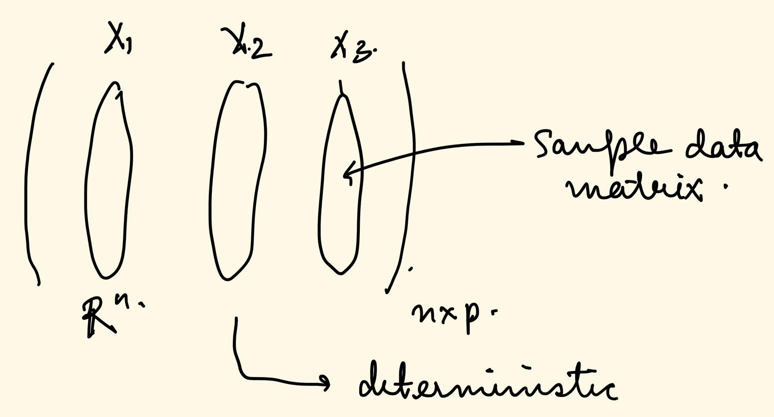
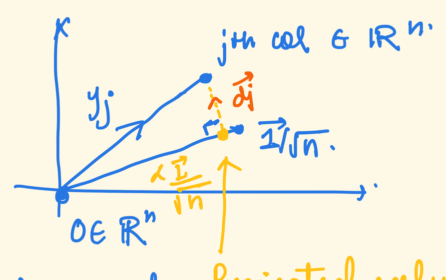
# Lec 6

Wednesday, 6 August 2025    10:07 AM




Columns are in $\mathbb{R}^n$ & p-such vectors.




$$y_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix} X_j \qquad j = 1, 2 \ldots, p.$$

orthogonal. Projected value of $\vec{y_j}$ on unit vector. $\vec{1}/\sqrt{n}$.

Consider a vector all component $= 1$ $\quad \vec{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$

A unit vector in the direction of $\vec{1}$ is $\frac{\vec{1}}{\sqrt{n}}$
(euclidian norm).

$\alpha \frac{\vec{1}}{\sqrt{n}}$ $\quad \alpha$ is the length of the vector.

Projec$^n$ of $\vec{y_j}$ on $\frac{\vec{1}}{\sqrt{n}} = \left(\vec{y_j} \cdot \frac{\vec{1}}{\sqrt{n}}\right) \frac{\vec{1}}{\sqrt{n}} = \frac{1}{n}(\vec{y_j} \cdot \vec{1})\, \vec{1}. = \frac{1}{n}(x_{1j} + x_{2j} + \cdots + x_{nj})\, \vec{1}$
$= \overline{x_j}\, \vec{1}$

$\overline{x_j}$: constant & a sample mean of the $j^{th}$ feature.

$$\overline{x_j}\,\vec{1} = \begin{pmatrix} \overline{x_j} \\ \overline{x_j} \\ \vdots \\ \overline{x_j} \end{pmatrix} \in \mathbb{R}^n.$$

$\vec{d_j}$, the $\perp^{ar}$ vector.

$$\boxed{\vec{y_j} = \overline{x_j}\,\vec{1} + \vec{d_j}}$$

$\vec{d_j} = \vec{y_j} - \overline{x_j}\,\vec{1}$

$$= \begin{pmatrix} x_{1j} - \overline{x_j} \\ x_{2j} - \overline{x_j} \\ \vdots \quad - \overline{x_j} \\ x_{nj} - \overline{x_j} \end{pmatrix} \in \mathbb{R}^n$$

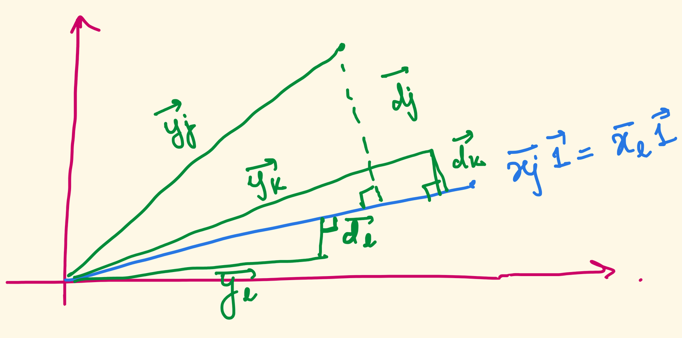
: deviation vector.
how much are we deviating from the mean vector.

$\vec{d_j} \cdot \vec{1} = \sum_{k=1}^{n} (x_{kj} - \overline{x_j}) = \sum_{k=1}^{n} x_{kj} - n\overline{x_j}$
$= n\overline{x_j} - n\overline{x_j} = 0.$

Expected $\because$ $(\overline{x_j}\,\vec{1}) \cdot \vec{d_j} = \overline{x_j}(\vec{1} \cdot \vec{d_j}) = 0.$
and they were orthogonal.

