06.05.18.

Udacity → Machine learning

Arthur Samuel:

The field of study that gives computers the ability to learn without being explicitly programmed.

Tom Mitchell:

The field of machine learning is concerned with the quest of how to construct computers program that automatically improve with experience. A com. prog. is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

i.e. Checkers Game:

E = the experience of playing

T = the task of playing checkers

P = the probability that the program will win the next game.

Machine learning algorithms:

1. Supervised learning
2. Unsupervised learning → Semi Supervised.
3. Reinforcement learning

Acerous Vs Non-Acerous

1. Supervised Learning:

SL is where you have input variables (x) and output variables (Y) and you use an algorithm to learn the mapping function from the input to the output.

$$Y = f(x)$$ (level data)

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict output variable (Y) for that data

Labels

Training Data

↓

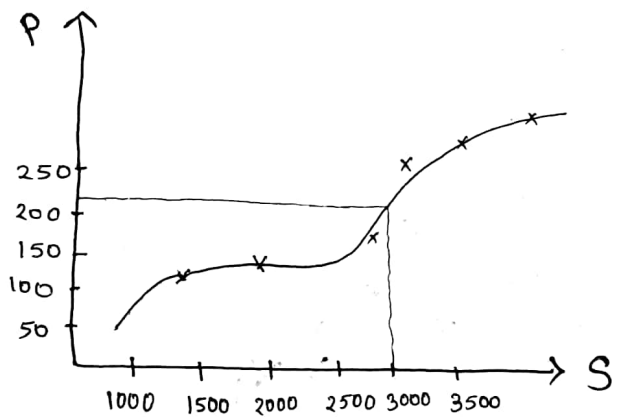
ML. Algo.

↓

New Data → Predictive Model → Prediction.

Test data

Model



$y = f(x)$

$P = f(S)$

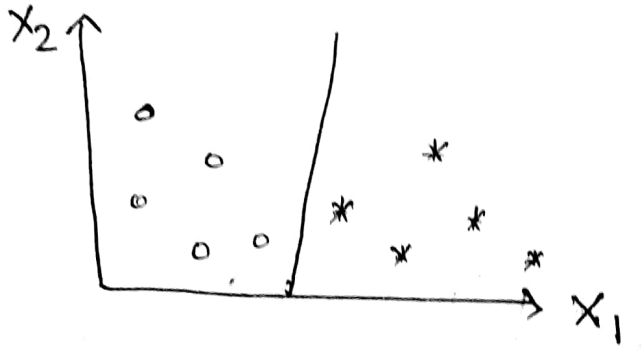
—x—

08.05.18

Supervised Learning → Regression
↳ Predict results within a continuous output

→ Classification
↳ Predict result in discrete output.

* Classification:



* Unsupervised Learning:

Here we have little or no idea what our result should look like.

input variable (x) but no Y	Underlying structure

Unsupervised Learning → Clustering
↳ want to discover the inherent grouping in the data such as group customers by purchasing behavior. (UL classification)
→ Association.

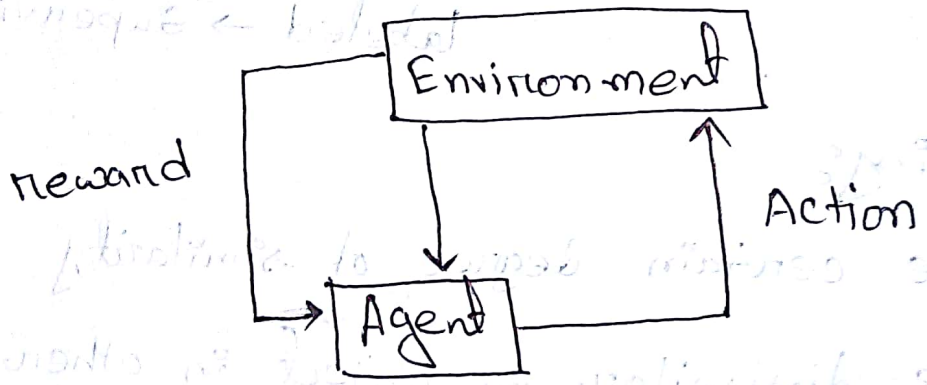
labeled → supervised

* Clustering:
  - Share certain degree of similarity.
  - more dissimilar to object in other clusters.

* Association:
  ↳ Discover rules that describe large portions of our data.
  ↳ such as: People that buy x also tend to buy Y.

* Reinforcement Learning:

↳ The goal is to develop a system (agent) that improve its performance based on interactions with the environment, no answer key.



