Two Perspectives of PCA

Yanhua Huang

Mar 2016

PCA (Principal Components Analysis) is a well-known dimension reduction algorithm. One perspective of its motivation is finding a linear subspace such that the variance of projected data is maximum, from which we can get the solution from the covariance matrix.

We discuss anther perspective that considers projection loss. Given N data x_i where $x_i \in \mathbb{R}^D$. Mark the projected data is $\hat{x}_i \in \mathbb{R}^M$ where $M < D$. We wish the empirical error $\mathcal{L} = \mathbb{E}[||x_i - \hat{x}_i||_2^2]$ is minimum.

With a collection of basis u_j in $\mathbb{R}^D$, we can get

$$x_i = \sum_{j=1}^{D} \langle x_i, u_j \rangle u_j \tag{1}$$

and

$$\hat{x}\_i = \sum_{j=1}^{M} z_{ij} u_j + \sum_{j=M+1}^{D} b_j u_j, \tag{2}$$

where b_j and z_{ij} is unknown variables. Make the partial derivative equal to zero, we can get $b_j = \langle \bar{x}, u_j \rangle$ and $z_{ij} = \langle x_i, u_j \rangle$. Take b_j and z_{ij} into $\mathcal{L}$ gets

$$\mathcal{L} = \mathbb{E}[|| \sum_{j=M+1}^{D} \langle x_i - \bar{x}, u_j \rangle u_j ||_2^2]. \tag{3}$$

Recall that the basis satisfies that $\langle u_j, u_j \rangle = 1$. With Lagrange multipliers, then $\lambda u_j = C u_j$ where C is the covariance matrix. It is clearly that Lagrange multipliers are the eigenvalues of C so that the largest M eigenvalues will reach the goal.