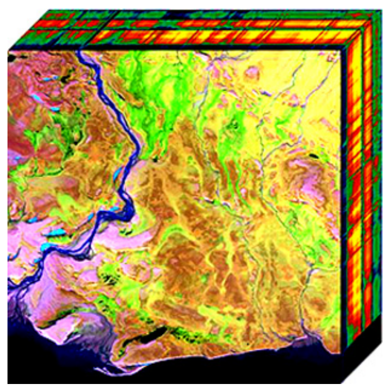
CSCI 1952Q: Algorithmic Aspects of Machine Learning (Spring 2024)

Written Assignment 2

Due at 1:00pm ET, Friday, Mar 1

1. (4 points) The goal of hyperspectral imaging is to obtain the spectrum for each pixel in the image of a scene, with the purpose of finding objects, identifying materials, or detecting processes. Whereas the human eye sees light in mostly three bands (long wavelengths – perceived as red, medium wavelengths – perceived as green, and short wavelengths – perceived as blue), spectral imaging divides 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 consists of a mixture of materials (e.g., soil, vegetation, etc., in satellite hyperspectral images). Different materials have different signature spectra (which are vectors in $\mathbb{R}^d$). The spectrum of a pixel is the convex combination of the spectra of its constituting materials. The goal of hyperspectral unmixing is to find the signature spectra of the materials, as well as the constituting materials for each pixel.

(1) Explain why hyperspectral unmixing can be viewed as an NMF problem $M = AW$. What are M, A, and W? If there are k different materials, what are the dimensions of M, A, and W?

(2) Translate the *separability* assumption into the context of hyperspectral unmixing.

(Hint: There are two possible translations because one can take the transpose of the matrices. However, all entries of the signature spectra vectors are strictly positive.)

[1] Explaination and figure are taken from Wikipedia.

2. (3 points) In this question, we 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.

3. (5 points) Let I be the identity matrix of appropriate dimensions. Recall that the (reduced) Singular Value Decomposition (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 $U \in \mathbb{R}^{m \times r}$, $\Sigma \in \mathbb{R}^{r \times r}$, $V \in \mathbb{R}^{n \times r}$, $U^\top U = I$, and $V^\top V = I$. Recall that the diagonal entries $\sigma_i = \Sigma_{i,i}$ are the singular values of M. Suppose $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$.

   (1) Suppose $m > r$. Disprove the following statement: $UU^\top = I$.

   (2) Prove that every column of V is an eigenvector of $M^\top M$.
   Recall that z is an eigenvector of A iff $Az = \lambda z$ for some scalar λ.

   (3) Prove that $\|M\|_F^2 = \sum_{i=1}^{r} \sigma_i^2$.

   (4) Prove that $\|M\|_2 = \sigma_1$.
   Recall that the spectral/operator norm of M is defined as $\|M\|_2 = \max_{x \in \mathbb{R}^n} \frac{\|Mx\|_2}{\|x\|_2}$.
   (Hint: Notice that by definition we have the following inequality $\|Ax\|_2 \leq \|A\|_2 \|x\|_2$. You can use $\|U\|_2 = \left\|V^\top\right\|_2 = 1$ without proving it.)

   (5) Suppose $y \in \mathbb{R}^m$ and $\|y\|_2 = 1$. Prove that $\left\|M^\top y\right\|_2$ is maximized when $y = u_1$, where u_1 is the first column of U.

4. (extra credit, 2 points) In this question, we are interested in computing the top 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 uniformly at random $x_0 \sim \{-1, 1\}^n$.
Let $k = O(\frac{1}{\epsilon} \log \frac{n}{\epsilon})$.
for $i = 1$ **to** k **do**
  $x_i \leftarrow Mx_{i-1}$
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 $\|x\|_2 = \sqrt{n}$.) Due to this reason and for simplicity, we assume $\langle x_0, v_1\rangle \geq \frac{1}{2}$ in Algorithm 1.

(1) Assume $\lambda_2 \leq (1-\epsilon)\lambda_1$ and prove that $\frac{x_k^\top M x_k}{x_k^\top x_k} \geq (1-\epsilon)\lambda_1$.

(2) Prove that $\frac{x_k^\top M x_k}{x_k^\top x_k} \geq (1-\epsilon)\lambda_1$.