1 Principal Component Analysis(PCA)

1.1 Settings

Let $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^p$ be a sample of n observations of p dimensional vectors. The first principal component of this sample is a real variable transformed from X

$$\mathbf{z}_1 = \mathbf{X}\mathbf{a}_1^T$$

where the vector $\mathbf{a}_1 = (a_{11}, a_{21}, \dots, a_{p1}) \in \mathbb{R}^p$ is chosen so that $\|\mathbf{a}_1\|_2 = 1$ and $var[\mathbf{z}_1]$ is maximized.
The k^{th} principal component of the sample is

$$\mathbf{z}_k = \mathbf{X}\mathbf{a}_k^T, \quad k = 2, \dots, p$$

where the vector $\mathbf{a}_k = (a_{1k}, a_{2k}, \dots, a_{pk}) \in \mathbb{R}^p$ is chosen so that $\|\mathbf{a}_k\|_2 = 1$, $var[\mathbf{z}_k]$ is maximized and $cov[\mathbf{z}_k, \mathbf{z}_l] = 0$ for $k > l \geq 1$.

Geometrically, at first step, PCA finds a unit vector $\mathbf{a}_1 \in \mathbb{R}^p$ that the projections of sample points on it has maximum variance. $\mathbf{z}_1$ is actually a variable of projection length of sample points on $\mathbf{a}_1$. Secondly, PCA finds the second component $\mathbf{z}_2$ in the same logic but with further restriction that $\mathbf{z}_2$ has no relation with $\mathbf{z}_1$, i.e. $cov[\mathbf{z}_2, \mathbf{z}_1] = 0$.

1.2 Computations

To find $\mathbf{a}_1$, first note that

$$\begin{aligned}
var[\mathbf{z}_1] =& \mathrm{E}[\mathbf{z}_1^2] - \mathrm{E}[\mathbf{z}_1]^2 \\
=& \sum_{i,j=1}^{p} a_{i1}a_{j1}\mathrm{E}[X_iX_j] - \sum_{i,j=1}^{p} a_{i1}a_{j1}\mathrm{E}[X_i]\mathrm{E}[X_j] \\
=& \sum_{i,j=1}^{p} a_{i1}a_{j1}cov[X_i, X_j] \\
=& (a_{11}, a_{21}, \dots, a_{p1}) \begin{pmatrix} cov[X_1, X_1] & cov[X_1, X_2] & \dots & cov[X_1, X_p] \\ cov[X_2, X_1] & cov[X_2, X_2] & \dots & cov[X_2, X_p] \\ \vdots & \vdots & \ddots & \vdots \\ cov[X_p, X_1] & cov[X_p, X_2] & \dots & cov[X_p, X_p] \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{pmatrix} \\
=& \mathbf{a}_1\Sigma\mathbf{a}_1^T
\end{aligned}$$

where Σ is the covariance matrix of X. The problem becomes

$$\max(\mathbf{a}_1\Sigma\mathbf{a}_1^T) \quad s.t. \quad \|\mathbf{a}_1\|_2 = 1$$

Let λ be Lagrange multiplier and let $g(\mathbf{a}_1) = \mathbf{a}_1\Sigma\mathbf{a}_1^T - \lambda(\mathbf{a}_1\mathbf{a}_1^T - 1)$. Then we need

$$\nabla g(\mathbf{a}_1) = \mathbf{a}_1\Sigma^T - \lambda\mathbf{a}_1 = (\Sigma\mathbf{a}_1^T - \lambda\mathbf{a}_1^T)^T = 0$$

Therefore, $\mathbf{a}_1^T$ is the eigenvector of Σ corresponding to eigenvalue $\lambda \equiv \lambda_1$. Furthermore, we have maximized $var[\mathbf{z}_1] = \mathbf{a}_1\Sigma\mathbf{a}_1^T = \lambda_1$, this means λ_1 is the largest eigenvalue of Σ.

For the second component, note that $cov[\mathbf{z}_2, \mathbf{z}_1] = \mathbf{a}_2\Sigma\mathbf{a}_1^T = \mathbf{a}_2\lambda_1\mathbf{a}_1^T = 0$. For another multipliers ϕ and λ and another $g(\mathbf{a}_2) = \mathbf{a}_2\Sigma\mathbf{a}_2^T - \lambda(\mathbf{a}_2\mathbf{a}_2^T - 1) - \phi\lambda_1\mathbf{a}_2\mathbf{a}_1^T$, we need

$$\nabla g(\mathbf{a}_2) = (\Sigma\mathbf{a}_2^T - \lambda\mathbf{a}_2^T - \phi\lambda_1\mathbf{a}_1^T)^T = 0$$

Multiply $\mathbf{a}_1$ both side, we get

$$\begin{aligned}&\mathbf{a}_1\Sigma\mathbf{a}_2^T - \mathbf{a}_1\lambda\mathbf{a}_2^T - \phi\lambda_1\mathbf{a}_1\mathbf{a}_1^T\\ =&(\mathbf{a}_2\Sigma\mathbf{a}_1)^T - \lambda(\mathbf{a_2}\mathbf{a}_1)^T - \phi\lambda_1 = 0\end{aligned}$$

So ϕ must be zero. Hence

$$\Sigma\mathbf{a}_2^T = \lambda\mathbf{a}_2^T$$

$\mathbf{a}_2^T$ is also an eigenvector and has $\lambda \equiv \lambda_2$ as its eigenvalue. And, again, $var[\mathbf{z}_2] = \mathbf{a}_2\Sigma\mathbf{a}_2^T = \lambda_2$ is the second largest eigenvalue.

In general,

$$var[\mathbf{z}_k] = \mathbf{a}_k\Sigma\mathbf{a}_k^T = \lambda_k$$

The k^{th} largest eigenvalue of Σ is the variance of the k^{th} PC.