— x —

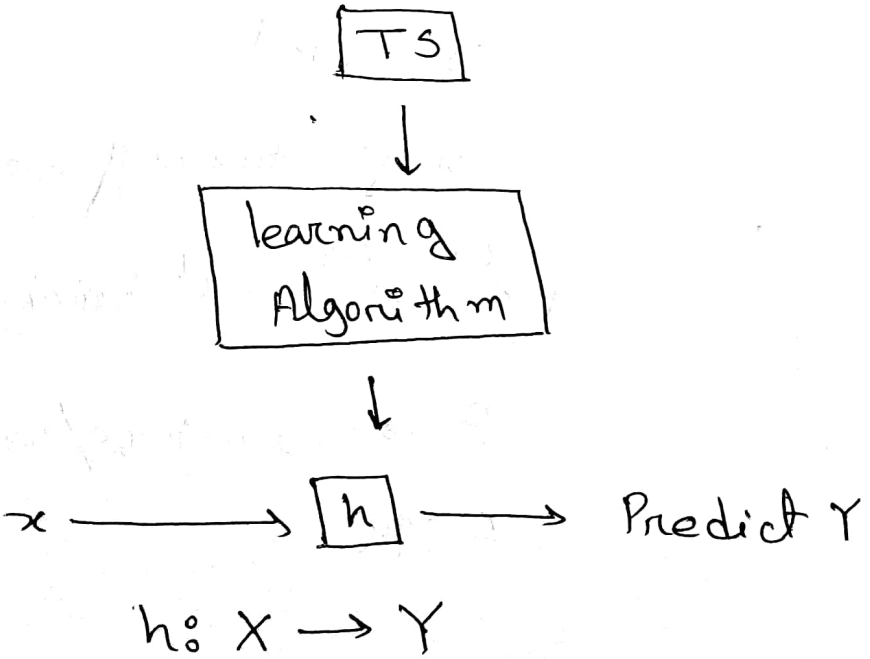
13.05.18.

* Linear regression with one variable:

Model Representation: [MR]

TS → Training Set

$Feet^2$	1000$s
2104	400
1600	330
2400	369
1416	540
⋮	⋮



$h: X \rightarrow Y$

*Univariant linear regression.

$h_\theta(x) = \theta_0 + \theta_1 x$

m = # No. of training example

n = # No. of variables

x = input

y = output

(x, y) = training example

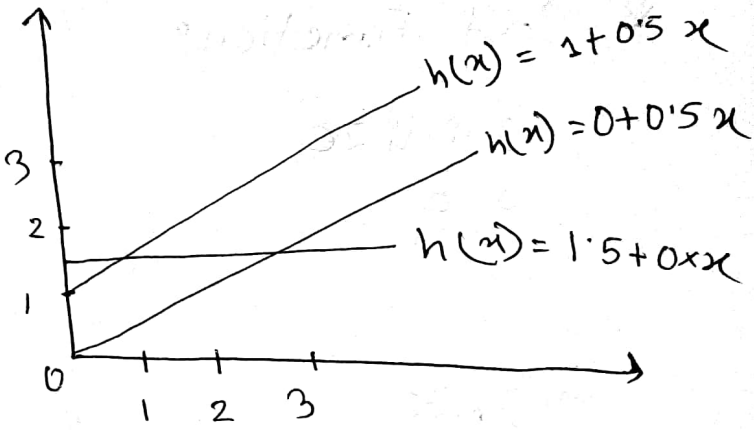
$(x^{(i)}, y^{(i)})$ = i^{th} training example

$\theta \rightarrow$ Parameters/weights

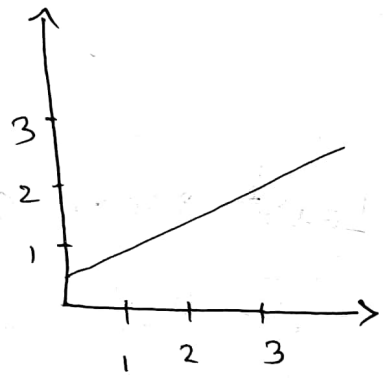
$\theta \in \mathbb{R}$

$h_\theta(x) = \theta_0 + \theta_1 x$

$\theta_0 = 1.5$	$\theta_0 = 0$	1
$\theta_1 = 0$	$\theta_1 = 0.5$	0.5



θ_0 =	0	0.5	1	1
θ_1 =	1	1	0.5	1



i.e.,

X	Y
0	4
1	(7)
2	7
3	8

$h_\theta(x) = \theta_0 + \theta_1 (x)$; $\theta_0 = \theta_1 = 2$

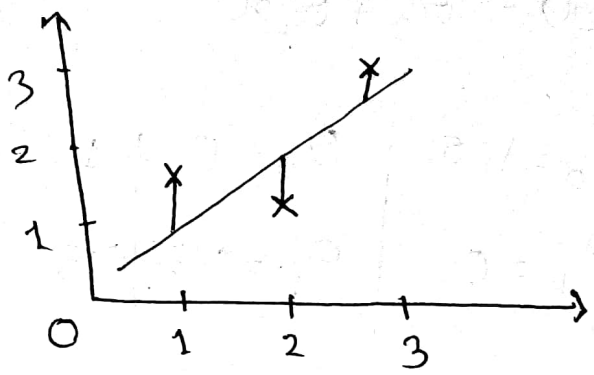
$= 2 + 2 * x$

if $x = 0 \rightarrow 2 + 2.0 = 2$

$1 \rightarrow 2 + 2 \times 1 =$ (4) X

* Cost Function:

