CORRELATION TESTS
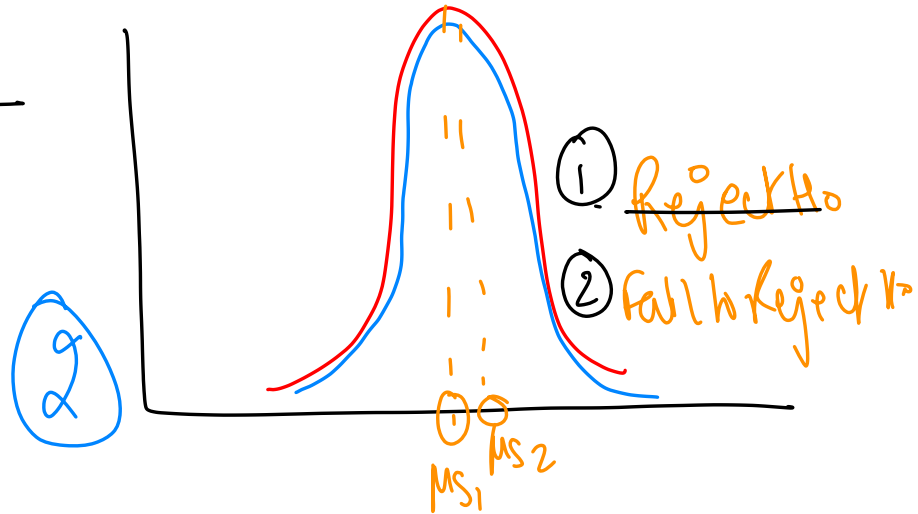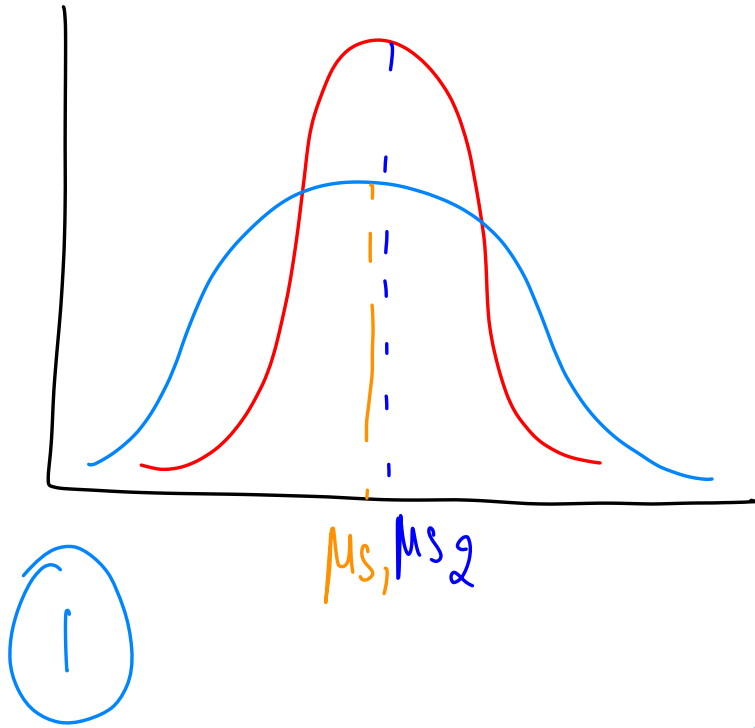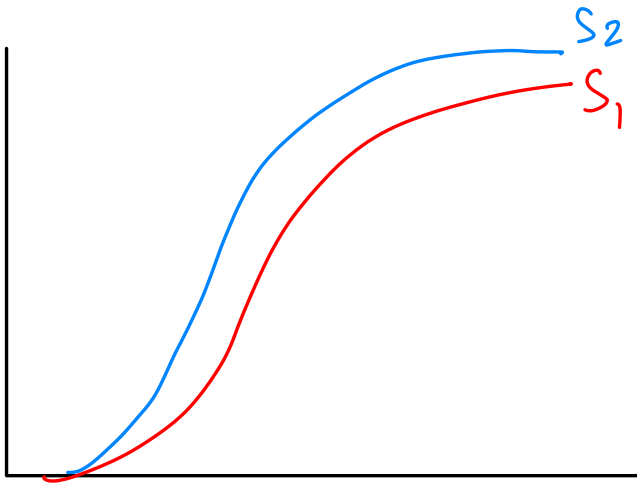
T test → Fail to Reject Ho

