Statistics is about data sets

a collection of samples $x_1, \dots, x_n$

real numbers, quantifying some measured quantity: temperature, height ....

$\boxed{\text{Definition}}$

- the mean of the data set is:

$$\mu = \frac{1}{n}(x_1 + \dots + x_n)$$

- the variance of the data set is

$$\Sigma = \frac{(x_1-\mu)^2 + \dots + (x_n-\mu)^2}{n-1}$$

Bessel's correction: use $n-1$ instead of n (explanation in lecture notes)

As $n \to \infty$, μ and Σ converge to the mean and variance of the probability distribution from which the samples are extracted

$$\mathbb{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \implies \mu = \frac{\mathbb{1}^T x}{\mathbb{1}^T \mathbb{1}} = \frac{x_1 + \dots + x_n}{n}$$

$n \times 1$ vectors

ratio appearing in the projection formula of x onto $\mathbb{1}$, i.e. $\mathbb{1} \cdot \frac{\mathbb{1}^T x}{\mathbb{1}^T \mathbb{1}}$

$$\Sigma = \frac{(x_1 - \mu)^2 + \dots + (x_n - \mu)^2}{n-1} = \frac{\| x - \mu \cdot \mathbb{1} \|^2}{n-1} = \frac{\left\| x - \mathbb{1} \frac{\mathbb{1}^T x}{\mathbb{1}^T \mathbb{1}} \right\|^2}{n-1} = \frac{\| P x \|^2}{n-1}$$

$$\Downarrow$$

$$\mathbb{1} \cdot \frac{\mathbb{1}^T x}{\mathbb{1}^T \mathbb{1}} = \mathrm{proj}_{\mathbb{1}} x$$

$$\Downarrow$$

$$x - \mathbb{1} \frac{\mathbb{1}^T x}{\mathbb{1}^T \mathbb{1}} = \mathrm{proj}_V x$$

where V is the orthogonal complement to $\mathbb{1}$

Projection matrix onto V

$$= I - \frac{\mathbb{1} \mathbb{1}^T}{\mathbb{1}^T \mathbb{1}} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} - \frac{1}{n} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{n-1}{n} & \cdots & \frac{-1}{n} \\ \vdots & \ddots & \vdots \\ \frac{-1}{n} & \cdots & \frac{n-1}{n} \end{bmatrix} =: P$$

Say you have two data sets $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

Def: the covariance of these data sets is:

$$\Sigma_{xy} = \frac{(x_1-\mu)(y_1-\nu) + \cdots + (x_n-\mu)(y_n-\nu)}{n-1}$$

μ → average of $x_1, \ldots, x_n$

ν → average of $y_1, \ldots, y_n$

(obs: Σ_{xx} is just the variance of the data set $\mathbf{x}$)

$$P\mathbf{x} = \begin{bmatrix} x_1-\mu \\ \vdots \\ x_n-\mu \end{bmatrix}, \quad P\mathbf{y} = \begin{bmatrix} y_1-\nu \\ \vdots \\ y_n-\nu \end{bmatrix}$$

Thm: $\boxed{\Sigma_{xy} = \frac{\mathbf{x}^T P \mathbf{y}}{n-1}}$

Pf: $\frac{\mathbf{x}^T P \mathbf{y}}{n-1} = \frac{\mathbf{x}^T P^T P \mathbf{y}}{n-1} = \frac{(P\mathbf{x})^T (P\mathbf{y})}{n-1} = \frac{(x_1-\mu)(y_1-\nu) + \cdots + (x_n-\mu)(y_n-\nu)}{n-1} = \Sigma_{xy}$

$$\Sigma_{xx} = \frac{\|P\mathbf{x}\|^2}{n-1} = \boxed{= \frac{\mathbf{x}^T P^T P \mathbf{x}}{n-1}}$$

b/c P is a projection matrix,

$$P^T = P \text{ and } P^2 = P$$

Cauchy - Schwatz : $\left|\Sigma_{xy}\right| \leq \sqrt{\Sigma_{xx}\Sigma_{yy}}$

Take m data sets $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$, ...