minimize
$\theta_0 \theta_1$



minimize
$\theta_0 \theta_1 \quad (h_\theta(x) - y)^2 \rightarrow \min_{\theta_0 \theta_1} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$

Least square Errors.

Least mean square Errror (LMS):

$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$

↓

Minimize
θ_0, θ_1

Hypothesis : $h_\theta(x) = \theta_0 + \theta_1 x$

Parameter : θ_0 and θ_1

Cost function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

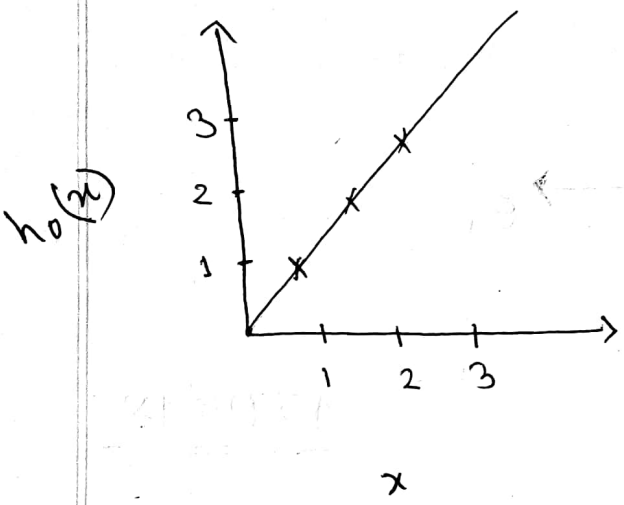
Goal: minimize
$\theta_0, \theta_1 \quad J(\theta_0, \theta_1)$

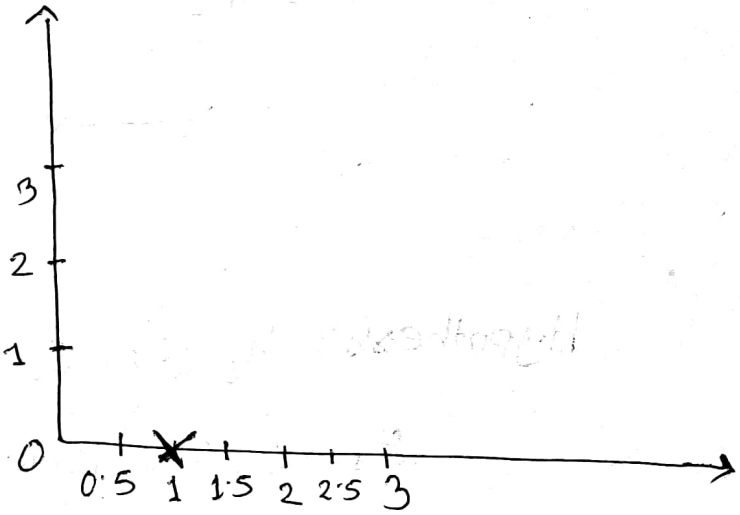
$h_\theta(x) = \theta_0 + \theta_1 x$ Let, $\theta_0 = 0$

$h_\theta(x) = \theta_1 x$

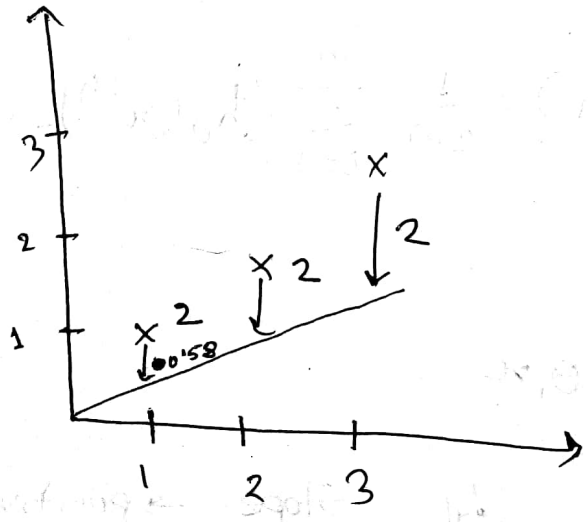
if $\theta_1 = 1$

then, $h_\theta(x) = x$ [$Y = X$]





