# Principle Component Analysis (PCA)

Suppose that $x_1, x_2, x_3, \ldots, x_m \in \mathbb{R}^n$ are a sequence of identically and independently distributed random vectors with mean

$$\mu = \mathbb{E}[x] = (\mathbb{E}[x^{(1)}], \ldots, \mathbb{E}[x^{(n)}])^T \in \mathbb{R}^n$$

and covariance matrix

$$C_{ij} = \mathbb{E}\left[(x^{(i)} - \mu^{(i)})(x^{(j)} - \mu^{(j)})\right] \Rightarrow C = \mathbb{E}\left[(x-\mu)(x-\mu)^T\right]$$

## What does PCA aim to do?

__View 1:__ Choose a new coordinate system for "feature space" $\mathbb{R}^n$ in which the data or random variable has independent entries/coords.

old features

new features $\quad y_j = U^T (x_j - \mu)$ mean correction

Find this to make $y_j$ entries i.i.d.

mean zero $\Rightarrow \mathbb{E}[y] = \mathbb{E}[U^T(x_j - \mu)] = U^T \mathbb{E}(x_j) - U^T \mu = 0$

$$\Rightarrow \mathbb{E}[yy^T] = \mathbb{E}[U^T(x-\mu)(x-\mu)^T U]$$

$$= U^T \mathbb{E}[(x-\mu)(x-\mu)^T] U$$

$$= U^T C U$$

Q: How should we choose the transform $U$?

$i = 1, \dots, n$ $\quad [v_1 \cdots v_n]$ $\quad \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

$$C v_i = \lambda_i v_i \Rightarrow CV = V\Lambda$$

$U = V$

$$\Rightarrow \mathbb{E}[yy^T] = V^T C V = V^T V \Lambda$$

<u>**Key Point:**</u> Covariance Matrix $C$ is symmetric

$$C_{ij} = \mathbb{E}[(x^{(i)} - \mu^{(i)})(x^{(j)} - \mu^{(j)})] = C_{ji} \Rightarrow C = C^T$$

$\Rightarrow$ $V$ is orthogonal so $V^T V = I$.

$$\Rightarrow \mathbb{E}[yy^T] = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Note that $\lambda_1 \geq 0, \dots, \lambda_n \geq 0$ and $C$ is pos. def.,

$$U = V \underbrace{\begin{bmatrix} 1/\sqrt{\lambda_1} & & \\ & \ddots & \\ & & 1/\sqrt{\lambda_n} \end{bmatrix}}_{\sqrt{\Lambda^{-1}}} \Rightarrow U^T C U = \sqrt{\Lambda^{-1}}\, V^T C V \sqrt{\Lambda^{-1}}$$
$$= \sqrt{\Lambda^{-1}}\, \Lambda \sqrt{\Lambda^{-1}}$$
$$= I$$

## Principal Components

After transformation with $U = V =$ eigenvecs of $C$,

old features

new features $\quad y_j = U^T (x_j - \mu)$ mean correction

Find this to make $y_j$ entries i.i.d.

$$\Rightarrow \quad \mathbb{E}[y] = 0 \quad \text{and} \quad \mathbb{E}[yy^T] = \Lambda$$

Convention: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

The $k^{th}$ **principal component** of $x$ is

$$y^{(k)} = u_k^T (x - \mu)$$

"Coordinate of $x - \mu$ along $u_k$."

View 2: The leading principle component(s) capture the "most" variance in the distribution of random vector in the following sense.

$$\underset{\|v\|=1}{\operatorname{argmax}} \mathbb{E}\left[v^T(x-\mu)(x-\mu)v^T\right] = \underset{\|v\|=1}{\operatorname{argmax}}\ v^T \mathbb{E}\left[(x-\mu)(x-\mu)^T\right] v$$

$\frac{v^TCv}{v^Tv}$ = "Rayleigh Quotient"

$$= \underset{\|v\|=1}{\operatorname{argmax}}\ v^T C v$$

The Rayleigh Quotient $\frac{v^TCv}{v^Tv}$ is maximized when $v = u_1$ (the leading eigenvector of $C$).

pf sketch

$$v^TCv = v^T \underbrace{\left(\sum_{j=1}^{n} \lambda_j u_j u_j^T\right)}^{C = U\Lambda U^T} v$$

$$= \sum_{j=1}^{n} \lambda_j (v^Tu_j)(u_j^Tv) = \sum_{j=1}^{n} \lambda_j |v^Tu_j|^2$$

$$\leq \lambda_1 \sum_{j=1}^{n} |v^Tu_j|^2 = \lambda_1 \underbrace{\|v\|^2}_{v^Tv}$$

$$u_1^TCu_1 = \sum_{j=1}^{n} \lambda_j |u_1^Tu_j|^2 = \lambda_1 \quad \checkmark$$

So $u_1$ maximizes the variance of $y^{(1)}$, the first principle component of $x$. The variance of $y^{(1)}$ is $\lambda_1$.

We can use similar principle to "find" the remaining rows of the transformation.

$$u_2 = \underset{\substack{\|v\|=1 \\ v^T u_1 = 0}}{\operatorname{argmax}} \; v^T C v$$

pf sketch

$$v^T C v = \sum_{j=1}^{n} \lambda_j |v^T u_j|^2 = \sum_{j=2}^{n} \lambda_j |v^T u_j|$$

$$\leq \lambda_2 \sum_{j=2}^{n} |v^T u_j|^2 \leq \lambda_2 (\overbrace{v^T v}^{=1})$$

$$u_2^T C u_2 = \sum_{j=1}^{n} \lambda_j |u_2^T u_j|^2 = \lambda_2$$

$\Rightarrow$ $u_2$ maximizes the variance of $y^{(2)}$, subject to $\|v\| = 1$, and $v^T u_1 = 0$.

$$u_k = \underset{\|v\|=1,\, v \in V_{k-1}^{\perp}}{\operatorname{argmax}} \; v^T C v \qquad V_{k-1}^{\perp} = \operatorname{span}\{u_1, u_2, \ldots, u_{k-1}\}^{\perp}.$$

## Courant-Fisher-Weyl Min-Max Principle

We can formulate the eigenvectors themselves as extrema of the Rayleigh-Quotient:

$$\lambda_k = \min_{\substack{\dim(M) \\ = k}} \max_{\substack{v \in M \\ \|v\|=1}} v^T C v \qquad (1 \leq k \leq n)$$