# 19 Population Principal Components Analysis

Suppose we have $m$ random variables $X_1, X_2, \ldots, X_m$. We wish to identify a set of weights $w_1, w_2, \ldots, w_m$ that maximizes

$$\operatorname{Var}\left(w_1 X_1 + w_2 X_2 + \cdots + w_m X_m\right).$$

However, this is unbounded, so we need to constrain the weights. It turns out that constraining the weights so that

$$\|\boldsymbol{w}\|_2^2 = \sum_{i=1}^m w_i^2 = 1$$

is both interpretable and mathematically tractable.

Therefore we wish to maximize

$$\operatorname{Var}\left(w_1 X_1 + w_2 X_2 + \cdots + w_m X_m\right)$$

subject to $\|\boldsymbol{w}\|_2^2 = 1$. Let $\boldsymbol{\Sigma}$ be the $m \times m$ population covariance matrix of the random variables $X_1, X_2, \ldots, X_m$. It follows that

$$\operatorname{Var}\left(w_1 X_1 + w_2 X_2 + \cdots + w_m X_m\right) = \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}.$$

Using a Lagrange multiplier, we wish to maximize

$$\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} + \lambda(\boldsymbol{w}^T \boldsymbol{w} - 1).$$

Differentiating with respect to $\boldsymbol{w}$ and setting to $\boldsymbol{0}$, we get $\boldsymbol{\Sigma} \boldsymbol{w} - \lambda \boldsymbol{w} = 0$ or

$$\boldsymbol{\Sigma} \boldsymbol{w} = \lambda \boldsymbol{w}.$$

For any such $\boldsymbol{w}$ and $\lambda$ where this holds, note that

$$\operatorname{Var}\left(w_1 X_1 + w_2 X_2 + \cdots + w_m X_m\right) = \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} = \lambda$$

so the variance is $\lambda$.

The eigendecompositon of a matrix identifies all such solutions to $\boldsymbol{\Sigma} \boldsymbol{w} = \lambda \boldsymbol{w}$. Specifically, it calculates the decompositon

$$\boldsymbol{\Sigma} = \boldsymbol{W} \boldsymbol{\Lambda} \boldsymbol{W}^T$$

where $\boldsymbol{W}$ is an $m \times m$ orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix with entries $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$.

The fact that $\boldsymbol{W}$ is orthogonal means $\boldsymbol{W}\boldsymbol{W}^T = \boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}$.

The following therefore hold:

- For each column $j$ of $\boldsymbol{W}$, say $\boldsymbol{w}_j$, it follows that $\boldsymbol{\Sigma} \boldsymbol{w}_j = \lambda_j \boldsymbol{w}_j$
- $\| \boldsymbol{w}_j \|_2^2 = 1$ and $\boldsymbol{w}_j^T \boldsymbol{w}_k = \boldsymbol{0}$ for $\lambda_j \neq \lambda_k$
- $\operatorname{Var}(\boldsymbol{w}_j^T \boldsymbol{X}) = \lambda_j$
- $\operatorname{Var}(\boldsymbol{w}_1^T \boldsymbol{X}) \geq \operatorname{Var}(\boldsymbol{w}_2^T \boldsymbol{X}) \geq \cdots \geq \operatorname{Var}(\boldsymbol{w}_m^T \boldsymbol{X})$
- $\boldsymbol{\Sigma} = \sum_{j=1}^m \lambda_j \boldsymbol{w}_j \boldsymbol{w}_j^T$
- For $\lambda_j \neq \lambda_k$,

$$\operatorname{Cov}(\boldsymbol{w}_j^T \boldsymbol{X}, \boldsymbol{w}_k^T \boldsymbol{X}) = \boldsymbol{w}_j^T \boldsymbol{\Sigma} \boldsymbol{w}_k = \lambda_k \boldsymbol{w}_j^T \boldsymbol{w}_k = \boldsymbol{0}$$

The $j$th **population principal component** (PC) of $X_1, X_2, \ldots, X_m$ is

$$\boldsymbol{w}_j^T \boldsymbol{X} = w_{1j} X_1 + w_{2j} X_2 + \cdots + w_{mj} X_m$$

where $\boldsymbol{w}_j = (w_{1j}, w_{2j}, \ldots, w_{mj})^T$ is column $j$ of $\boldsymbol{W}$ from the eigendecomposition

$$\boldsymbol{\Sigma} = \boldsymbol{W} \boldsymbol{\Lambda} \boldsymbol{W}^T.$$

The column $\boldsymbol{w}_j$ are called the **loadings** of the $j$th principal component. The **variance explained** by the $j$th PC is $\lambda_j$, which is diagonal element $j$ of $\boldsymbol{\Lambda}$.