# Principle Component Analysis (PCA)

May 30, 2022

Consider a $n \times 1$ random column vector $\mathbf{x}$ with zero mean (i.e. $\mathbb{E}[\mathbf{x}] = \mathbf{0}$) Our aim is to find the linear combination of its component, such that the variance is maximised.

First of all, we denote a real number $y$ as the linear combination of $\mathbf{x}$ elements with weight determined by $\mathbf{w}$, a $n \times 1$ column vector

$$y = \mathbf{w}^T \mathbf{x} \tag{1}$$

Variance of $y$ is given by

$$\begin{aligned}
\sigma_y^2 &= \mathbb{E}[y^2] - \mathbb{E}[y]^2 \\
&= \mathbb{E}[y^2] \qquad (\because \mathbb{E}[\mathbf{x}] = 0) \\
&= \mathbf{w}^T \mathbb{E}[\mathbf{x}\mathbf{x}^T]\mathbf{w} \\
&= \mathbf{w}^T \mathbf{C_x} \mathbf{w} \qquad (\because \mathbb{E}[\mathbf{x}] = 0)
\end{aligned} \tag{2}$$

where $\mathbf{C_x}$ is the covariance matrix of $\mathbf{x}$

There is an issue to maximise the variance of $y$, $\sigma_y^2$. As you can see, we if scale the weighting vector by a factor of $\alpha > 1$ (i.e. $\mathbf{w} \to \alpha\mathbf{w}$), $\sigma_y^2$ would be increased by $\alpha^2$. Therefore, it makes more sense to consider a normalised weighting vector with unit length. Hence, our constraint would be $||\mathbf{w}||^2 = 1$.

For constrained optimisation with equality, we apply Lagrange multiplier method and our objective function $J$ is given by

$$\begin{aligned}
J &\equiv \mathbf{w}^T \mathbf{C_x} \mathbf{w} - \lambda\big(||\mathbf{w}||^2 - 1\big) \\
&= \sum_{i,j=1}^{n} w_i (C_x)_{ij} w_j - \lambda\big(\sum_{i=1}^{n} w_i^2 - 1\big)
\end{aligned} \tag{3}$$

Taking derivative of $J$ with respective to $w_k$, we have

$$\begin{aligned}
\frac{\partial J}{\partial w_k} &= 0 \\
\sum_{j=1}^{n} (C_x)_{kj} w_j + \sum_{i=1}^{n} w_i (C_x)_{ik} - 2\lambda w_k &= 0 \\
\sum_{j=1}^{n} (C_x)_{kj} w_j + \sum_{i=1}^{n} w_i (C_x)_{ki} - 2\lambda w_k &= 0 \qquad (\because \mathbf{C_x}^T = \mathbf{C_x}) \\
2\sum_{i=1}^{n} w_i (C_x)_{ki} &= 2\lambda w_k \\
(\mathbf{C_x}\mathbf{w})_k &= \lambda w_k \qquad \text{for all k} = 1, 2, ..., \text{n}
\end{aligned} \tag{4}$$

In other words, the weight $\mathbf{w}$ that maximises the objective function $J$ is governed by

$$\mathbf{C_x}\mathbf{w} \;=\; \lambda\mathbf{w} \tag{5}$$

which implies that the optimal $\mathbf{w}$ is an eigen-vector of the covariance matrix $\mathbf{C_x}$, and the Lagrange multiplier is an eigen-value of the covariance matrix $\mathbf{C_x}$.

The maximised variance of $y$ is given by

$$\begin{aligned}\sigma_y^2 &= \mathbf{w}^T \mathbf{C_x} \mathbf{w} \\ &= \lambda \mathbf{w}^T \mathbf{w} \\ &= \lambda \qquad (\because ||\mathbf{w}|| = 1)\end{aligned} \tag{6}$$

which is an eigen-value of the covariance matrix $\mathbf{C_x}$.

For a covariance matrix $\mathbf{C_x}$ which is a $n \times n$ semi-definite matrix, it has $n$ non-negative eige-values $\lambda_i$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} \geq \lambda_n \geq 0 \tag{7}$$

To truly maximise the variance of $y$, we can choose an eigen-vector $\mathbf{w}$ that corresponds to the largest eigen-value $\lambda_1$. We call this eigen-vector $\mathbf{w}_1$, the first principle component.