Recitation 4

DS-GA 1013 Mathematical Tools for Data Science

1. Consider the equation $A\vec{x} = \vec{b}$ where $A \in \mathbb{R}^{m\times n}$, $\vec{x} \in \mathbb{R}^n$, and $\vec{b} \in \mathbb{R}^m$.

    1. Give conditions on $A, \vec{b}$ so that there is always an $\vec{x}$ satisfying the equation.
    2. When is the solution unique?
    3. Under what conditions does $A^T A\vec{x} = A^T\vec{b}$ have a solution?

Solution:

1. $\vec{b} \in \text{col}(A)$.
2. $\text{null}(A) = \{0\}$, or equivalently, that A has full column rank.
3. It always has a solution. We will prove this by showing $\text{col}(A^T A) = \text{col}(A^T)$. As $A^T A\vec{x} = A^T(A\vec{x})$ we see that $\text{col}(A^T A) \subseteq \text{col}(A^T)$. To complete the proof we will show that $\dim(\text{col}(A^T A)) = \dim(\text{col}(A^T))$. We begin by showing $\text{null}(A) = \text{null}(A^T A)$. To see this note that $A\vec{x} = 0$ implies $A^T A\vec{x} = 0$ and $A^T A\vec{y} = 0$ implies

$$0 = \vec{y}^T A^T A\vec{y} = \|A\vec{y}\|_2^2.$$

This proves $\text{null}(A) = \text{null}(A^T A)$, which implies $\text{rank}(A) = \text{rank}(A^T A)$ since

$$\text{rank}(A) + \dim(\text{null}(A)) = n = \text{rank}(A^T A) + \dim(\text{null}(A^T A)).$$

Thus

$$\dim(\text{col}(A^T)) = \text{rank}(A) = \text{rank}(A^T A) = \dim(\text{col}(A^T A)).$$

2. Let $A \in \mathbb{R}^{m\times m}$ with SVD $A = USV^T$.

    1. Assuming A is invertible, give the SVD for A^{-1}.
    2. Give the SVD for A^T.
    3. What is the relationship between the SVD of a symmetric matrix and the diagonal factorization given by the spectral theorem?
    4. Give the SVD for $A = \mathcal{P}_{\mathcal{S}}$, the orthogonal projection onto the subspace $\mathcal{S} \subseteq \mathbb{R}^m$.

Solution:

1. $A^{-1} = VS^{-1}U^T$
2. $A^T = VSU^T$
3. By the spectral theorem, we have $A = VSV^T$. If A is positive semidefinite then this is also an SVD. Otherwise, we can compute an SVD by $A = \widetilde{V}|S|V^T$ where $\widetilde{V}_{:,i} = V_{:,i}$ if $S_{ii} \geq 0$ and $\widetilde{V}_{:,i} = -V_{:,i}$ if $S_{ii} < 0$. Here $|S|$ is obtained from S by taking the absolute value of each entry.

4. Let U be a matrix whose columns form an orthonormal basis for $\mathcal{S}$. Then $A = UU^T = UIU^T$. Note here that I is a $k \times k$ matrix where k is the dimension of $\mathcal{S}$.

3. Let $A \in \mathbb{R}^{m \times n}$. Find maximizers $\vec{x} \in \mathbb{R}^m, \vec{y} \in \mathbb{R}^n$ solving

$$\begin{array}{ll} \text{maximize} & \vec{x}^T A \vec{y} \\ \text{subject to} & \|\vec{x}\|_2 = 1, \\ & \|\vec{y}\|_2 = 1. \end{array}$$

Also give the maximum value obtained.

Solution: Compute the SVD $A = USV^T$. The maximizers are given by $\vec{x} = U_{:,1}$ and $\vec{y} = V_{:,1}$ with maximum value given by $\sigma_1 = S_{11}$.

To see this is the maximum, note that

$$\vec{x}^T A \vec{y} \leq \|\vec{x}\| \|A\vec{y}\| = \|A\vec{y}\| \leq S_{11},$$

by Cauchy-Schwarz.

4. Let $A \in \mathbb{R}^{m \times n}$ have SVD $A = U\Sigma V^T$, where $\Sigma \in \mathbb{R}^{r \times r}$ and $\text{rank}(A) = r$. Compute the eigenvalues and the condition number of the block matrix

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}.$$

When computing the condition number, you can assume A is square and invertible.