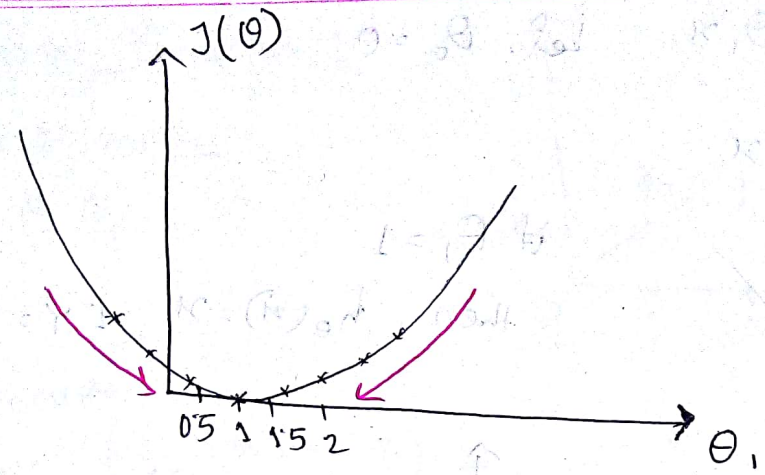
As distance 0 (1-1=0)



$\theta_1 = 0.5$

$$J(0.5) = \frac{1}{2m}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right]$$

$$= \frac{1}{2\times 3}(\quad) \qquad = 0.58$$



—x—

15.05.18

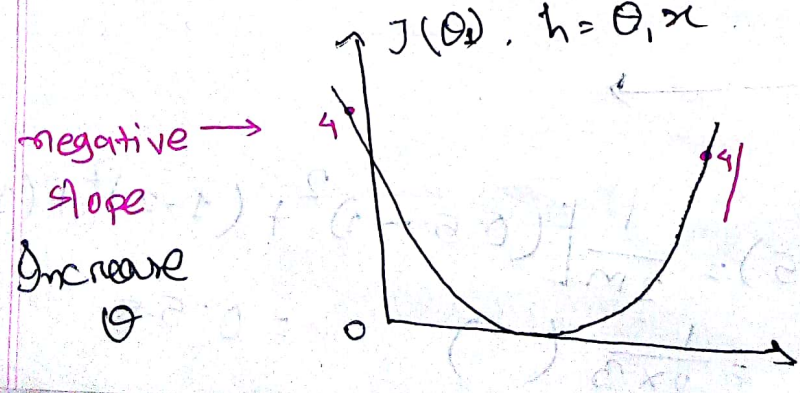
Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

θ_1, θ_0 → Parameters

Cost function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$

Goal: Minimize

θ_0, θ_1

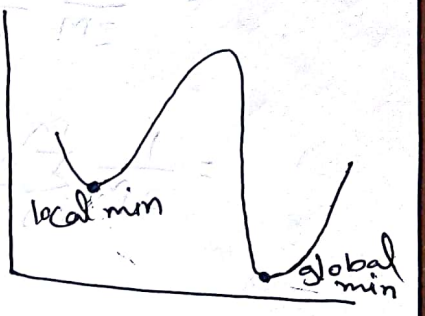


Gradient Descent:

→ start with θ_0, θ_1 (any value)

→ Keep changing θ_0, θ_1, until we have minimum $J(\theta)$ (Hopefully)

→



problem in gradient descent

//we get more global solution

repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \quad ; \quad j = 0, 1$$

}

repeat

α = learning rate

{

$$\theta_0 = \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta) = \theta_0 - \alpha \frac{\partial \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2}{\partial \theta_0}$$

$$\theta_1 = \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta) = \theta_1 - \alpha \frac{\partial \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2}{\partial \theta_1}$$

}

When $j=0$,

$$\theta_0 \rightarrow \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$= \frac{1}{2m} \frac{\partial}{\partial \theta_0} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$= \frac{1}{2m} \sum_{i=1}^{m} 2 \cdot \left(h_\theta(x^{(i)}) - y^{(i)} \right) \frac{\partial}{\partial \theta_0} \left(h_\theta(x^{(i)}) - y^{(i)} \right)$$

$$= \frac{1}{2m} \cdot 2 \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \frac{\partial}{\partial \theta_0} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right)$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \cdot 1$$

$$\therefore \theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)$$

When, $j=1$,

$$\theta_1 = \frac{\partial}{\partial \theta_1} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$= \frac{1}{2m} \frac{\partial}{\partial \theta_1} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$= \frac{1}{2m^2} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) \frac{\partial}{\partial \theta_1} \left(h_\theta(x^{(i)}) - y^{(i)}\right)$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) \frac{\partial}{\partial \theta_1} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

$$\therefore \theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

repeat

{

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

} $\theta_0 - \theta_0 \geq \varepsilon$ or $J(\theta_1) - J(\theta_0) \geq \varepsilon$ ... //Condition to stop.

update simultaneously:

{ $temp0 = \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta)$

$temp1 = \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta)$

$\theta_0 = temp0$

$\theta_1 = temp1$

}

Batch Gradient descent:

↓ Solution

Incremental / Stochastic Gradient descent:

(Fast, but possibility of getting global solution is less.)

