The standard orthonoral basis for $\mathbb{R}^n$ is

$\{e_1, \ldots, e_n\}$

where $e_i = 1$ at position i
$= 0$ at position $j \neq i$

$e_1 \ (1, 0, 0 \ldots)$
$e_2 \ (0, 1, 0 \ldots)$
$e_3 \ (0, 0, 1, \ldots)$
$\vdots$

A vector $\vec{v}$ is an eigen~~value~~ vector of an $n \times n$ matrix A
if $A\vec{v} = \lambda\vec{v}$ for some $\lambda \in \mathbb{R}$
λ is called the eigenvalue corresponding to $\vec{v}$.

The eigenvalues are solutions of the equation $\det(A - \lambda I) = 0$

where $I = \begin{pmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \end{pmatrix}$ is the identity matrix

5

The <u>transpose</u> of an $n \times n$ matrix A is the $n \times n$ matrix A^T defined by

$$(A^T)_{i,j} = (A)_{j,i}$$

<u>Ex</u>: $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$ $n=2$

A square matrix A is <u>orthogonal</u> if

$$A^T A = I$$

$A = (v_1 \cdots v_n)$ v_i column vectors

$$A^T \begin{pmatrix} v_1^T \\ v_2^T \\ v_3^T \\ \vdots \\ v_n^T \end{pmatrix}$$

$A^T A = I$ iff $v_i \circ v_i = 1$

$v_i \circ v_j = 0$ for $i \neq j$

because in matrix multiplication $v_i^T v_j = v_i \circ v_j$

The column vectors of an $n \times n$ orthogonal matrix are an orthonormal basis for $\mathbb{R}^n$.

END OF LINEAR ALGEBRA REVIEW

Suppose we have an $n \times r$ data matrix $\hat{X}$ where n is the number of observations

r is the number of variates

We define a new $n \times n$ matrix X by subtracting from each entry of $\hat{X}$ the sum of entries in its column divided by n:

$$(X)_{i,j} = (\hat{X})_{i,j} - \frac{1}{n}\sum_{k=1}^{n}(\hat{X})_{k,j}$$

X has columns which sum to 0:

$$\sum_{i=1}^{n} X_{i,j} = \sum_{i=1}^{n}(\hat{X})_{i,j} - \sum_{k=1}^{n}(\hat{X})_{k,j} = 0$$

Ex: $\hat{X} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ $X = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}$

$\frac{1+3}{2} = 2$ $\frac{2+4}{2} = 3$ centering values

Now, $\frac{1}{n-1} X^T X$ is the sample covariance matrix of the variates.

Note that $X^T X$ is an $r \times r$ matrix and $\left(\frac{1}{n-1} X^T X\right)_{i,j}$ is an estimator of the covariance between the ith and jth variate.

<u>FACT</u>: $X^T X$ can ~~be~~ always be diagonalized, meaning that there is an othonormal basis for $\mathbb{R}^r$ consisting of eigenvalues of $X^T X$.

This is because $X^T X$ is symmetric

$$\left((X^T X)^T = X^T (X^T)^T = X^T X \right)$$

$(AB)^T = B^T A^T$

and any ~~m~~ matrix that symmetric is diagonalizable.

We call the elements of the orthonormal basis of eigenvalues of $X^T X$ components.

It often happens that the largest few eigenvalues of $X^T X$ are smaller than the others. The components corresponding to these larger eigenvalues are principal.

These principal components explain most of the variability.

Now we will explain how to get an approximation to X using the principal components.

Let W be the $r \times r$ matrix whose columns are the vectors (i.e. components) in the orthonormal basis for $X^T X$ ordered left to right from largest to smallest eigenvalue.

We have $W^T W = I = W W^T$

so $$\begin{aligned} X &= X I \\ &= X W W^T \\ &= P W^T \end{aligned}$$

where $P = XW$

[Remark: $P^T P$ is a diagonal matrix whose diagonal entries are the eigenvalues of the components.]

If we zero-out the columns of P corresponding to non principal components we get a new matrix $\tilde{P}$.

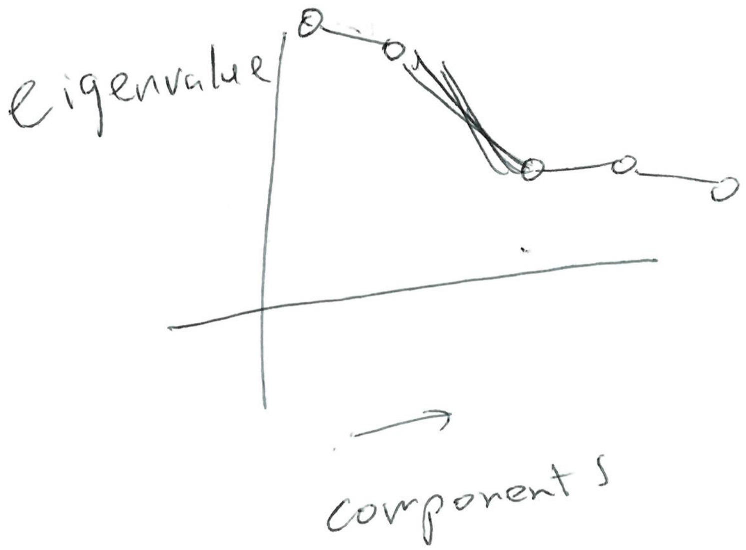
Then $\tilde{X} = \tilde{P} X^T$

is a good approximation to X which only uses the principal components.

