random variable X, Y $\rightsquigarrow$ covariance matrix

$$K = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}$$ symmetric positive semi-def

$$\sum_{x,y} \operatorname{Prob}(X = x \text{ and } Y = y) \cdot \begin{bmatrix} x - \mu \\ y - \nu \end{bmatrix} \begin{bmatrix} x - \mu & y - \nu \end{bmatrix}$$

random variables

expected value of X

expected value of Y

random variables $X_1, \dots, X_n$ $\rightsquigarrow$ covariance matrix

$$K = \begin{bmatrix} \Sigma_{X_1X_1} & \cdots & \Sigma_{X_1X_n} \\ \vdots & \ddots & \vdots \\ \Sigma_{X_nX_1} & \cdots & \Sigma_{X_nX_n} \end{bmatrix}$$ symmetric positive semi-def.

$$\sum_{x_1, \dots, x_n} \operatorname{Prob}\begin{pmatrix} X_1 = x_1 \\ X_2 = x_2 \\ \vdots \\ X_n = x_n \end{pmatrix} \begin{bmatrix} x_1 - \mu_1 \\ \vdots \\ x_n - \mu_n \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 & \cdots & x_n - \mu_n \end{bmatrix}$$

$E[X_1]$ $E[X_n]$

<u>Def</u>. a random vector $\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ where $X_1, \dots, X_n$ are random variables

e.g. $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ takes value $\begin{bmatrix} 3 \\ 5 \\ -7 \end{bmatrix}$ with prob $\frac{2}{3}$ and value $\begin{bmatrix} -2 \\ 9 \\ 6 \end{bmatrix}$ with prob $\frac{1}{3}$

The expected value (average / mean) is defined as:

$$E[\mathbf{X}] = \sum_{\text{vectors } v} \text{Prob}(\mathbf{X} = v) \cdot v$$

$$E[\mathbf{X}] = \frac{2}{3} \begin{bmatrix} 3 \\ 5 \\ -7 \end{bmatrix} + \frac{1}{3} \begin{bmatrix} -2 \\ 9 \\ 6 \end{bmatrix} = \dots$$

The variance of $\mathbf{X}$ is defined as:

$$\underset{n \times n \text{ matrix}}{E}\Big[\underbrace{(\mathbf{X} - E[\mathbf{x}])}_{n \times 1}\underbrace{(\mathbf{X} - E[\mathbf{x}])^T}_{1 \times n}\Big] = \sum_{\substack{\text{vectors} \\ v}} \text{Prob}(\mathbf{X} = v) \cdot (v - E[\mathbf{x}])(v - E[\mathbf{x}])^T$$

continued on next slide

$$= \sum_{\substack{x_1, \dots, x_n \\ \text{(entries of } v)}} \text{Prob}\begin{pmatrix} X_1 = x_1 \\ \vdots \\ X_n = x_n \end{pmatrix} \cdot \begin{bmatrix} x_1 - E[X_1] \\ \vdots \\ x_n - E[X_n] \end{bmatrix} \begin{bmatrix} x_1 - E[X_1] \dots x_n - E[X_n] \end{bmatrix}$$

Upshot: the $\underbrace{\text{"variance" of } \mathbf{X}}_{E\left[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^T\right]}$ is the $n \times n$ covariance matrix of the n random variables which make up $\mathbf{X}$

$\forall$ vector of real numbers $c = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$

$\forall$ random vector $\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$

$\Longrightarrow X = c^T \cdot \mathbf{X} = c_1 X_1 + \dots + c_n X_n$

is a random variable

what is its variance?

the variance of $X = c^T \cdot \mathbf{X}$ is $\quad (\mu = c^T \cdot \boldsymbol{\mu})$

$$E\big[\underbrace{(X-\mu)}_{1\times 1}\underbrace{(X-\mu)^T}_{1\times 1}\big] = E\left[(c^T\mathbf{X} - c^T\boldsymbol{\mu})(\mathbf{X}^T\cdot c - \boldsymbol{\mu}^T\cdot c)\right] = c^T \cdot \underbrace{E\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\right]}_{\text{"the variance of } \mathbf{X}\text{"}} \cdot c$$

a.k.a the covariance matrix of the entries of $\mathbf{X}$

$c^T \cdot K \cdot c$

the energy of the vector c and symmetric matrix K

Ex: $\quad \mathbf{X} = \begin{bmatrix} \text{temp} \\ \text{pressure} \end{bmatrix} \qquad c = \begin{bmatrix} 5 \\ -7 \end{bmatrix} \qquad \begin{bmatrix} K_{TT} & K_{TP} \\ K_{PT} & K_{PP} \end{bmatrix}$

$X = c^T \cdot \mathbf{X} = 5 \cdot \text{temp} - 7 \cdot \text{pressure} \rightsquigarrow$ the variance of X $= [5 \quad -7] \cdot K \cdot \begin{bmatrix} 5 \\ -7 \end{bmatrix}$

X has variance 0 precisely if $c^T K c = 0$, which can only happen if K is singular, i.e. K is not pos. definite

Example $\mathbf{X} = \begin{bmatrix} X \\ Y \end{bmatrix} \rightsquigarrow K = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}$

$Z = X + Y$, what is the variance of Z?

$$= \underbrace{\begin{bmatrix} 1 & 1 \end{bmatrix}}_{c^T} \underbrace{\begin{bmatrix} X \\ Y \end{bmatrix}}_{\mathbf{X}}$$

$$= c^T \cdot K \cdot c = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\Sigma_{X+Y, X+Y} = \Sigma_{XX} + \Sigma_{YY} + 2\Sigma_{XY}$$

cross-term

(e.g. if $X = Y$

$$\Sigma_{2X,2X} = \Sigma_{XX} + \Sigma_{XX} + 2\Sigma_{XX}$$

$$\Sigma_{2X,2X} = 4 \cdot \Sigma_{XX}$$

Def: Principal Component Analysis (PCA)

is just a fancy way of saying "diagonalize the $\underbrace{\text{covariance matrix } K}_{\text{of random variables } X_1,\dots,X_n}$"

$$K = \underbrace{Q}_{\text{orthogonal matrix}} \overbrace{D}^{\text{diagonal}} Q^T$$

$Q^T K Q = D$

$$= \left[q_1 | \cdots | q_n \right] \begin{bmatrix} d_1 & & 0 \\ & d_2 & \\ 0 & & \ddots & \\ & & & d_n \end{bmatrix} \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix}$$

$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \rightsquigarrow \mathbf{Y} = Q^T \cdot \mathbf{X}$

(entries $Y_1,\dots,Y_n$ of $\mathbf{Y}$ will just be linear combinations of the entries $X_1,\dots,X_n$ of $\mathbf{X}$)

$\nu = E[\mathbf{Y}] = Q^T \cdot E[\mathbf{X}] = Q^T \cdot \mu$

the covariance matrix of $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = Q^T \mathbf{X}$ is $E\left[(\mathbf{Y}-\nu)(\mathbf{Y}-\nu)^T\right] = E\left[Q^T(\mathbf{X}-\mu)(\mathbf{X}-\mu)^T \cdot Q\right] = Q^T K Q = D = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{bmatrix}$

Hence the random variables $Y_1,\dots,Y_n$ are uncorrelated!