Loop {

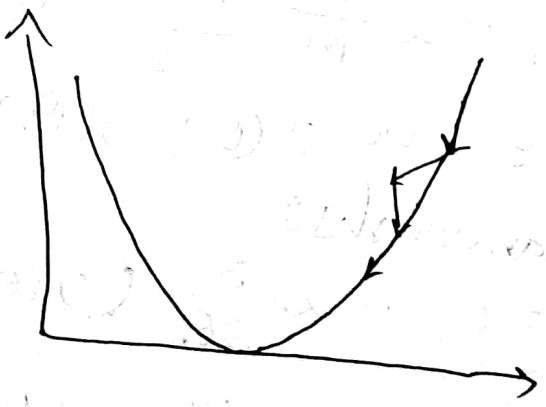
for i = 1 to m {

$$\theta_j = \theta_j - \alpha\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)}$$

}

}

; for j = 0, 1



—x—

"Special class"

17.05.18.

eliademy.com

Python

Introduction to Machine Learning and Decision Tree

aktarhossain@daffodilvarsity.edu.bd

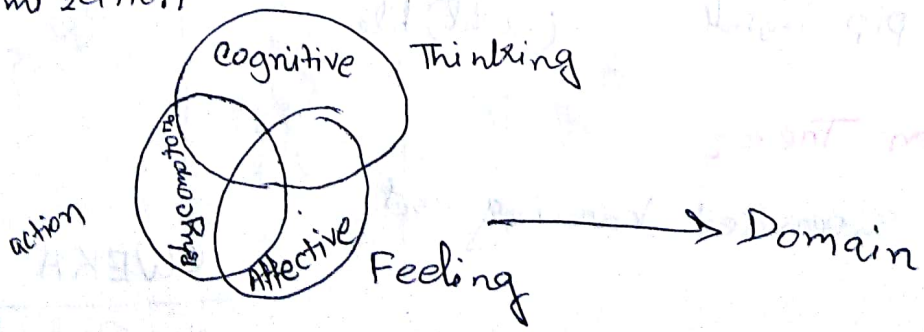
Industry 4.0

* Three components of ML:

Representation

Evaluation

Optimization



Bloom's Taxonomy of learning

Supervised or inductive learning:

markany.com

$2^{**}100$ → Power.

every single type is tuple

Python Basics.pdf.

Everything in python is an object.

1. get-pip.py

Install pip installer

C:\> Python get-pip.py (Intenet)

Pallindrome

2. Install Numpy & sci-pi

C:\> pip install ---(.whl) file

Decision Tree:

Supervised learning set

"WEKA" → DM Tool

New Zealand Waikato University

20.05.18.

x_0	x_1 feet^2	x_2 # bedrooms	x_3 # floor	x_4 Age	Price
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1535	3	2	30	315
1	852	2	1	36	178

n = # features

$x^{(i)}$

$$x^{(2)} = \begin{bmatrix} 1416 \\ 3 \\ 2 \\ 40 \end{bmatrix} \begin{matrix} \rightarrow x_1 \\ \rightarrow x_2 \\ \rightarrow x_3 \\ \rightarrow x_4 \end{matrix}$$

$$x^{(2)}_3 = 2$$

$$x^{(i)} = \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \\ x_3^{(i)} \\ x_4^{(i)} \end{bmatrix} \in \mathbb{R}^4$$

$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

$$h_\theta(x) = \theta_0 + \theta_1 x + \dots + \theta_n x_n$$

Let, $x_0 = 1$ for all $i = 1$ to m

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \dots + \theta_n x_n$$

$$= \sum_{j=0}^{n} \theta_j x_j = \theta^T x \quad \text{(no. of features)}$$

$\theta^T x$

$$x^{(i)} = \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix} \in \mathbb{R}^{n+1} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

$n+1$ ~~dimen~~ dimensional vector

$(n+1) \times 1$ matrix

$$\theta^T x = \begin{bmatrix} \theta_0 & \theta_1 & \dots & \theta_n \end{bmatrix} \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix}$$

* Gradient Descent:

$$J(\theta_0, \theta_1, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(\theta^T x^{(i)} - y^{(i)} \right)^2 \quad (\text{no. of examples } m)$$

$$= \frac{1}{2m} \sum_{i=1}^{m} \left(\left(\sum_{j=0}^{n} \theta_j x_j^{(i)} \right) - y^{(i)} \right)^2$$

when, $n = 1$

Repeat {

$$\theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)$$

$$\theta_1 = \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x^{(i)}$$

}

when $n \geq 2$

Repeat {

$$\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)}$$

}

update simultaneously.