$$H_0 : \mu_1 = \mu_2$$

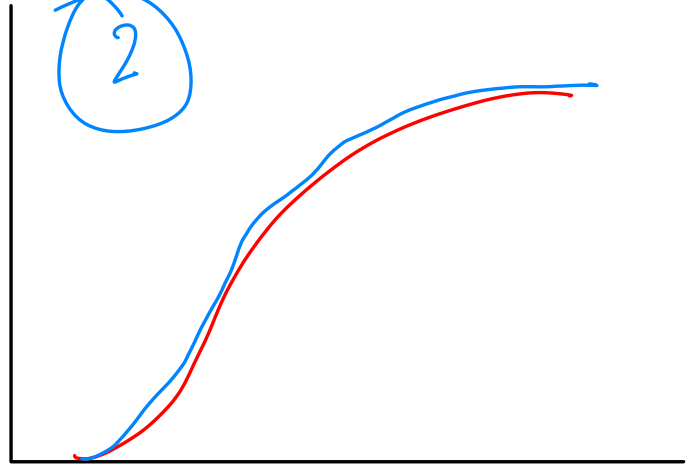$$H_a : \mu_1 \neq \mu_2$$

① Reject Ho

② Fail to Reject Ho

Graph ①: MS_1, MS_2

Graph ②: $MS_1 \quad MS_2$

① (graph with curves labeled S_2 (blue) and S_1 (red))
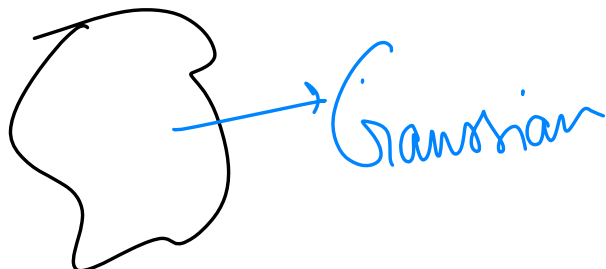
② (graph with two overlapping curves)

KS Test → Kolmogorov Smirnoff Test.

S1 $\longleftrightarrow$ Gaussian $\rightarrow$ KS Test

KS Test $\rightarrow$ S1 $\quad$ $\rightarrow$ Gaussian
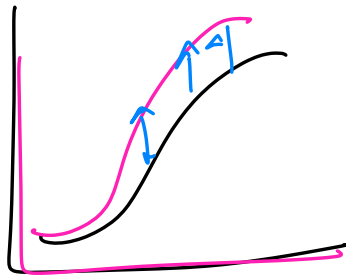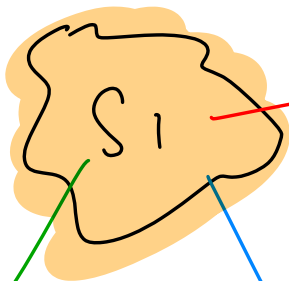
CDFs

S_1 — Gaussian ✓ ✗

Gaussian

2?

