Revision Session

Week 3

Story so far...

PCA

- Dimensionality reduction technique
- Transform the original feature space into lower dimension space.
- orthogonal features

Kernel PCA

Clustering

- Unsupervised learning

Goal:

Data: $\{x_1, x_2, \ldots, x_n\}$
Partition the dataset into K different partitions.

How many configurations are possible??

Which partition is better??

- We need a metric to quantify the goodness of the partition.

Lloyd's Algorithm (K-means algorithm)

Task: Partition the data into k cluster.

- **Step 1:** Choose K different cluster means (or centroids) randomly.
- **Step 2** Assign each data point to one of the K clusters.
  - Find the distance of a point from each cluster mean.
  - Assign to the cluster which is at minimum distance.
- **Step 3** Calculate the cluster mean again.
  - Find the mean of data points that belongs to a cluster.
- **Step 4** Repeat step 2 and 3 until convergence
  - convergence = when cluster mean does not change

More formal way:

Data = $x_1, x_2, \ldots, x_n$

Cluster Indicator = $z_1^t, z_2^t, \ldots, z_n^t$

$$z_i^t \in \{1, 2, \ldots, ?\}$$

z_i^t = function that maps the cluster index to ith data point in t^{th} iteration.

Lloyd's algorithm

- Initialization: $z_1^0, z_2^0, \ldots, z_n^0$
- Until convergence:
  - compute means for all K clusters: $\mu_k^t = \dfrac{\sum\limits_{i=1}^{n} x_i \mathbb{1}\left(z_i^t = k\right)}{\sum\limits_{i=1}^{n} \mathbb{1}\left(z_i^t = k\right)}$
  - Reassignment: $z_i^{t+1} = \arg\min\limits_k ||x_i - \mu_k^t||^2$

- Will the initialization affect the final clusters??
- **Objective: To minimize** $||x_i - \mu_{z_i}||^2$.

Better ways to initialize the cluster means

- **step 1:** Randomly choose the first cluster mean.
- **Step 2:** For choosing the next cluster mean, give higher weightage to the farthest point.
- **Step 3** Similary, for choosing t^{th} cluster mean, give higher weightage to farthest point from the nearest already chosen means.

how to choose K

Do we really know how many natural clusters are there in the dataset?

The objective : $\min ||x_i - \mu_{z_i}||^2$

- For $K = n$, What will be the objective function value?
- Should we choose high value of K?

We have to make sure that we don't end up choosing higher values of K.

- Penalize k

And therefore, objective function becomes

$$||x_i - \mu_{z_i}||^2 + P(K)$$

- $P(K)$ is a function of K such that P(K) increases as K increases.

Elbow method

- Start with some random K (Possibly the lesser value)
- Find the objective function value for final clusters
- Repeat for different values of K
- Choose elbow

Nature of clusters

As per the algorithm for $K = 2$, a point x goes in cluster 1 if $||x - \mu_1||^2 \leq ||x - \mu_2||^2$

Solving it, we have

$$x^T(\mu_2 - \mu_1) \leq \frac{||\mu_2||^2 - ||\mu_1||^2}{2}$$

Convergence:

$$\sum_{i=1}^{n} ||x_i - \mu^t_{z_i^{t+1}}||^2 < \sum_{i=1}^{n} ||x_i - \mu^t_{z_i^t}||^2$$

Fact:

$$\arg\min_v ||x_i - v||^2 = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

$$\sum_{i=1}^{n} ||x_i - \mu_{z_i^{t+1}}^{t+1}||^2 \leq \sum_{i=1}^{n} ||x_i - v||^2$$

$$\Rightarrow \sum_{i=1}^{n} ||x_i - \mu_{z_i^{t+1}}^{t+1}||^2 \leq \sum_{i=1}^{n} ||x_i - \mu_{z_i^{t+1}}^{t}||^2$$

$$\sum_{i=1}^{n} ||x_i - \mu_{z_i^{t+1}}^{t+1}||^2 \leq \sum_{i=1}^{n} ||x_i - \mu_{z_i^{t+1}}^{t}||^2 < \sum_{i=1}^{n} ||x_i - \mu_{z_i^{t}}^{t}||^2$$

$$\Rightarrow F(z_1^{t+1}, z_2^{t+1}, \ldots, z_n^{t+1}) < F(z_1^t, z_2^t, \ldots, z_n^t)$$

Objective function never repeats its value. It means partitions never repeat itself while reassinments.