Component Analysis

Thomas Bayes

April 26, 2018

Principle Component Analysis I

Find one direction which minimize reconstruction error,

$$\text{minimize}\, ||\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{W}^\top||^2$$

$$\begin{aligned}
& ||\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{W}^\top||^2 \\
= \;& \mathrm{tr}((\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{W}^\top)(\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{W}^\top)^\top) \\
= \;& \mathrm{tr}(\mathbf{X}\mathbf{X}^\top) - 2\mathrm{tr}(\mathbf{X}\mathbf{W}\mathbf{W}^\top\mathbf{X}^\top) + \mathrm{tr}(\mathbf{X}\mathbf{W}\mathbf{W}^\top\mathbf{W}\mathbf{W}^\top\mathbf{X}^\top) \\
= \;& \mathrm{const} - \mathrm{tr}(\mathbf{X}\mathbf{W}\mathbf{W}^\top\mathbf{X}^\top) \\
= \;& \mathrm{const} - \mathrm{tr}(\mathbf{W}^\top\mathbf{X}^\top\mathbf{X}\mathbf{W}) \\
= \;& \mathrm{const} - \mathrm{const} \cdot \mathrm{tr}(\mathbf{W}^\top\mathbf{\Sigma}\mathbf{W})
\end{aligned}$$

This is equivalent to maximizing the total variance of $\mathbf{X}$ on the projected space $\mathbf{X}\mathbf{W}$.

Principle Component Analysis II

Now we show that the solution is the first k eigenvectors of $\mathbf{\Sigma}$. Assume $\mathbf{X}_{n\times d}$ is full rank d, SVD decomposition of $\mathbf{X}_{n\times d} = \mathbf{U}_{n\times n}\mathbf{S}_{n\times d}\mathbf{V}_{d\times d}^{\top}$.

$$\mathbf{\Sigma} = \mathbf{X}^{\top}\mathbf{X} = \mathrm{const}\cdot(\mathbf{V}\mathbf{S}^{\top}\mathbf{U}^{\top}\mathbf{U}\mathbf{S}\mathbf{V}^{\top}) = \mathrm{const}\cdot(\mathbf{V}\mathbf{S}^{\top}\mathbf{S}\mathbf{V}^{\top})$$

We denote $\mathbf{S}^{\top}\mathbf{S}$ as $\mathbf{\Lambda}$, $\mathrm{tr}(\mathbf{W}^{\top}\mathbf{\Sigma}\mathbf{W}) = \mathrm{tr}(\mathbf{W}^{\top}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\top}\mathbf{W})$. Note that $\mathbf{R} = \mathbf{V}^{\top}\mathbf{W}$ is also a orthonormal matrix. So

$$\mathrm{tr}(\mathbf{W}^{\top}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\top}\mathbf{W}) = \mathrm{tr}(\mathbf{R}^{\top}\mathbf{\Lambda}\mathbf{R}) = \sum_{i=1}^{d}\lambda_i\sum_{j=1}^{k}\mathbf{R}_{ij}^2 \leq \sum_{i=1}^{k}\lambda_i$$

We just pick the largest k eigen values and get the result.

Principle Component Analysis III

So from minimizing reconstruction error using a linear combination subspace, we have showed that it is equivalent to maximizing the total variance of the projection on linear subspace and the correponding subspace is formed of exactly the k largest eigenvectors.(Note that we do not use any other assumpition except linearity.)

Independent Component Analysis I

In PCA we mentioned earlier, we want to find a linear combination subspace for which the reconstruction error is minimized.

Linear Discriminant Analysis I