1, 1\}^n$. (Note that this is not a strong correlation because $\|x\|_2 = \sqrt{n}$.) For simplicity, we assume $\langle x_0, v_1\rangle \geq \frac{1}{2}$ in Algorithm 1.

   (1) Suppose $\lambda_2 \leq (1 - \epsilon)\lambda_1$. Prove that $\frac{x_k^\top Mx_k}{x_k^\top x_k} \geq (1 - \epsilon)\lambda_1$.

   (2) Prove that $\frac{x_k^\top Mx_k}{x_k^\top x_k} \geq (1 - \epsilon)\lambda_1$ without the assumption in Part (1).