Sample

New class

percentiles

Gaussian

- −2
- −1.9
- −1.9
- −1.7
- −1.6
- ⋯
- 0
- ⋯
- 2

−2 −1 0 1 2

1st % 1%
2nd% 2nd %
100% 3rd %
100%
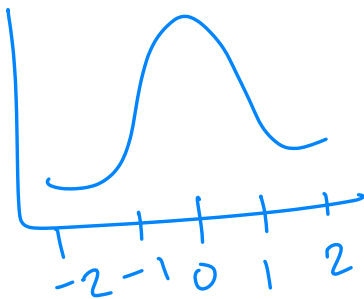
Sample
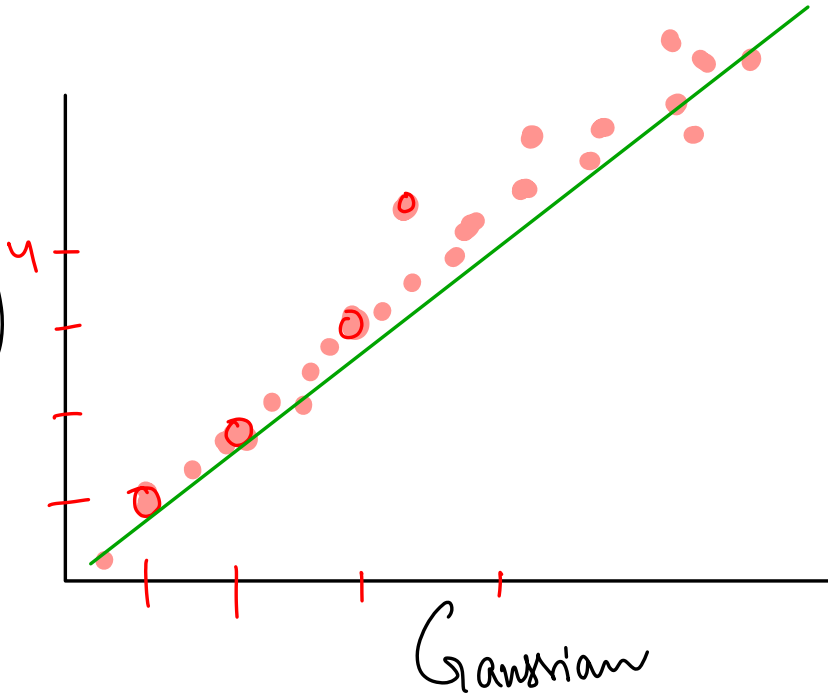
Gaussian

If these points
are lying on a straight
line → Distribution
is Gaussian.

QQ Plot → Quantile - Quantile Plot.

S_1

x_1 100% x_7
x_6

x_2 x_5

x_4

x_3 x_3

x_2

x_4 x_1

x_5

x_6

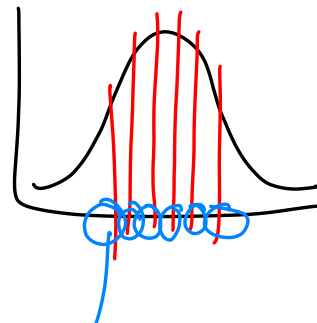x_7 $\rightarrow 68, 95, 99\%$

y_1 y_2 y_3 y_4 y_5 y_6 y_7

$y_1 \, y_2 \, y_3 y_4 \, y_5 y_6 \, y$

y

y

Gau

y percur
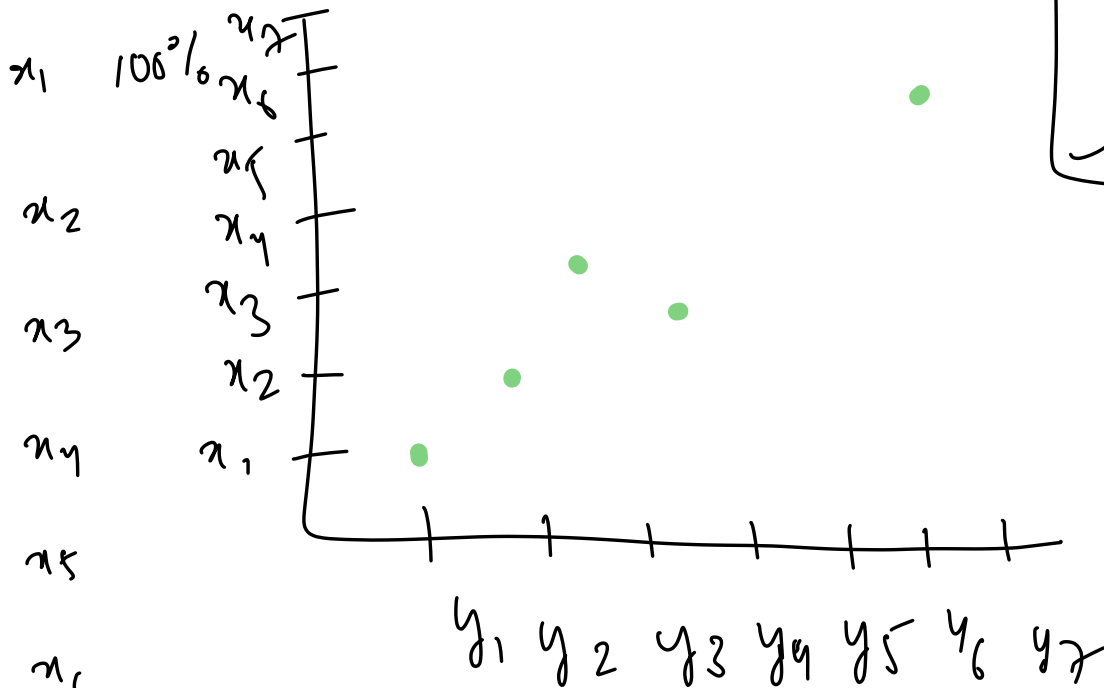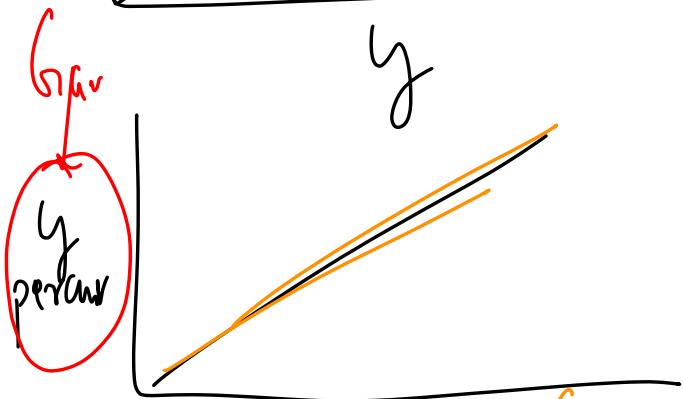
y quantile   GaussianQuant

