CS 4786 Diagnostic Assignment P1

Lingfeng Cheng (lc674)

March 15, 2016

Q1

One simple heuristic to choose K for both PCA and CCA is to use the following criterion

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \geq \text{Threshold} \tag{1}$$

where λ is the eigen value of the corresponding covariance matrix. Threshold is typically 0.9 or 0.95. The reason for this choice is the goal for both PCA and CCA is to find the fewest principal components, i.e. eigen vectors, that retain the maximum variance/correlation. And according to the definition of PCA and CCA, if we list the eigen values in a descending order, the corresponding top eigen vectors represent orthogonal principal components in which the variance/correlation varies the most. Therefore, following Equation 1 actually means find the fewest coordinates/principal components so that the **Threshold** portion of the variance/correlation is retained.

The heuristic for finding K for random projection (RP) is to follow the JL Lemma.

$$K \approx \frac{log(\frac{n}{\delta})}{\epsilon^2} \tag{2}$$

As long as we define properly the number of data points n, the confidence interval $1-\delta$ and the closeness ϵ between the distance of two original points and projected points, we can calculate K using Equation 2

Q2

$$\begin{aligned}
\sum_{x_t \in C_j} \|x_t - x_s\|^2 &= \sum_{x_t \in C_j} \|x_t - r_j + r_j - x_s\|^2 \\
&= \sum_{x_t \in C_j} \|x_t - r_j\|^2 + \sum_{x_s \in C_j} \|r_j - x_s\|^2 + 2 \sum_{x_t \in C_j} (x_t - r_j)^\top (r_j - x_s) \\
&= \sum_{x_t \in C_j} \|x_t - r_j\|^2 + |C_j| \|r_j - x_s\|^2 + 2|C_j| (\frac{1}{C_j} \sum_{x_t \in C_j} x_t - r_j)^\top (r_j - x_s)
\end{aligned}$$

With the above analysis, we have

$$\begin{aligned}
M_2 &= \sum_{j=1}^{K} \frac{1}{|C_j|} \sum_{x_s, x_t \in C_j} \|x_s - x_t\|^2 \\
&= \sum_{j=1}^{K} \frac{1}{|C_j|} (\sum_{x_s \in C_j} (\sum_{x_t \in C_j} \|x_t - r_j\|^2 + |C_j| \|x_s - r_j\|^2)) \\
&= 2 \sum_{j=1}^{K} \frac{1}{|C_j|} (|C_j| \sum_{x_t \in C_j} \|x_t - r_j\|^2) \\
&= 2 \sum_{j=1}^{K} \sum_{x_t \in C_j} \|x_t - r_j\|^2 \\
&= 2M_1
\end{aligned}$$

In this way, we have proved that minimizing M_1 is equivalent to minimizing M_2, and M_1 is one half of M_2.

Q3

Consider the case when we have 2 clusters, let $c_j = -1$ if x_j belong to cluster 0 and $c_j = 1$ if x_j belong to cluster 1. According to the lecture notes, minimizing the inter-cluster cuts can be formulated as

$$\begin{aligned}
&min \quad c^\top L c \\
&= c^\top (D - A) c \\
&= c^\top D c - c^\top A c
\end{aligned}$$

And the first term is a constant, i.e. the sum of diagonal component of matrix D, therefore, the minimization problem is equivalent to

$$min \quad -c^\top A c$$

or

$$max \quad c^\top A c$$

And this is exactly maximization of inner-cluster number of edges.

Other algorithms: Enlightened by what we did in class, we can run the following optimization problem

$$\begin{aligned}
&max \quad c^\top A c \\
&s.t. \quad \|c\|_2 = \sqrt{n}
\end{aligned}$$

Find the eigen vectors $v_1, ..., v_n$ of A in a descending order, and pick the K eigen vectors with largest eigen values to get $y_1, ..., y_n$, and then use K-means clustering on $y_1, ..., y_n$.