BÀI 9 - PHÉP CHIẾU NGẪU NHIÊN

Bài 1

Giả sử chúng ta có một ma trận dữ liệu $X = [x_1, \ldots, x_n] \in \mathbb{R}^{p\times n}$. Đặt $d_{ij} = \|x_i - x_j\|_2$ là khoảng cách giữa x_i và x_j trong không gian Euclide. Bài toán Multidimensional Scaling cố gắng tìm $Y = [y_1, \ldots, y_n] \in \mathbb{R}^{k\times n}$ sao cho tổng bình phương nhỏ nhất

$$\min_{Y=[y_1,\ldots,y_n]} \sum_{i,j} \left(\|y_i - y_j\|_2^2 - d_{ij}^2\right)$$

với điều kiện $\sum_i y_i = 0$. Đặt

$$K = -\frac{1}{2}HDH$$

với $D = [d_{ij}^2]$ và $H = I - \frac{1}{n}11^T$ là ma trận centering. Chứng minh rằng bài toán cực tiểu trong (1) tương đương với:

$$\min_{Y\in\mathbb{R}^{k\times n}} \|Y^TY - K\|_F^2$$

trong đó $\|\cdot\|_F$ là chuẩn Frobenius.

Lời giải

1. **Biểu diễn tổng bình phương chênh lệch:**

$$\sum_{i,j} \left(\|y_i - y_j\|_2^2 - d_{ij}^2\right) = \sum_{i,j} \left(\|y_i\|_2^2 - 2\langle y_i, y_j\rangle + \|y_j\|_2^2 - d_{ij}^2\right)$$

Với điều kiện $\sum_i y_i = 0$, ta có:

$$\sum_i \|y_i\|_2^2 = \mathrm{Tr}(Y^TY)$$

$$\sum_{i,j} \langle y_i, y_j\rangle = \mathrm{Tr}(Y^TY)$$

Do đó, bài toán trở thành:

$$\min_Y \mathrm{Tr}(Y^TY) - 2\mathrm{Tr}(Y^TY) + \mathrm{Tr}(Y^TY) - \sum_{i,j} d_{ij}^2$$

$$= \min_Y -2\mathrm{Tr}(Y^TY) - \sum_{i,j} d_{ij}^2$$

2. **Biểu diễn ma trận K:**
Từ định nghĩa của K:

$$K = -\frac{1}{2}HDH$$

Ma trận $H = I - \frac{1}{n}11^T$ là ma trận centering, do đó:

$$HDH = (I - \frac{1}{n}11^T)D(I - \frac{1}{n}11^T)$$

H là ma trận đối xứng và $H^2 = H$, ta có thể biến đổi HDH sao cho:

$$HDH = H(D - \frac{1}{n}D11^T - \frac{1}{n}11^TD + \frac{1}{n^2}11^TD11^T)$$

$$= D - \frac{1}{n}D11^T - \frac{1}{n}11^TD + \frac{1}{n^2}11^TD11^T$$

Do đó:

$$K = -\frac{1}{2}\left(D - \frac{1}{n}D11^T - \frac{1}{n}11^TD + \frac{1}{n^2}11^TD11^T\right)$$

3. **Biểu diễn tối thiểu hóa chuẩn Frobenius:**
Bài toán tối thiểu hóa $\min_Y \|Y^TY - K\|_F^2$ có thể được viết lại thành:

$$\min_Y \|Y^TY - K\|_F^2 = \min_Y \mathrm{Tr}((Y^TY - K)(Y^TY - K)^T)$$

$$= \min_Y \mathrm{Tr}(Y^TYY^TY - 2Y^TYK + K^TK)$$

Do K đối xứng, $K^T = K$:

$$= \min_Y \mathrm{Tr}(Y^TYY^TY) - 2\mathrm{Tr}(Y^TYK) + \mathrm{Tr}(K^2)$$

Vì $\mathrm{Tr}(K^2)$ là hằng số, bài toán tối ưu trở thành:

$$\min_Y \mathrm{Tr}(Y^TYY^TY) - 2\mathrm{Tr}(Y^TYK)$$

Do đó, bài toán tối ưu trong không gian Frobenius tương đương với bài toán ban đầu trong (1).