percentile } y

x
percent

y (Gaussian)

percent
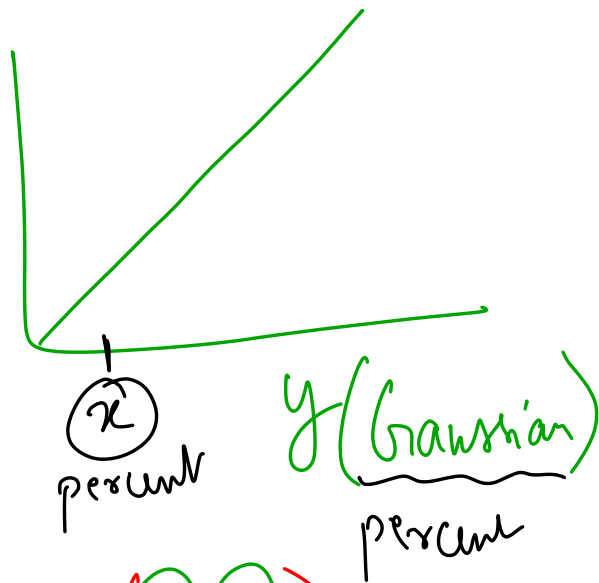
KS (S_1, S_2)

KS (S_1, G) Special case.

Covariance, Correlation

Height (inches)	Weight (kg)
68	72
62	58
64	67
61	72
70	79
66	61
61	68
65	64
71	80
72	79
$\bar{h} = 66$	$\bar{w} = 70$
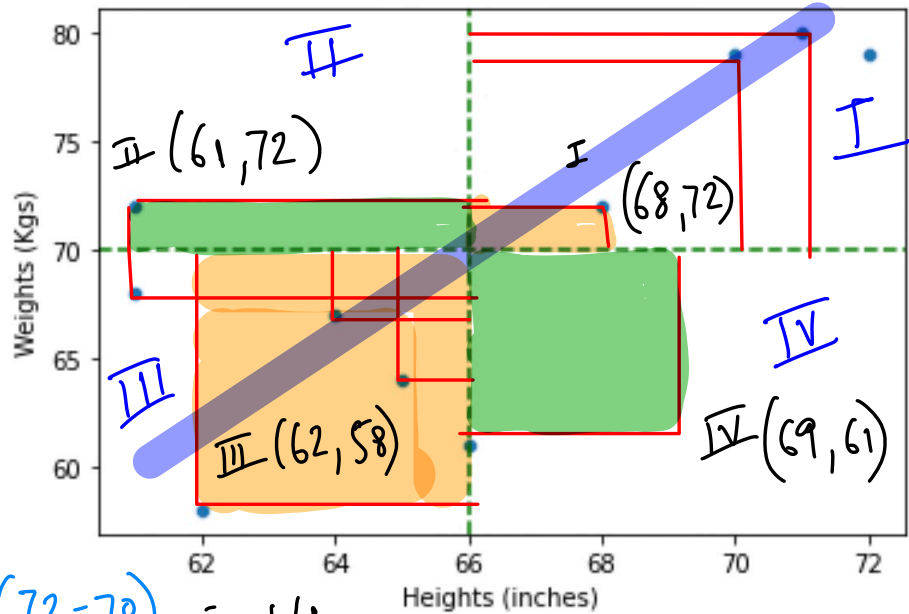
$I \quad (68 - 66)(72 - 70) = +4$
$\qquad +2 \qquad +2$