Assignment

Derivation of → Next class

$$\theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_0^{(i)}$$

$$\theta_1 = \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_1^{(i)}$$

$$\vdots$$

$$\theta_n = \theta_n - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_n^{(i)}$$

Gradient Descent in practice:

→ Feature Scaling

→ Mean ~~nor~~ normalization

→ Debug

→ select α

x_1 = size = 100 − 2000 feet2

x_2 = # bedrooms = 2 − 5

* θ will descent quickly on small range

So, set input value in same range (roughly).

$-1 \leq x^{(i)} \leq 1$

$-0.5 \leq (x^{(i)}) \leq 0.5$

$\boxed{-3 \leq x^{(i)} \leq 3}$

Feature Scaling:

$$x_j = \frac{x_j}{\text{range}} = \frac{2104}{2400} \approx 1 \quad ; \text{for } x_1$$

$$x_2 = \frac{5}{3} = 1.5$$

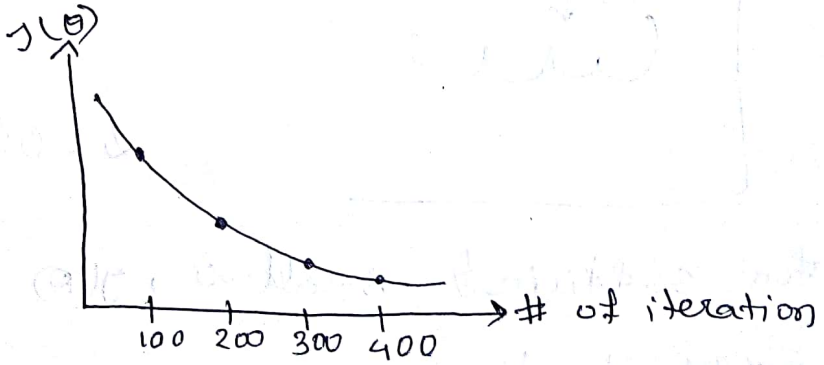
Mean normalization:

$$x_j = \frac{x_j - \mu_j \text{ (mean avg.)}}{s_j \text{ (range)}}$$

$$\mu_j = \frac{5+3+3+2}{4}$$

$$x_2^{(1)} = \frac{5 - 3.25}{3}$$

Debug:



$J(\theta) = 5$ (2)

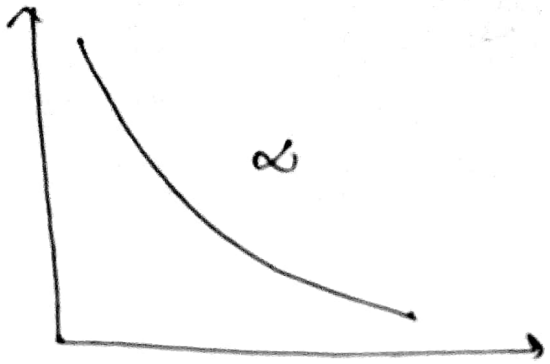
$J(\theta) = 4.99999$ (3)

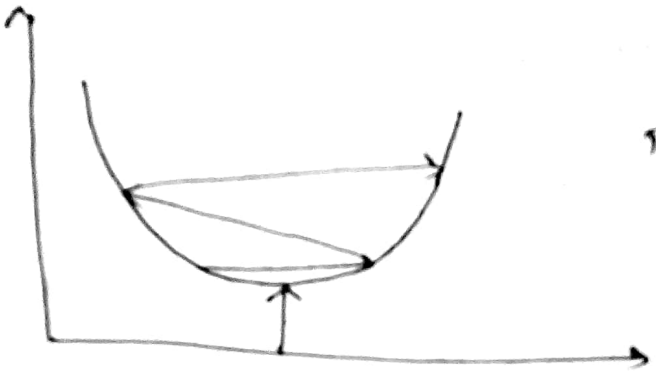
Change learning rate, α

→ Threshold value $\approx 10^{-3}$

* If the change is below threshold value then we can assume to get optimal solution.

