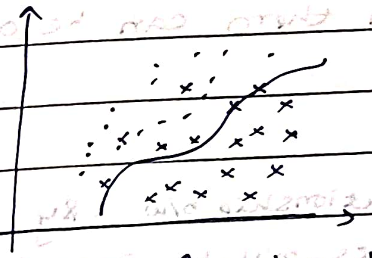
M6 LEC I REGRESSION ANALYSIS

Linear Regression, MSE and polynomial regression

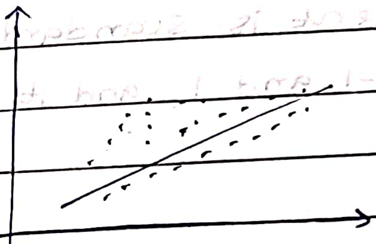
* classification vs Regression

classification : Predicts a discrete value / class label.



Ex: spam filtering, Image classification.

Regression : Type of ML that predicts a continuous value.



Ex: a house's [Area, age] (x) vs its Price (y)

* Covariance from PCA

→ Covariance tells us about the amount of dependency between two variables.

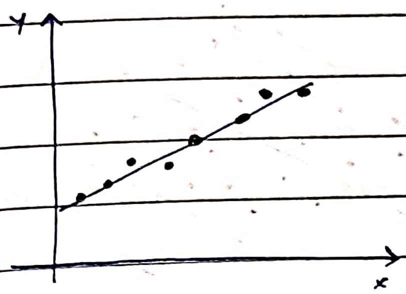
$$cov(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)}$$ feature

$cov(x, Y) > 0$ → x and Y are positively related

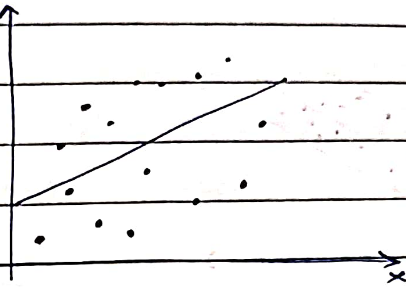
$cov(x, Y) < 0$ → x and Y are inversely related

$cov(x, Y) = 0$ → x and Y are independent

* Scatter Plots of Data with various correlation coefficients

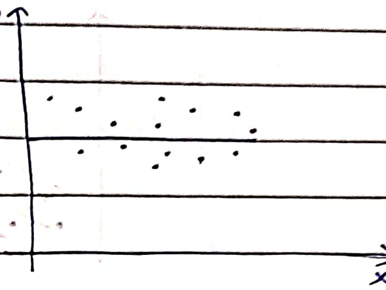


$r = +1$



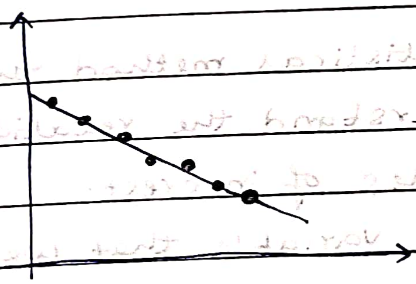
$r = +0.3$

- relation is there
- increasing only for some values
- relationship is not obvious here and not strong.

