# PCA : Principal Components Analysis Derivation.

## Linear Algebra Review

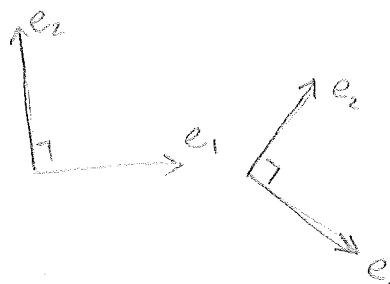
Notation : $X^t$ the transpose of a matrix $X$.

$\langle x, y \rangle$ the dot product between $x$ and $y$.

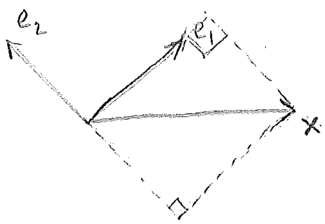
"$x$ is a $p$-vector" means "$x$ is a vector with $p$ components." no hyphen!

An orthogonal basis $\{e_1, e_2, \dots, e_p\}$ of $p$-vectors is a collection of $p$ vectors satisfying:

$$\langle e_i, e_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$



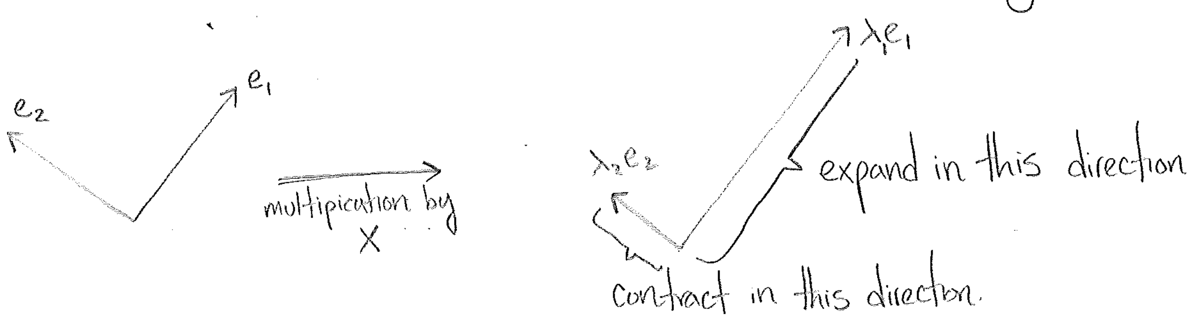

If $x$ is **any** vector : $\underbrace{x = \sum_{i=1}^{p} \langle x, e_i \rangle e_i}_{\text{Easy to reconstruct any vector } x \text{ from its projections onto } e_i\text{s.}}$




An eigenvector of a matrix $X$ is a vector $v$ satisfying

$$Xv = \lambda v$$

this is a number, it is called an eigenvalue of $X$.



A square matrix $X$ is called <u>symmetric</u> if $X^t = X$.

<u>Sophisticated Technology # 1</u>

If $X$ is a symmetric matrix then there is an orthogonal basis, each vector of which is an eigenvector of $X$.

$\{e_1, e_2, \dots, e_p\}$ an orthogonal basis with $Xe_i = \lambda_i e_i$.

eigenvalues!

A symmetric maxrix is called <u>positive semi-definate</u> if $v^t X v$ for **every** vector $v$.

<u>Sophisticated Technology # 2</u>

All the eigenvalues of a positive semi-definate are non-negative numbers.

Ok. **LETS GO!**

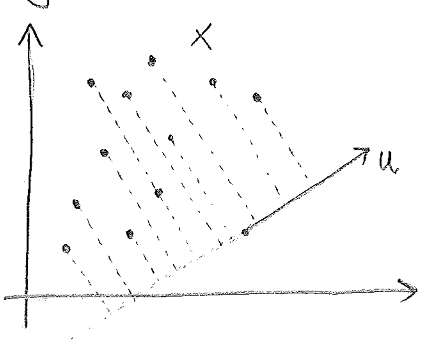
**<u>Setup</u>**

$X$ is a dataset ($n \times p$ matrix)

$X_i$ is a row in $X$.

$u$ is a unit vector ($p$-vector)

**<u>Problem</u>**: Find the vector $u$ that maximizes the variance of $X$ projected onto $u$.




<u>Observations</u>: Translating $X$ does not change the variance of the projection onto a fixed vector $u$.

$\Longrightarrow$ Can assume the mean of $X$ is at the orgin.

"$X$ is centered."