$\|d_j\|^2 = (x_{1j} - \overline{x_j})^2 + (x_{2j} - \overline{x_j})^2 + \cdots + (x_{nj} - \overline{x_j})^2.$
Euclidian norm $= \sum_{k=1}^{n} (x_{kj} - \overline{x_j})^2$

$= n s_{jj}$ $\quad$ (when n very large, n or n-1 doesn't matter)
↪ jth sample variance.

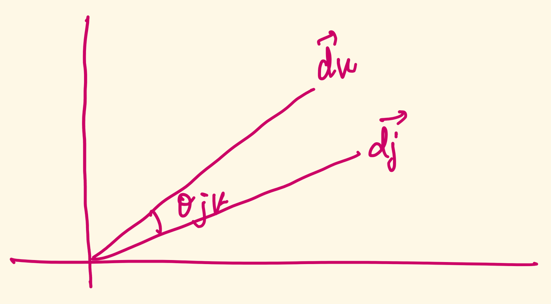



$\overline{x_j}\,\vec{1} = \overline{x_l}\,\vec{1} = \overline{x_k}\,\vec{1}$ (for this example, they may vary).

If n same, if $\vec{d_x}$ small $\Rightarrow$ std dev small, given n is same.

For heavy tailed dist, std deviation not a very good metric as it captures the central tendency.

Think of two deviation vectors $\vec{d_j}$ & $\vec{d_k}$ in $\mathbb{R}^n$




$$\cos(\theta_{jk}) = \frac{\vec{d_k} \cdot \vec{d_j}}{\|\vec{d_k}\|\, \|\vec{d_j}\|}.$$

$$= \frac{\vec{d_k} \cdot \vec{d_j}}{\sqrt{n s_{kk}}\, \sqrt{n s_{jj}}} = \frac{\vec{d_k} \cdot \vec{d_j}}{n\sqrt{s_{kk} \cdot s_{jj}}}.$$

$$= \frac{\frac{1}{n} \sum_{r=1}^{n} (x_{rj} - \overline{x_j})(x_{rk} - \overline{x_k})}{\sqrt{s_{kk}}\, \sqrt{s_{jj}}}$$ → covariance of ~~jth~~ $x_j$ & $x_k$ in sample.

$$= \frac{s_{kj}}{\sqrt{s_{kk}}\, \sqrt{s_{jj}}} = \frac{\text{cov}(x_k, x_j) \text{ in sample matrix}}{\sqrt{s_{kk}}\, \sqrt{s_{jj}}}$$

$$= \frac{\text{Cov}(X_j, X_k)}{SD(x_k) \cdot SD(x_j)} = \rho_{jk}$$ (correlation coefficient b/w jth & kth feature based on the sample data)

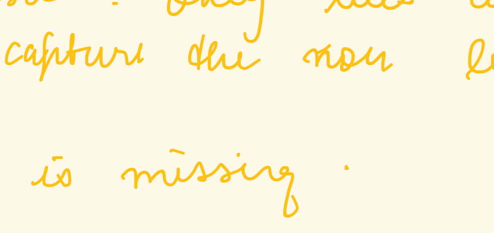
$\rho_{jk}$ might change if sample changed

if $\theta = 90^\circ$ $\cos(\theta) = 0 \Rightarrow \rho_{jk} = 0 \Rightarrow$ jth & kth features are not correlated

$\theta = 0^\circ$ $\cos(\theta) = 1 \Rightarrow \rho_{jk} = 1. \Rightarrow$ perfectly co-related.
$\theta = \pi$ $\cos(\theta) = -1 \Rightarrow \rho_{jk} = -1 \Rightarrow$ jth & kth feature ~~are~~ have perfect -ve corelated

If there's linear dependency b/w $x_j$ & $x_k$ feature, we can fit a linear line fitting both the patterns


non-linear




← can't be captured by line.

Co-relation not a good metric as it does not guarantee indepence b/w the features or vice versa. Only tells about the linear relationship. Doesn't capture the non linear rel$^n$ ship.

Uncorelated or linearly indep. is missing.

why do we use n or n-1 while calc std dev:-

$\Rightarrow$ p=1 (in one r.v. X), we know that.

$\boxed{E(\bar{X}) = \mu.}$ pop mean.
sample mean

$\boxed{E(s^2) = \sigma^2}$ pop variance
sample variance.

$$s^2 = \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \bar{x})^2$$

Any statistic is called an unbiased estimator of a population parameter say $\theta$ if $E(T) = \theta$.
if sample change → T changes but $E(T) = \theta$ should hold irrespectively.

$\bar{X}$ : sample mean.
$s^2$ : sample variance
are unbiased estimators of pop mean & pop. variance resp.