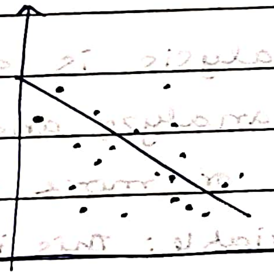


y value is constant no matter how much x changes

no relation b/w x & y



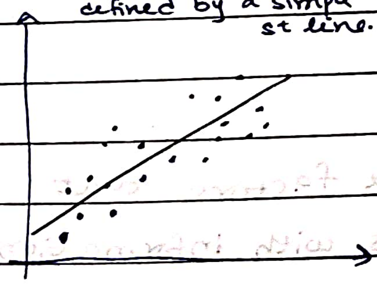
$r = -1$

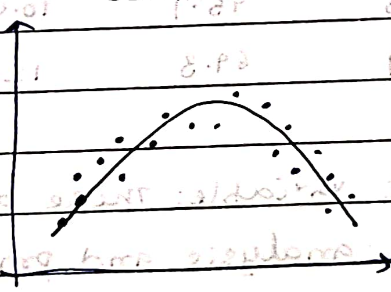


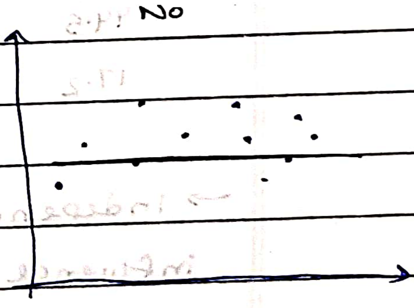
$r = -0.6$

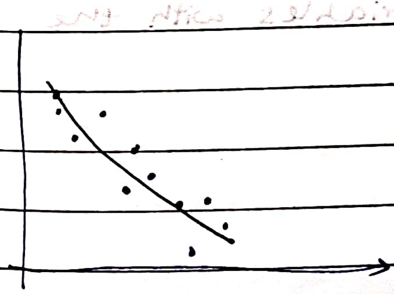
* linear correlation

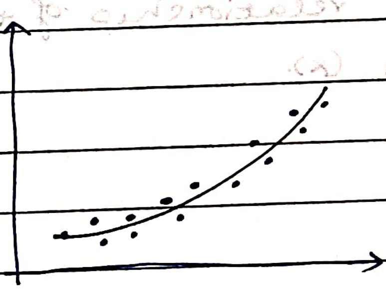
Relationship: Linear defined by a simple st line. | Curvilinear | No

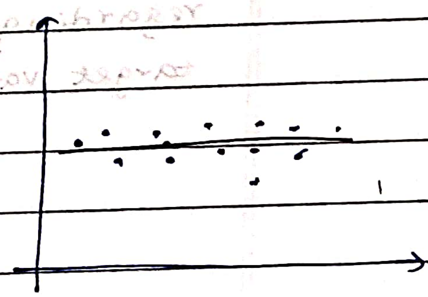






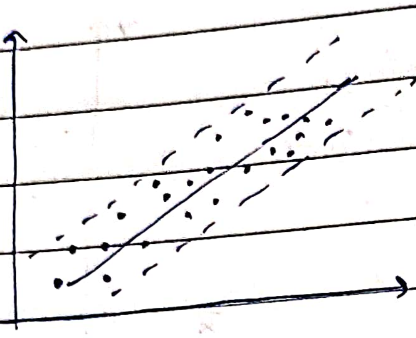


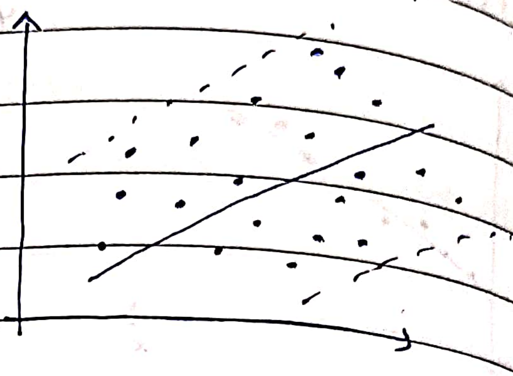




strong relationships

weak relationships





* Regression Analysis

→ The two variables (x_i, y_i) are treated as equals in correlation.

→ Regression analysis is a statistical method that helps us to analyze and understand the relationship between two or more variables of interest.

→ Dependent Variable: This is the variable that we are trying to forecast (Y)

TV	Radio	Newspaper	(Y) Sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	12

→ Independent variable: These are the factors that influence the analysis and provide us with information regarding the relationship of the variables with the target variable (X).

* Correlation Coefficient

→ Assuming a linear relationship between the variou
the relative strength between them can be observ

→ Normalization & correlation
↓
compute relationship b/w x & y
but your values will be standardiz
(dependency range [-1, 1])

→ Pearson's correlation coefficient is standardized covariance ranging between -1 and 1, and it is unitless.

$$r = \frac{cov(x, y)}{\sqrt{var x}\sqrt{var y}}$$

- $r = 1$ Perfect +ve linear correlation
- $r = -1$ Perfect -ve linear correlation
- $r = 0$ No linear correlation

(we assumed in PCA;
all variables when they are related
⇒ consider it as linear relationship.

* Linear Regression

→ Linear regression is a predictive model used for finding the linear relationship b/w a dependent variable and one or more independent variables.

$$y = mx + b$$

dependent variable

independent variable

y-intercept

slope
(Gradient, determines changes in y, per unit change in x)

NOTE:

Correlation does not imply causation

x is increasing because y is increasing. X NOT TRUE

Even if two variables are correlated, it does not necessarily mean that changes in one variable cause changes in the other.

Ex: In summers; swimming & ice-creams