"PCA as maximizing variance"

Consider projection onto a line given by a unit vector u

For a data point x_i, the projection onto line along u is $(x_i^T u)u$

Dataset $\mathcal{D} = \{x_1 --- x_n\}$ $x_i \in \mathbb{R}^d$, Mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

x_i's projection is $(x_i^T u)u$, Mean projected value $(\bar{x}^T u)u$

Variance is $(x_i^T u - \bar{x}^T u)^2$

Summing the variance over all points, we obtain

$$\boxed{\frac{1}{n}\sum_{i=1}^{n}(x_i^T u - \bar{x}^T u)^2} \longrightarrow \text{maximize this over } u$$

Notice that $$\frac{1}{n}\sum_{i=1}^{n}\left((x_i - \bar{x})^T u\right)^2 = \frac{1}{n}\sum_{i=1}^{n} u^T (x_i - \bar{x})(x_i - \bar{x})^T u$$

$$= u^T C u, \text{ where } \quad C = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$$

Goal: $\max_u u^T C u$ s.t. $u^T u = 1$

Using Lagrangian as before, it can be argued that the maximizer of $u^T C u$ is an eigenvector of C corresponding to the largest eigenvalue of C.

A Calculus argument:

$$\max_u \quad u^T C u \quad \text{s.t. } u^T u = 1 .$$

Or, equivalently, $\displaystyle \max_u \frac{u^T C u}{u^T u}$

Let $u = \left(u^{(1)}, \ldots, u^{(d)} \right)$

$$\frac{\partial\, u^T u}{\partial u^{(i)}} = \frac{\partial}{\partial u^{(i)}} \left({u^{(1)}}^2 + \cdots + {u^{(d)}}^2 \right) = 2\, u^{(i)}$$

$$\frac{\partial\, u^T C u}{\partial u^{(i)}} = \frac{\partial}{\partial u^{(i)}} \left(\sum_{i=1}^{d} \sum_{j=1}^{d} C_{ij}\, u^{(i)} u^{(j)} \right) = 2 \sum_j C_{ij}\, u^{(j)} = 2\left((Cu)^{(i)} \right)$$

(i,j)th entry of matrix C

ith co-ordinate of u

jth co-ordinate of vector u

$$\frac{\partial}{\partial u^{(i)}} \left(\frac{u^T C u}{u^T u} \right) = u^T u \, 2\, (Cu)^{(i)} - (u^T C u)\, 2\, u^{(i)} = 0 \quad \text{—} (*)$$

in vector form, $(*) \Leftrightarrow$ $\quad u^T u \; Cu = (u^T C u)\, u \iff Cu = \left(\frac{u^T C u}{u^T u} \right) u$

(or) $Cu = \lambda u$

So, the maximizer of $\frac{u^T C u}{u^T u}$ is an eigenvector of C and $\max_u \frac{u^T C u}{u^T u} = \lambda$

$\downarrow$ an eigenvalue of C

To maximize $\frac{u^T C u}{u^T u}$, choose λ to be the largest eigenvalue of C & let u be the corresponding eigenvector.

The logic can be extended to the case when $m > 1$

For instance, in 2d-case, we need u_1, u_2 s.t. $\|u_1\|=1$, $\|u_2\|=1$, $u_1^T u_2 = 0$ and projected variance is maximized

It can be shown that picking the eigenvectors corresponding to the top-2 eigenvalues would maximize projected variance & so on.

In general, from a "maximizing variance" viewpoint, PCA does pick top-m eigenvalues of C, with corresponding eigenvectors $\{u_1, \ldots u_m\}$ $u_i \rightarrow$ Principal directions, Projected values $\rightarrow$ Principal components

Revisiting the example: Data: $\left\{ \underbrace{\begin{pmatrix} -1 \\ -1 \end{pmatrix}}_{x_1}, \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{x_2}, \underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{x_3} \right\}$

Projections: $\left\{ \underbrace{\begin{pmatrix} -1 \\ -1 \end{pmatrix}}_{\tilde{x}_1}, \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{\tilde{x}_2}, \underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{\tilde{x}_3} \right\}$ Recall $u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$$x_1^T u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = -\sqrt{2}$$

$$x_2^T u_1 = 0, \quad x_3^T u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \sqrt{2}$$

$$\text{Projected variance} = \frac{1}{3} \left((x_1^T u_1)^2 + (x_2^T u_1)^2 + (x_3^T u_1)^2 \right) = \frac{4}{3}$$

This solution can be arrived at more directly. (How?)

Recall eigenvalues of C are $\frac{4}{3}$ and 0

$\max_{u_1} u_1^T C u_1$ s.t. $u_1^T u_1 = 1$ is achieved by $u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ corresponding to eigenvalue $\frac{4}{3}$

Or, equivalently, Projected variance $= \lambda_1 = \frac{4}{3}$.