Def. the covariance matrix of $x, y, z, \ldots$ is :

$$K = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{xz} & \cdots \\ \Sigma_{yx} & \Sigma_{yy} & \Sigma_{yz} & \cdots \\ \Sigma_{zx} & \Sigma_{zy} & \Sigma_{zz} & \cdots \\ \vdots & & & \ddots \end{bmatrix}$$

m×m symmetric matrix

$$\underbrace{A}_{n \times m \text{ matrix}} = \begin{bmatrix} x_1 & y_1 & z_1 & \\ \vdots & \vdots & \vdots & \cdots \\ x_n & y_n & z_n & \end{bmatrix} \Rightarrow PA = \begin{bmatrix} x_1-\mu & y_1-\nu & z_1-\xi & \\ \vdots & \vdots & \vdots & \cdots \\ x_n-\mu & y_n-\nu & z_n-\xi & \end{bmatrix}$$

μ: mean of x $\quad$ ν: mean of y $\quad$ ξ: mean of z

Thm: $\boxed{K = \dfrac{A^T P A}{n-1}}$

$$\frac{\underbrace{(PA)^T}_{m\times n} \underbrace{PA}_{n\times m}}{n-1} = \frac{\begin{bmatrix} x_1-\mu & \cdots & x_n-\mu \\ y_1-\nu & \cdots & y_n-\nu \\ & \vdots & \end{bmatrix}\begin{bmatrix} x_1-\mu & y_1-\nu & \\ \vdots & \vdots & \cdots \\ x_n-\mu & y_n-\nu & \end{bmatrix}}{n-1} = \overbrace{\frac{\begin{bmatrix} (x_1-\mu)^2+\dots+(x_n-\mu)^2 & (x_1-\mu)(y_1-\nu)+\dots \\ (x_1-\mu)(y_1-\nu)+\dots & (y_1-\nu)^2+\dots+(y_n-\nu)^2 \\ & \vdots \end{bmatrix}}{n-1}}^{K}$$

PCA : diagonalize covariance matrix

$$\frac{A^T P A}{n-1} = K = Q D Q^T, \quad Q^{-1} = Q^T$$

sample matrix

diagonal matrix

$$\Downarrow$$

$$\frac{Q^T A^T P A Q}{n-1} = D$$

$$\Downarrow$$

$$\frac{(AQ)^T P A Q}{n-1} = D$$

$$\Downarrow$$

$$\frac{B^T P B}{n-1} = D$$

Let $B = A \cdot Q = \begin{bmatrix} x_1 & y_1 & \\ \vdots & \vdots & \cdots \\ x_n & y_n & \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & \cdots \\ q_{21} & q_{22} & \\ \vdots & & \ddots \end{bmatrix}$

$$= \begin{bmatrix} x_1 q_{11} + y_1 q_{21} + \cdots & x_1 q_{12} + y_1 q_{22} + \cdots \\ x_2 q_{11} + y_2 q_{21} + \cdots & \vdots \\ \vdots & \vdots \\ x_n q_{11} + y_n q_{21} + \cdots & x_n q_{12} + y_n q_{22} + \cdots \end{bmatrix}$$

The columns of B are mutually uncorrelated

$\Uparrow$

the covariance matrix of B is diagonal

$$x = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 7 \\ 8 \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} 4 \\ 3 \\ 5 \\ 6 \\ 9 \end{bmatrix}$$

PCA: find linear combinations of x and y which are uncorrelated

$$A = [x \mid y] = \begin{bmatrix} 2 & 4 \\ 3 & 3 \\ 4 & 5 \\ 7 & 6 \\ 8 & 9 \end{bmatrix}$$

$$PA = \frac{1}{5} \begin{bmatrix} -14 & -7 \\ -9 & -12 \\ -4 & -2 \\ 11 & 3 \\ 16 & 18 \end{bmatrix}$$

$$K = \frac{A^T PA}{5-1} = \frac{(PA)^T PA}{5-1} = \frac{1}{20} \begin{bmatrix} 134 & 107 \\ 107 & 106 \end{bmatrix}$$

option 1: $\boxed{K = QDQ^T}$ compute 2×2 matrix $Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$

$$D = \begin{bmatrix} 6 + \frac{\sqrt{11645}}{20} & 0 \\ 0 & 6 - \frac{\sqrt{11645}}{20} \end{bmatrix}$$

define $B = AQ = \begin{bmatrix} \underbrace{\begin{matrix} 2q_{11} + 4q_{21} \\ 3q_{11} + 3q_{21} \\ \vdots \end{matrix}} & \underbrace{\begin{matrix} 2q_{12} + 4q_{22} \\ 3q_{12} + 3q_{22} \\ \vdots \end{matrix}} \end{bmatrix}$

mutually uncorrelated data sets

option 2: $PA = \boxed{U \Sigma V^T}$, where $\Sigma = \begin{bmatrix} \sqrt{24 + \frac{\sqrt{11645}}{5}} & 0 \\ 0 & \sqrt{24 - \frac{\sqrt{11645}}{5}} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

$$K = \frac{(U\Sigma V^T)^T (U \Sigma V^T)}{5-1} = \frac{V \Sigma^T U^T U \Sigma V^T}{5-1}$$

$$= \frac{V \Sigma^T \Sigma V^T}{5-1}$$

$Q = V$, $D = \dfrac{\Sigma^T \Sigma}{5-1}$