Projection of $x_i$ onto $u$: $\langle x_i, u \rangle$

Mean of the projections: $\frac{1}{n}\sum_{i=1}^{n} \langle x_i, u \rangle = \left\langle \frac{1}{n}\sum_{i=1}^{n} x_i, u \right\rangle = \langle 0, u \rangle = 0.$

Variance of the projections: $\frac{1}{n}\sum_{i=1}^{n} \langle x_i, u \rangle^2 \underset{\uparrow}{=} \frac{1}{n} u^t X^t X u$

Wait, what?

column vector

$$Xu = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} u = \begin{pmatrix} \langle x_1, u \rangle \\ \langle x_2, u \rangle \\ \vdots \\ \langle x_n, u \rangle \end{pmatrix}$$

$$u^t X^t = (Xu)^t = \left( \langle x_1, u \rangle, \langle x_2, u \rangle, \cdots, \langle x_n, u \rangle \right)$$

So: $$u^t X^t X u = \left( \langle x_1, u \rangle, \cdots, \langle x_n, u_n \rangle \right) \begin{pmatrix} \langle x_1, u \rangle \\ \vdots \\ \langle x_n, u \rangle \end{pmatrix} = \sum_{i=1} \langle x_i, u \rangle^2$$

✓

Problem Restatement #1: Find the unit vector $u$ maximizing $u^t \Omega u$, where $\Omega = \frac{1}{n} X^t X$ is the sample covariance matrix.

Facts:

① $\Omega$ is symmetric:
$$\Omega^t = \frac{1}{n}(X^t X)^t = \frac{1}{n} X^t X = \Omega$$

② $\Omega$ is positive semi-definate:
$$v^t \Omega v = \sum_{i=1}^{n} \langle x_i, v \rangle^2 \geq 0$$

This means:

$\Omega$ has $p$ orthogonal eigenvectors $\{e_1, e_2, \cdots, e_p\}$

The associated eigenvalues are all non-negative numbers:
$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$$

The vector we are seeking can be written as a linear combination of the e's

$$u = \sum_{i=1}^{p} a_i e_i$$

Because $u$ is a unit vector, the a's satisfy a constraint

$$1 = \langle u, u \rangle = \sum_{i,j=1}^{p} a_i a_j \langle e_i, e_j \rangle = \sum_{i=1}^{p} a_i^2$$

Let's plug this into the thing we are maximizing:

$$\left.\begin{aligned} u^t \Omega u &= u^t \Omega (a_1 e_1 + a_2 e_2 + \cdots + a_p e_p) \\ &= u^t (a_1 \Omega e_1 + a_2 \Omega e_2 + \cdots + a_p \Omega e_p) \\ &= u^t (a_1 \lambda_1 e_1 + a_2 \lambda_2 e_2 + \cdots + a_p \lambda_p e_p) \end{aligned}\right\} \text{e's are eigenvectors of } \Omega$$

$$= (a_1 e_1^t + a_2 e_2^t + \cdots + a_p e_p^t)(a_1 \lambda_1 e_1 + a_2 \lambda_2 e_2 + \cdots + a_p \lambda_p e_p)$$

$$= \sum_{j,k=1}^{p} a_j a_k \lambda_k \underbrace{e_j^t e_k}_{\text{dot product } \begin{cases} 0 \text{ when } j \neq k \\ 1 \text{ when } j = k \end{cases}}$$

$$= \sum_{j=1}^{p} a_j^2 \lambda_j$$

<u>Restatement #2</u>: Find $(a_1, a_2, \cdots, a_p)$ which satisfy $\sum_{j=1}^{p} a_j^2 = 1$ and maximizes $\sum_{j=1}^{p} a_j^2 \lambda_j$

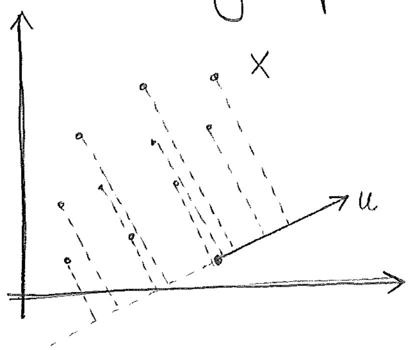
<u>Solution to Restatement #2</u>:

If we replace every eigenvalue in $\sum a_j^2 \lambda_j$ with $\lambda_1$, the value gets larger

$$\sum_{j=1}^{p} a_j^2 \lambda_j \leq \sum_{j=1}^{p} a_j^2 \lambda_1 = \lambda_1 \sum_{j=1}^{p} a_j^2 = \lambda_1$$

So $\lambda_1$ is the maximal value, which is achieved by $a = (1, 0, \cdots, 0)$.

Solution to original problem:




$u$ is the first eigenvector of the covariance matrix

$$\Omega = \frac{1}{n} X^t X \quad (X \text{ is centered})$$

The variance of the projected dataset is the first eigenvalue $\lambda_1$.