CSE 250B: Section 8 - Sharad Vikram

1. k-medians Clustering

   Instead of calculating the mean for each cluster to determine its centroid, k-medians clustering calculates the median, where the median of a set of data $D = \{x_1, \ldots, x_n\}$ is

   $$\arg\min_{y \in \mathbb{R}^d} \sum_{i=1}^{n} ||x_i - y||_1$$

   (a) Please write down the objective function for k-medians clustering. Suppose you have data $\{x_i\}_{i=1}^N$ and cluster centers $\{z_k\}_{k=1}^K$.

   (b) What is the iterative algorithm to solve the clustering problem.

2. Kernelized k-means

   Suppose we have a dataset $\{x_i\}_{i=1}^N, x_i \in \mathbb{R}^d$ that we want to split into K clusters. Furthermore, suppose we know a priori that this data is best clustered in a large feature space $\mathbb{R}^m$, and that we have a feature map $\phi : \mathbb{R}^d \to \mathbb{R}^m$. How should we perform clustering in this space?

   (a) Write the objective for k-means clustering in the feature space (using the squared L_2 norm in the feature space). Do so by explicitly constructing cluster centers $\{\mu_k\}_{k=1}^K$ with all $\mu_k \in \mathbb{R}^m$.

   (b) Write an algorithm that minimizes the objective in (a).

   (c) Write an algorithm that minimizes the objective in (a) without explicitly constructing the cluster centers $\{\mu_k\}$. Assume you are given a kernel function $\kappa(x, y) = \phi(x)^T \phi(y)$.

   Hint: the cluster assignment for data point x_i can be written as $\arg\min_k f(x_i, k) = ||\phi(x_i) - \mu_k||_2^2$ and the cluster center can be written as a function of data points, i.e.

   $$\mu_k = \frac{1}{|S_k|} \sum_{x \in S_k} \phi(x)$$

3. Derivation of PCA

In this question we will derive PCA. PCA aims to find the direction of maximum variance among a dataset. You want the line such that projecting your data onto this line will retain the maximum amount of information. Thus, the optimization problem is

$$\max_{u:\|u\|_2=1} \frac{1}{n}\sum_{i=1}^{n}\left(u^T x_i - u^T \bar{x}\right)^2$$

where n is the number of data points and $\bar{x}$ is the sample average of the data points.

(a) Show that this optimization problem can be massaged into the form

$$\max_{u:\|u\|_2=1} u^T \Sigma u$$

where $\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$.

(b) Show that the maximizer for this problem is equal to v_1, where v_1 is the eigenvector corresponding to the largest eigenvalue λ_1 of Σ. Also show that optimal value of this problem is equal to λ_1.