Choosing which components are principal

We can choose the principal components the be the ones with largest eigenvalues containing at least 90% of the sum of the eigenvalue.

We can make a *scree diagram* which is a plot of the eigenvalues from largest to smallest.



looks like a mountain

Choose the principal components to make the mountain; the other components make the "scree".

Example

$n=2$

$r=2$

$\frac{4+2}{2}=3 \qquad \frac{10+6}{2}=8$

$$\hat{X}=\begin{pmatrix} 2 & 10 \\ 4 & 6 \end{pmatrix}$$

$$X=\begin{pmatrix} 2-3 & 10-8 \\ 4-3 & 6-8 \end{pmatrix}=\begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix}$$

$$X^TX=\begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}\begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix}=\begin{pmatrix} 2 & -4 \\ -4 & 8 \end{pmatrix}$$

symmetric

$$\det(X^TX-\lambda I)$$

$$=\det\begin{pmatrix} 2-\lambda & -4 \\ -4 & 8-\lambda \end{pmatrix}$$

$$=(2-\lambda)(8-\lambda)-16=0$$

$$\Rightarrow \lambda^2-10\lambda+16-16=0$$

$$\lambda^2-10\lambda=0$$

$$\lambda(\lambda-10)=0$$

$$\Rightarrow \lambda=0,\ \lambda=10$$

Let us find the components.

For $\lambda = 10$

$$\begin{pmatrix} 2-10 & -4 \\ -4 & \cancel{8}\,0-10 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$-8x - 4y = 0$$

$$-4x - 2y = 0$$

Let (arbitrarily) $x = 1$

then $y = -2$

$\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ is an eigen value of $X^T X$

~~wea~~ if v is ~~an~~ eigenvalue $\dfrac{v}{\sqrt{v \cdot v}}$ is an eigen value with unit length

because $\left(\dfrac{v}{\sqrt{v \cdot v}}\right) \cdot \left(\dfrac{v}{\sqrt{v \cdot v}}\right) = \dfrac{v \cdot v}{v \cdot v} = 1$

$$\frac{\begin{pmatrix} 1 \\ -2 \end{pmatrix}}{\sqrt{\begin{pmatrix} 1 \\ -2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -2 \end{pmatrix}}} = \frac{\begin{pmatrix} 1 \\ -2 \end{pmatrix}}{\sqrt{1+4}} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

For $\lambda = 0$

$$\begin{pmatrix} 2 & -4 \\ -4 & 8 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$2x - 4y = 0$$
$$-4x + 8y = 0$$

If $x = 2$, $y = 1$

$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ is an eigen value of $X^T X$

$\frac{1}{\sqrt{2^2 + 1^2}}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ is a component of $X^T X$.

$$W = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$$

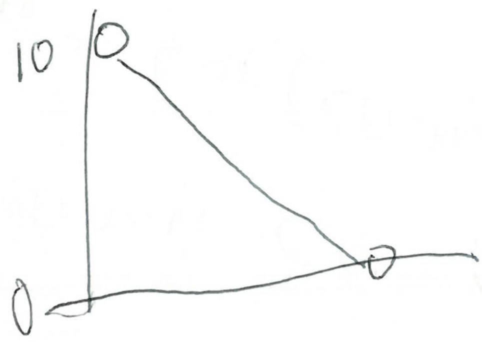
~~Larger~~ Component with larger eigenvalue goes the left.

$$P = XW = \begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix} \cdot \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$$

$$= \frac{1}{\sqrt{5}}\begin{pmatrix} -5 & 0 \\ 5 & 0 \end{pmatrix}$$

larger eigenvalue takes 100% of sum of eigenvalues >90%

and



$\frac{1}{\sqrt{5}}\begin{pmatrix}1\\-2\end{pmatrix}$ is the only principal component

The second column of P is already zeroed.

$$P^TP = \frac{1}{5}\begin{pmatrix}-5 & 5\\0 & 0\end{pmatrix}\begin{pmatrix}-5 & 0\\5 & 0\end{pmatrix}$$

$$= \frac{1}{5}\begin{pmatrix}50 & 0\\0 & 0\end{pmatrix} = \begin{pmatrix}10 & 0\\0 & 0\end{pmatrix}$$