ECE M146 Discussion 8
Introduction to Machine Learning
Instructor: Lara Dolecek
TA: Zehui (Alex) Chen

1. In class, you learned that the direction that maximize the variance of the projection onto a one-dimensional space is the eigenvector that corresponds to the largest eigenvalue of the data covariance matrix $S = \frac{1}{N}X^TX$. Formally, the solution to the following maximization problem

$$\underline{\max_{u_1} u_1^T S u_1} \quad \text{subject to } \underline{\|u_1\|^2 = 1},$$

is the eigenvector that corresponds to the largest eigenvalue of S.

Suppose u_2 is orthogonal to u_1 and have unit norm. We want to maximizes the variance of the data projected on u_2. Show that the optimal u_2 is defined by the second eigenvectors of the data covariance matrix S that corresponds to the second largest eigenvalues.

$$\max \; u_2^T S u_2$$

$$\text{s.t. } \|u_2\|^2 = 1 \quad \cdots \quad \lambda_2$$

$$u_2^T u_1 = 0 \quad \cdots \quad \eta_2$$

$$L(u_2, \lambda_2, \eta_2) = u_2^T S u_2 + \lambda_2(1 - u_2^T u_2) + \eta_2 u_2^T u_1$$

$$\nabla_{u_2} L(u_2, \lambda_2, \eta_2) = 2Su_2 - 2\lambda_2 u_2 + \eta_2 u_1 = 0$$

$$2\underline{\underline{u_1^T S}}\, u_2 - 2\lambda_2 u_1^T u_2 + \eta_2 u_1^T u_1 = 0$$

$$\cancel{2\lambda_1 u_1^T u_2} - \cancel{2\lambda_2 u_1^T u_2} + \eta_2 \|u_1\|^2 = 0$$

$$\boxed{\eta_2 = 0}$$

$$\therefore \; S u_2 = \lambda_2 u_2$$

$$\max_{u_2} \; \lambda_2$$

Let λ_2 be the 2-nd largest eigenvalue, u_2 be the corresponding eigenvector.

In class:

Maximum Variance Formulation

2. Minimum Error Formulation of PCA

$x_1, \ldots, x_N \in R^D$

Want to find basis vectors $\{u_i\}$ $\quad u_i^T u_j = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$

$$\tilde{x}_n = \sum_{i=1}^{M} z_{ni} u_i + \sum_{i=M+1}^{D} b_i u_i$$

$$\min_{u_1 \ldots u_D, z_{ni}, b_i} J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \tilde{x}_n \|_2^2$$

$$\|a+b+c\|^2 = \|a\|^2 + \|b\|^2 + \|c\|^2 + 2a^Tb + 2a^Tc + 2b^Tc$$

① Find the optimal z_{nj}

$$J = \frac{1}{N} \cancel{\sum_{n=1}^{N}} \| x_n - \sum_{i=1}^{M} z_{ni} u_i - \sum_{i=M+1}^{D} b_i u_i \|^2$$

$$J = z_{nj}^2 u_j^T u_j - 2 z_{nj} u_j^T x_n + \text{constant}$$

$$\frac{\partial J}{\partial z_{nj}} = 2 z_{nj} - 2 u_j^T x_n = 0 \Rightarrow z_{nj} = x_n^T u_j$$

② Find the optimal b_j

$$J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \sum_{i=1}^{M} z_{ni} u_i - \underline{\sum_{i=M+1}^{D} b_i u_i} \|^2$$

$$J = \frac{1}{N} \sum_{n=1}^{N} [b_j^2 - 2 b_j u_j^T x_n] + \text{constant}$$

$$\frac{\partial J}{\partial b_j} = 2 b_j - \frac{2}{N} \sum_{n=1}^{N} x_n^T u_j = 0$$

$$\Downarrow$$

$$b_j = \underline{\underline{\frac{1}{N} \sum_{n=1}^{N}}} x_n^T u_j = \bar{x}^T u_j$$

③ Find $\{u_i\}$ $\quad z_{nj} = x_n^T u_j$; $b_j = \bar{x}^T u_j$

$$\tilde{x}_n = \sum_{i=1}^{M} (x_n^T u_i) u_i + \sum_{i=M+1}^{D} (\bar{x}^T u_i) u_i$$

$$x_n = \sum_{i=1}^{D} (x_n^T u_i) u_i$$

$$J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \tilde{x}_n \|^2$$

$$= \frac{1}{N} \sum_{n=1}^{N} \left\| \sum_{i=M+1}^{D} (x_n - \bar{x})^T u_i u_i \right\|^2 \qquad S = \frac{1}{N} \bar{X}^T \bar{X}$$

$$= \frac{1}{N} \sum_{n=1}^{N} \sum_{i=M+1}^{D} (\underbrace{x_n^T u_i}_{\text{projected data}} - \underbrace{\bar{x}^T u_i}_{\text{mean projected data}})^2 = \sum_{i=M+1}^{D} u_i^T S u_i$$

In the minimum error formulation :

$$\min_{u_i} \sum_{i=M+1}^{D} u_i^T S u_i$$

In the maximum variance formulation ;

$$\max_{u_i} \sum_{i=1}^{M} u_i^T S u_i$$