Select α:



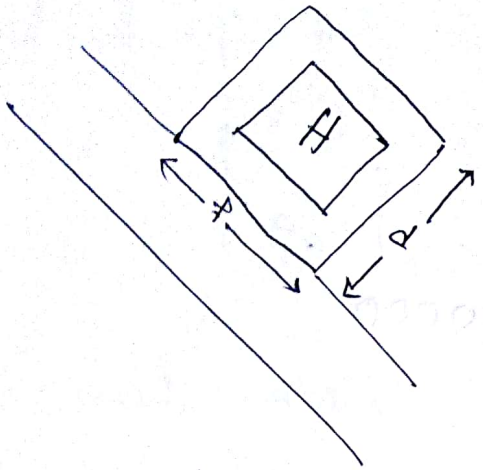


reduce the value of α

$\alpha = 0.001, 0.003, 0.01, 0.03$

* for sufficient small α, $J(\theta)$ increase for every iteration.

22.05.18.

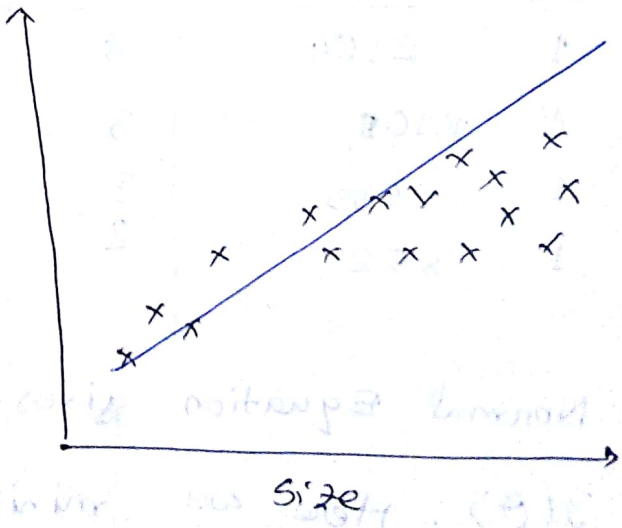


$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

(x_1: f, area; x_2: d)

$$= \theta_0 + \theta_1 x$$

*Polynomial Regression



$x_1 = size$

$x_2 = size^2$

$x_3 = size^3$

$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$h_\theta(x) = \theta_0 + \theta_1 x^2$$

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

Feature Scaling:

$1 - 1000$

$x_1 = 1000$

$x_2 = 1000000$

$x_3 = 10^9$

$h_\theta(x) = \theta_0 + \theta_1 \text{ size} + \theta_2 \sqrt{\text{size}}$

Normal Equation:

x_0	feet^2	#bedrooms	#floor	Age	Price
1	2104	5	1	45	316
1	1416	3	2	40	420
1	1535	3	2	30	290
1	852	2	1	35	330

* Normal Equation gives other way of minimizing $J(\theta)$. Here we minimize $J(\theta)$ explicitly taking its derivatives w.r.to the θ_j and setting them to zero. This allow us to find the optimal θ without iteration.

$$\theta = (X^T X)^{-1} X^T y$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} = \mathbb{R}^{n+1} \qquad \begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \text{output}.$$

X = input matrix

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1535 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 35 \end{bmatrix}$$

$$x^{(1)} = \begin{bmatrix} 1 \\ 2104 \\ 5 \\ 1 \\ 45 \end{bmatrix} \qquad x^{(2)} = \begin{bmatrix} 1 \\ 1416 \\ 3 \\ 2 \\ 40 \end{bmatrix}$$

$$X = \begin{bmatrix} x_i^{(1)T} \\ x_i^{(2)T} \\ \vdots \\ x_i^{(m)T} \end{bmatrix}$$

X X^T

$m \times (n+1)$ $(n+1) \times m$ ⟶ dimension

4×5

Feature scaling not needed

Gradient Descent

1. Need to choose α.
2. Need many iteration.
3. $O(kn^2)$
4. Works well if n is large

Normal Equation

1. No need to choose α.
2. No iteration needed
3. $O(n^3)$ [$(X^T X)^{-1}$]

[if no. of feature is huge, then complexity would increase]

4. Slow if n is large

[n = no. of features]

* If $n \geq 1000$ it is better to use gradient descent

$(X^T X)^{-1}$ not invertible,

- 1. Redundant features
- 2. Delete feature $n \geq m$
- On regularization

* Assignment → finding weight from height using Python. Before → 5th June.

Sklearn linear regression

udacity

↓

Introduction to machine learning.

