Proof for PCA for Lecture 5

October 15, 2019

1 Proof for PCA

Suppose that the data matrix is $X = \begin{bmatrix} x_1^T \\ \cdots \\ x_n^T \end{bmatrix}$, i.e., each row is a data vector (already centered by mean).

Let $U = [u_1, \cdots, u_k]$ be the projection matrix, whose columns are the basis of the low-dimensional space.

Then, the coordinate of data in the projected space is $Z = XU$, i.e., each row is the coordinate of a data point.

Using the coordinates and the basis, we can reconstruct the data in the original space as $X' = ZU^T = XUU^T$.

So the sum of projection error is $\|X' - X\|_{fro}^2$, where the $\|\cdot\|_{fro}^2$ is the square of the Frobenius norm of a matrix, defined as the sum of squares of all elements in X.

Our goal is to solve the optimization problem:

$$\begin{aligned} &\text{minimize}_U \quad \|XUU^T - X\|_{fro}^2 \\ &\text{subject to} \quad U^T U = I \end{aligned} \tag{1}$$

The Lagrangian of the problem is

$$\begin{aligned} L &= \|XUU^T - X\|_{fro}^2 - \text{tr}(\Lambda(U^T U - I)) \\ &= \text{tr}[(XUU^T - X)(XUU^T - X)^T] - \text{tr}[\Lambda(U^T U - I)] \end{aligned} \tag{2}$$

here, $\text{tr}(\cdot)$ is the trace of a matrix.

Taking gradient w.r.t U,

$$\nabla_U L = -2X^T XU + 2U\Lambda \tag{3}$$

Set $\nabla_U L = 0$, and we get $X^T XU = U\Lambda$.

Because X^X is a symmetric matrix, it is diagnalizable. Additionally, assume that $X = P\Sigma Q^T$ by SVD, then $X^T X = Q\Sigma^2 Q^T$, which implies that all the eigenvalues of $X^T X$ must be a squared number, thus positive.

So U's columns are the eigenvectors of $X^T X$, which are also the right singular vectors of X.

Substitute $X^T XU = U\Lambda$ into the objective, we will have the projection error to be $\text{tr}(X^T X) - \sum_{i=1}^k \lambda_i$. By the positivity of λ_i's, we should choose the largest k eigenvalues to minimize the projection error, and choose the corresponding eigenvectors as the basis.