Solution: There are $m + n - 2r$ eigenvalues with value 0, and the remaining have the values $\pm\sigma_1, \ldots, \pm\sigma_r$. Note that

$$\begin{bmatrix} 0 & U\Sigma V^T \\ V\Sigma^T U^T & 0 \end{bmatrix} \begin{bmatrix} \vec{u}_i \\ \vec{v}_i \end{bmatrix} = \sigma_i \begin{bmatrix} \vec{u}_i \\ \vec{v}_i \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & U\Sigma V^T \\ V\Sigma^T U^T & 0 \end{bmatrix} \begin{bmatrix} -\vec{u}_i \\ \vec{v}_i \end{bmatrix} = \sigma_i \begin{bmatrix} \vec{u}_i \\ -\vec{v}_i \end{bmatrix},$$

where $\vec{u}_i$ is the ith column of U, and $\vec{v}_i$ is the ith column of V. This gives $2r$ eigenvectors with eigenvalues $\pm\sigma_1, \ldots, \pm\sigma_r$. The other $m + n - 2r$ eigenvectors have the form

$$\begin{bmatrix} \vec{x} \\ 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 \\ \vec{y} \end{bmatrix}$$

where $\vec{x} \in \text{null}(U^T)$ and $\vec{y} \in \text{null}(V^T)$, and have eigenvalue 0. The condition number is σ_1/σ_n, the same as the condition number of A.

5. Let $X \in \mathbb{R}^{n \times p}$ denote a matrix whose **rows** are datapoints $\vec{x}_1^T, \ldots, \vec{x}_n^T \in \mathbb{R}^p$ with $p \leq n$. Suppose you are only given access to $G = XX^T$. How would you compute the first $k < p$ principal components of $\vec{x}_i$, for $i = 1, \ldots, n$?

Solution: First we have to center the data. If $\tilde{X}$ denotes the matrix of centered data then we have

$$\tilde{X} = X - \vec{\mathbb{1}}\vec{\mathbb{1}}^T X/n = (I - \vec{\mathbb{1}}\vec{\mathbb{1}}^T/n)X,$$

where $\vec{\mathbb{1}}$ is the all ones vector. Thus we can center G to obtain $\tilde{G}$ given by

$$\tilde{G} = (I - \vec{\mathbb{1}}\vec{\mathbb{1}}^T/n)G(I - \vec{\mathbb{1}}\vec{\mathbb{1}}^T/n) = \tilde{X}\tilde{X}^T.$$

Let $\tilde{X} = USV^T$ where $S \in \mathbb{R}^{p\times p}$. The principal directions are the eigenvectors of $\tilde{X}^T\tilde{X}$, which are the columns of V. The matrix of principal components is then given by

$$\tilde{X}V = US.$$

But we can compute this since

$$\tilde{G} = \tilde{X}\tilde{X}^T = US^2U^T$$

can be obtained from the spectral decomposition of $\tilde{G}$: if the spectral decomposition is $\tilde{G} = W\Lambda W^T$ then $US = W\sqrt{\Lambda}$. To obtain k principal components, we take the first k columns of US.

6. Generalizing the previous example, suppose there is a (possibly unknown) mapping $\Phi : \mathbb{R}^p \to \mathcal{H}$ where $\mathcal{H}$ is some Hilbert space. Suppose we are given a known, relatively easy to compute function $K : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ (called a *kernel*) such that

$$K(\vec{x}, \vec{y}) = \langle \Phi(\vec{x}), \Phi(\vec{y})\rangle_{\mathcal{H}}.$$

   1. Given a dataset $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^p$, if we compute the matrix $G \in \mathbb{R}^{n\times n}$ where $G_{ij} = K(\vec{x}_i, \vec{x}_j)$, what property of matrices must G have?
   2. Suggest a method for using K to perform a modified version of PCA (called Kernel PCA).
   3. One such kernel (the RBF or Gaussian kernel) is given by $K(\vec{x}, \vec{y}) = \exp(-\|\vec{x} - \vec{y}\|^2/\sigma^2)$. What does the fact that K is always positive-valued say about $\mathcal{H}$?
   4. The dth degree polynomial kernel is given by $K(\vec{x}, \vec{y}) = (1 + \vec{x}^T\vec{y})^d$. Give a space $\mathcal{H}$ corresponding to K. What is the dimension of $\mathcal{H}$?

Solution:

1. G must be positive semidefinite. If G constructed in this way is positive semidefinite for any choice of $\vec{x}_1, \ldots, \vec{x}_n$ we say the kernel K is positive definite. This condition is actually sufficient for a map Φ and a Hilbert space $\mathcal{H}$ to exist corresponding to K.
2. Use our previous exercise on PCA applied to G.
3. That Φ need not be surjective, and that no pairs of vectors in the image of Φ are orthogonal.
4. The space of all monomials up to degree d. This has dimension $\binom{p+d}{d}$ (the number of ways of choosing p degrees that sum to d). Thus we are computing inner products on a potentially large dimensional space quickly, but only on the image of Φ.