pdf is defined over n dimensional euclidean space. ($P(x)$)

$\underline{x} \in \mathbb{R}^n$

↑

column vector

$p(\underline{x}) \geq 0 \quad \forall \underline{x} \in \mathbb{R}^n$

$$\int_{\mathbb{R}^n} p(x)\,dx = 1$$

$\hookrightarrow \approx (dx_1 dx_2 dx_3 \ldots dx_n)$

~~eg~~. One example of pdf is gaussian or Normal distribution

$$p(x) = \frac{1}{(\sqrt{2\pi})^n |\Sigma|^{1/2}} \exp\Big\{ \underbrace{-\tfrac{1}{2} \underset{1\times n}{(x-\mu)'} \underset{n\times n}{\Sigma^{-1}} \underset{n\times 1}{(x-\mu)}}_{\text{scalar } (1\times 1)} \Big\}$$

if $|\Sigma|$ can be negative or zero, then $\sqrt{|\Sigma|}$ will be complex or not defined. Thus $|\Sigma|$ ~~has~~ has to be strictly positive i.e. $\boxed{|\Sigma| > 0.}$

$\underline{\mu}$ - Mean vector

$\underset{n\times n}{\Sigma}$ - Variance Covariance Matrix / Dispersion Matrix

exponential of a scalar will be greater than zero. Thus, we need to ensure $|\Sigma|$ should be strictly positive

$x_1, x_2, \ldots x_n \in \mathbb{R}$

Mean of n observation $\quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Variance of n observation $\quad = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Covariance is between two variables

There should be some quantity to find relationship between X & Y such that shown in plots.

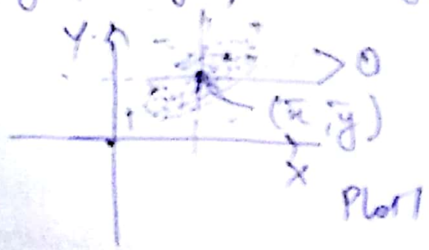
If we multiply new x & y values found using new ones

Covariance $\quad = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) > 0$ for case 1

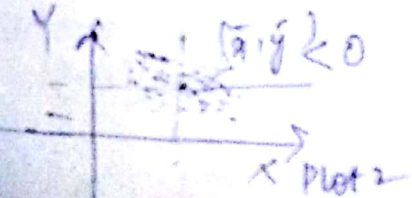
$Cov(X,Y)$

(X, Y) ← height, weight

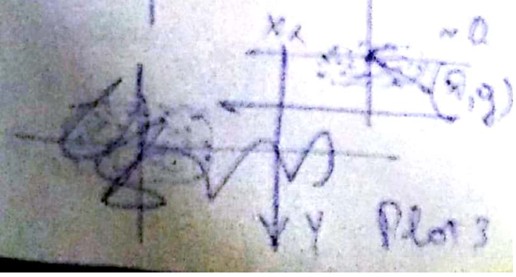
$(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$



Plot 1



Plot 2



Plot 3

Clustering vs Classification

problem of classification to classify these two triangles to distinguish between these two.

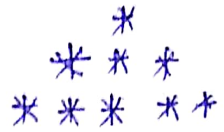


fig 1

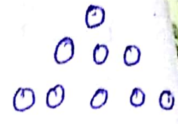


fig 2

We need property to classify or distinguish these two triangles. Let it be element i.e. star & circle of fig 1 fig 2.

If I can extract these two element then I can classify these two objects.

If we choose element then we cannot classify correctly given fig 3.

fig 3

Thus we need more than one feature to distinguish the diff. objects.

Challeges in problem of classification

→ To distinguise objects which are different but actually look same. for example human faces, cars.

~~Two clas~~ We have n number of features. (2)

$Cov(X,X)$ is variance of X.

Variance Covariance Matrix.

$$\begin{bmatrix} Cov(X_1,X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & Cov(X_2,X_2) & \cdots & Cov(X_2,X_n) \\ \vdots & & & \\ Cov(X_n,X_1) & Cov(X_n,X_2) & \cdots & Cov(X_n,X_n) \end{bmatrix}$$

symmetric matrix →

as $Cov(A,B) = Cov(B,A)$

If n feature are $X_1 X_2 X_3 \ldots X_n$ then the Variance Covariance matrix of these n features is denoted by Σ & is defined as above

Variance Covariance Matrix is positive definite.

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$$

distance b/w x & y $\quad d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$

euclidean distance

for a person lets say $\begin{pmatrix} 160 \text{ cm} \\ 70 \text{ kg} \end{pmatrix}$ another person $\begin{pmatrix} 168 \text{ cm} \\ 74 \text{ kg} \end{pmatrix}$

distance between two

Thus $d(x,y) = \sqrt{8^2 + 4^2}$
$= \sqrt{64+16} = \sqrt{80}$

Let us assume we ~~use~~ another metric for measurement

now $x = \begin{pmatrix} 1600 \text{ mm} \\ 70 \text{ kg} \end{pmatrix} \quad y = \begin{pmatrix} 1680 \text{ mm} \\ 74 \text{ kg} \end{pmatrix}$

Now $d(x,y) = \sqrt{6400+16} = \sqrt{6416}$

But distance should not change

$$d^2(x,y) = (x_i - y_i)^2$$

$$= (x_1 - y_1 \quad x_2 - y_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \end{pmatrix}$$

for generalization

$$d^2(x,y) = (x_1 - y_1 \quad x_2 - y_2) \begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix} \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \end{pmatrix}$$

w_i will change if unit change

Better Generalization

$$d^2(x,y) = (x_1 - y_1 \quad x_2 - y_2) \begin{pmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{pmatrix} \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \end{pmatrix}$$

$d^2(x,y) \geq 0$, for this $w_{ij} > 0$ & $x_i \neq y_i$

for eg.

$$\begin{pmatrix} x_1 - y_1 & x_2 + y_2 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \end{pmatrix} > 0$$

iff $x_1 \neq y_1$
or
$x_2 \neq y_2$

$A_{n \times n}$ is said to be positive definite if

$\underline{a}' A \underline{a} > 0 \quad \forall \; \underline{a} \neq \vec{0}$

Variance covariance Matrix can be shown to be non-negative definite but most of time it is positive definite

$A_{n \times n}$ is said to positive semidefinite or Non-Negative definite if

$\underline{a}' A a \geq 0 \quad \forall \; \underline{a}$

Gaussian Random Variables are extremely useful in machine learning and statistics for two reasons

→ Common for modelling noise in statistical algos.

→ Gaussian RV are convineent for many manipulations.

Scanned by CamScanner