$II \quad (61 - 66)(72 - 70) = -10$
$\qquad -5 \qquad +2$

$III \quad (62 - 66)(58 - 70) = +48$
$\qquad -4 \qquad -12$

$IV \quad (69 - 66)(61 - 70) = -27$
$\qquad +3 \qquad -9$

I & III   +ve

II & IV   −ve

$$\text{Covariance} \quad \frac{1}{n} \sum \left(x_i - \bar{x} \right) \left(y_i - \bar{y} \right)$$

+ve

inches   kgs

$$\frac{20.4 \quad \text{inch} \times \text{kgs}}{\sigma_x \sigma_y \quad \text{inch} \times \text{kgs}}$$

Ice Cream Sales

Rains
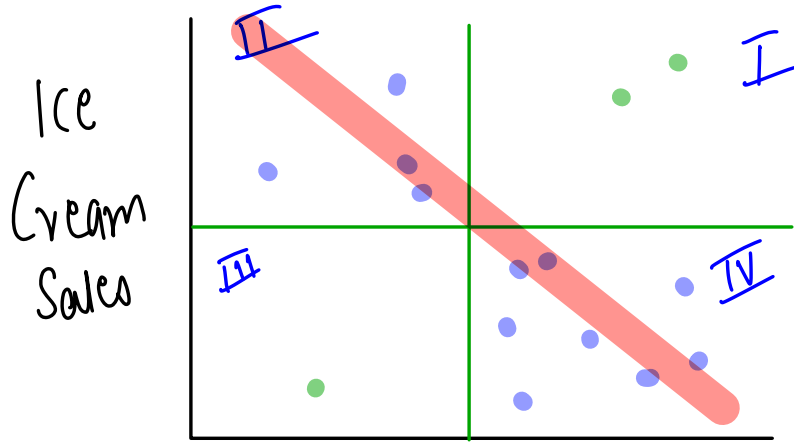
Covariance = -ve
-vely correlated

Chai Pakoda

Rains
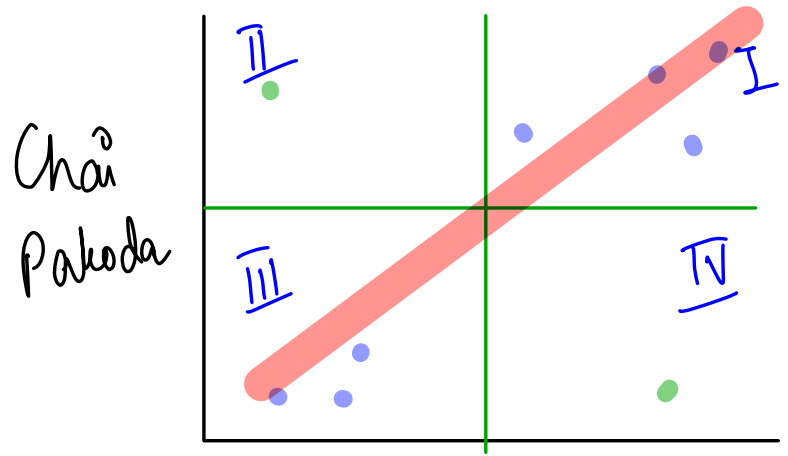
Covariance = +ve
+vely correlation
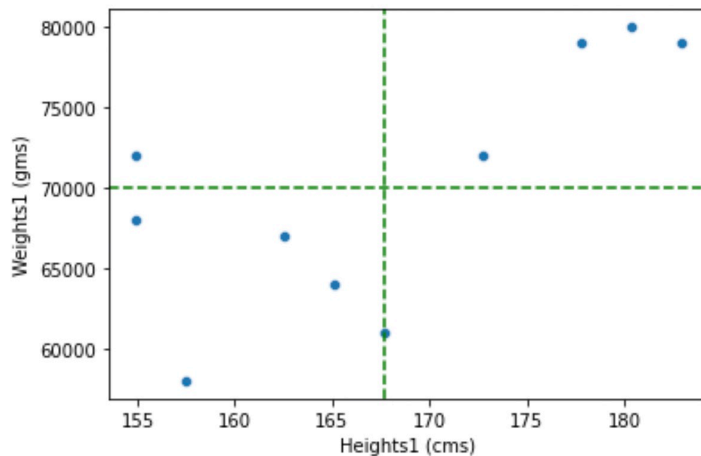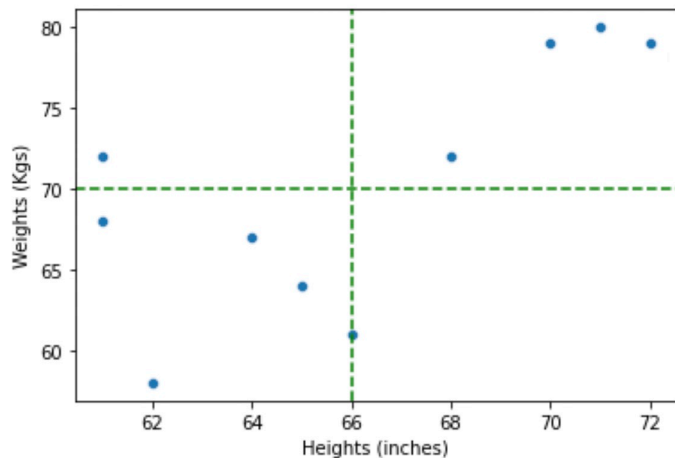
$$\text{Correlation} = \frac{1}{n} \sum \frac{(x-\bar{x})}{\sigma_x} \frac{(y-\bar{y})}{\sigma_y}$$

