Lec 3

Wednesday, 30 July 2025 10:08 AM

Population $\underset{p\times 1}{\mu}$ $\underset{p\times p}{\Sigma}$.

$$s_{ij} = \frac{1}{n} \sum \underset{p\times 1}{(x_i - \bar{x}_i)} \underset{p\times 1}{(x_j - \bar{x}_j)^T}$$

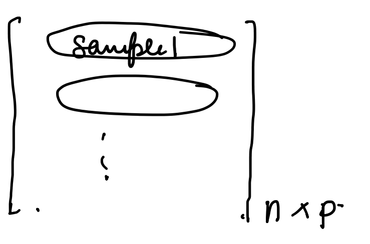
Sample $\underset{p\times 1}{\bar{x}}$ $\underset{p\times p}{S}$

$$s_{ij} = \frac{1}{n} \sum_{r=1}^{n} (x_{ri} - \bar{x}_i)(x_{rj} - \bar{x}_j).$$

$\underset{\sim}{x_j}$ $\begin{matrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{matrix}$. $\underset{\sim}{\bar{x}} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{pmatrix}$, $\bar{x}_j = \dfrac{\sum_{r=1}^{n} \underset{\sim}{x_{rj}}}{n}$.

unless sample data given the formulas are r.v.
practically they're deterministic.

Collect sample & sample data matrix.



In $\mathbb{R}^p$, we have the sample points.
So in finite dims, all distances are equiv. (all norms are equiv. in f. dim).

2 sample pts x, y then
(euclidean distance) –
$d(x,y) = (x-y)^T(x-y) = \sqrt{(x_1-y_1)^2 + \cdots + (x_p-y_p)^2}$

$$\begin{bmatrix} \underset{\text{(meters)}}{x_1} & \underset{\text{(hrs)}}{x_2} & \underset{\text{(grams)}}{x_3} \\ & & \\ & & \end{bmatrix}$$

Flaw 1
diff. units of features; we need to make data collection unitless.

Suppose $p=3$

	4.1	48	.1
x	1.9	99	2.9
y	6.0	5.1	3.0.

$d(x,y) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + (x_3-y_3)^2}$.
$= 48.17.$

Flaw 2.
x feature very dominant; hides the properties of other features.

Flaw 3 doesn't capture the dependency of one feature on other.

→ solved by providing weights to the components ($(x_i-y_i)^2$).

$$d(x,y) = \sqrt{\left(\frac{x_1-y_1}{\underset{\uparrow\ \sqrt{\text{variance}}}{\sqrt{s_{11}}}}\right)^2 + \cdots + \left(\frac{x_p-y_p}{\sqrt{s_{pp}}}\right)^2}$$

(similar to $X \sim N(\mu, \sigma^2)$
$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
$\frac{x-y-0}{\sqrt{s_{11}}}$)

This makes the components unitless.

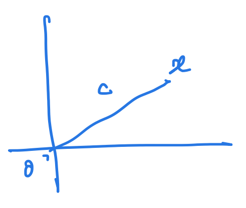
Now, each component of vector x, y is given a weight in distance calculation b/w x & y. & wt is $\frac{1}{s_{ii}}$ $1 \le i \le p$.

Dependency structure b/w feature in d-formula-

$$d(x,0) = \sqrt{\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}} + \cdots \frac{x_p^2}{s_{pp}}} = c \ge 0$$

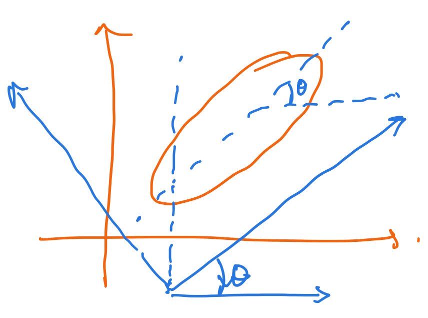
$x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$.

$$\frac{x_1^2}{c^2 s_{11}} + \frac{x_2^2}{c^2 s_{22}} + \cdots + \frac{x_p^2}{c^2 s_{pp}} = 1.$$

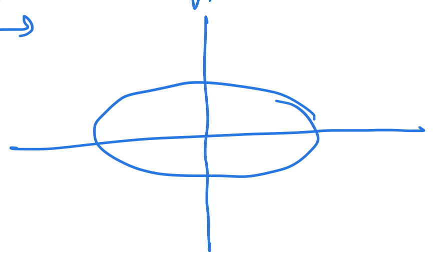


standard ellipse.

p- axis in ellipse (tilted ellipse)
semilength $= c\sqrt{s_{ii}}$.
higher variance ⇒ elongated on that sides.



→ rotate & shift to origin



Let $p=2$. (tilted ellipse bcz of dependency b/w two features).
$x = (x_1, x_2)$.
Rotation matrix.

$$\underset{\text{standard frame}}{\begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix}} = \overset{P}{\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}} \underset{\text{tilted frame}}{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} = \begin{pmatrix} \cos\theta\, x_1 + \sin\theta\, x_2 \\ -\sin\theta\, x_1 + \cos\theta\, x_2 \end{pmatrix}$$

↑
orthonormal matrix.
$P^T P = I$

$$x_1^* - \bar{x}_1^* = \cos\theta(x_1-\bar{x}_1) + \sin\theta(x_2-\bar{x}_2)$$
$$x_2^* - \bar{x}_2^* = -\sin\theta(x_1-\bar{x}_1) + \cos\theta(x_2-\bar{x}_2).$$

Shifting range to mean.

$\mathrm{Var}(x_1^*) = \cos^2\theta\, \mathrm{Var}(x_1) + \sin^2\theta\, \mathrm{Var}(x_2) + 2\sin\theta\cdot\cos\theta \cdot \mathrm{cov}(x_1, x_2).$
$\mathrm{Var}(x_2^*) =$ ______ similarly.

$$d(\overset{x^*}{\bar{p}}, p) = \sqrt{\left(\frac{x_1^* - \bar{x}_1^*}{\sqrt{s_{11}^*}}\right)^2 + \left(\frac{x_2^* - \bar{x}_2^*}{\sqrt{s_{22}^*}}\right)^2}.$$

$$= \sqrt{\underbrace{a_{11}}_{\text{constants in terms of } \sin\theta\ \&\ \cos\theta}(x_1-\bar{x}_1)^2 + 2a_{12}(x_1-\bar{x}_1)(x_2-\bar{x}_2) + a_{22}(x_2-\bar{x}_2)^2} = \langle x-\bar{x}, A(x-\bar{x})\rangle.$$

↑
rota^n.

$(x-\bar{x})^T A (x-\bar{x})$ $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$.

we need the dist to be +ve, for which we need to have. A as. symmetric +ve semi definite, to define the inner product.

The role of A matrix is played by S^{-1}.
∵ S is symmetric ⇒
S is always +ve semi definite?
iff.
all minors ≥ 0.

$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$.

$s_{11} > 0.$
$s_{11}\cdot s_{22} - s_{12}\cdot s_{21}$
$s_{11}\cdot s_{22} - (s_{12})^2 \ge 0.$
$\sqrt{s_{11} s_{22}} > s_{12}.$

$$\frac{s_{12}}{\sqrt{s_{12} s_{22}}} < 1$$

S positive semi def.
so one eigen val can be zero.
so inverse not psble
but for all practical purpose we take S to be positive definite.

$d(x,0) = x^T S^{-1} x$ → Mahalanobis dist.