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?

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 $\mathcal{S} \subseteq \mathbb{R}^m$.

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.

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 $\operatorname{rank}(A) = r$. 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.

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 $i = 1, \ldots, n$?

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 $\mathcal{H}$?