'rho'

$$\rho_{xy} = \frac{Cov(x,y)}{\sigma_x \, \sigma_y}$$

$$-1 \leq \rho_{xy} \leq 1$$

① Categorical v/s Categorical → Chi Square

① Goodness of fit
② Test of Independence

② Numerical v/s Numerical → Correlation Tests.

③ Numerical v/s Categorical → t test   → KS Test
                                            QQ Plot
2 Categories

④ Numerical v/s Categorical → ANOVA   → QQ Plot
                                          KS test
>2 Categories
                                          Shapiro
                                          Kruskal
                                          Levene

$$\begin{array}{|c|c|c|c|}
\hline
1 & n_1 & n_2 & 10 \\
\hline
2 & n_3 & n_4 & 10 \\
\hline
3 & 4 & n_5\,3 & 10 \\
\hline
6 & 9 & 15 & 30 \\
\hline
\end{array}$$

$$n_2 + n_3 = x_2 - n_1$$
$$n_4 + n_5 = x_3 -$$
$$n_2 + n_4 = y_2 -$$
$$n_3 + n_5 = y_3 -$$

$$n_2 + n_4 = 12$$
$$n_1 + n_3 = 5$$
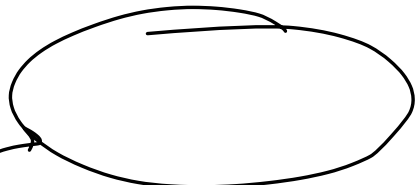$$n_1 + n_2 = 9$$
$$n_3 + n_4 = 8$$
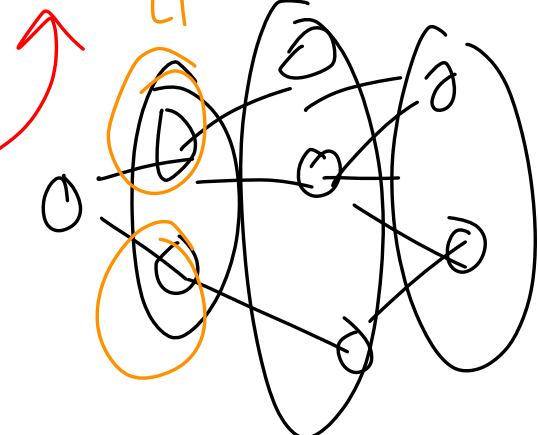
OHE

Count Vectoriser

TF-IDF

Word2Vec → Glove

Transformer

Vectors

L1 = 2
L2 = 3
L3 = 2

L1   L2   L3

Into to ML &NN

ML 1.1 1.2   2

NN | C V | NLP

MLOPS