———×———

27.05.18

Classification :

Email : Spam / Not

Online Tran : Normal / Not

Tummer : Malignant / Benign

Binary classification : $y \in \{0, 1\}$

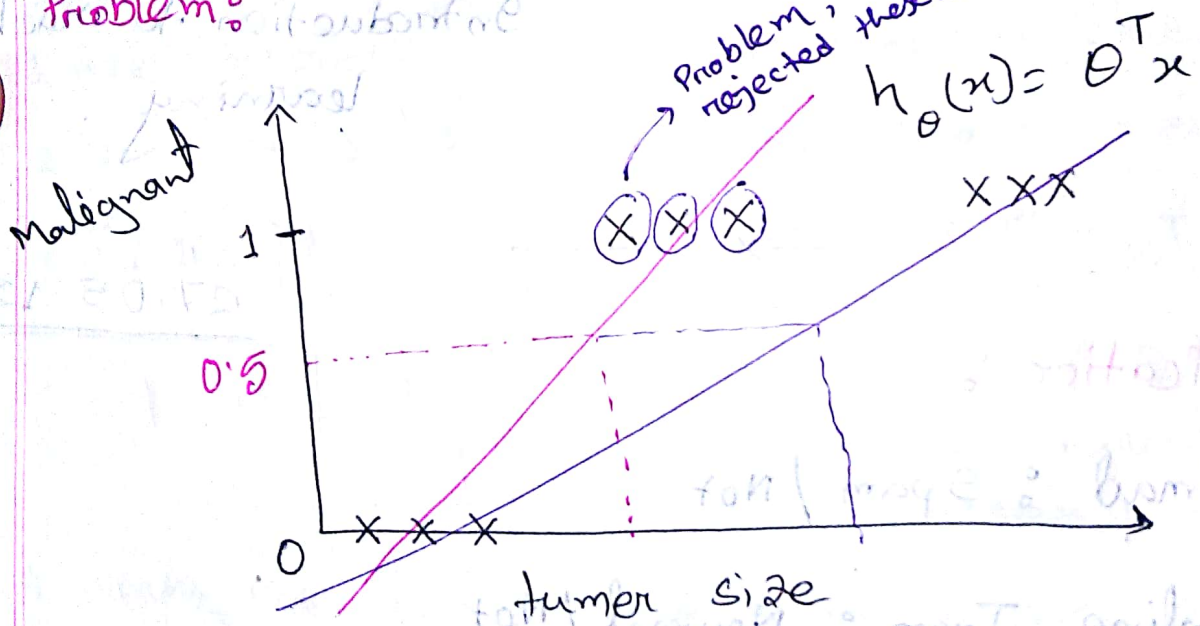
Multiclass classification : $y \in \{0, 1, 2, 3, \ldots\}$

0 : Negative class / Absence of something

1 : Positive class / Presence of something

* Problem with Linear Regression for classification

Problem:



Threshold value : 0.5

$h_\theta(x) \geq 0.5, \quad y = 1$

$h_\theta(x) < 0.5, \quad y = 0$

Problem:

$Y = 0$ or 1

$h_\theta(x)$ using Linear Regression may produce value

< 0 or > 1

* Solution :- →

Logistic Regression (Logistic function, sigmoid function)
[classification algorithm]

$$\frac{1}{1 + e^{-x}}$$

always produces result -

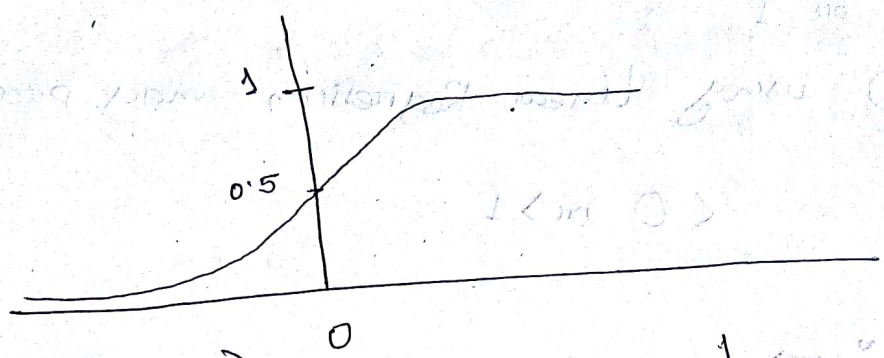
$$0 \leq h_\theta(x) \leq 1$$

Hypothesis:

$$h_\theta(x) = g(z) \quad ; \quad z = \theta^T x \;/\; \theta_0 + \theta_1 x$$

$$= \frac{1}{1 + e^{-z}}$$

$$= \frac{1}{1 + e^{-\theta^T x}}$$



$g(z)$

$g(0) = 0.5$

$g(\infty) = 1$

$g(-\infty) = 0$

$$\frac{1}{1+e^{-z}}$$

$$\frac{1}{1+e^{-\infty}}$$

$h_\theta(x)$

→ Estimate Probability that $y = 1$ on input x.

If $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumor\_size} \end{bmatrix}$ $= h_\theta(x) = 0.7$

$h(x) = P(y = 1 \mid x, \theta)$ Probability that $y = 1$, given x parameterized by θ

$y = 0$ or 1

$P(y = 1 \mid x, \theta) + P(y = 0 \mid x, \theta) = 1$

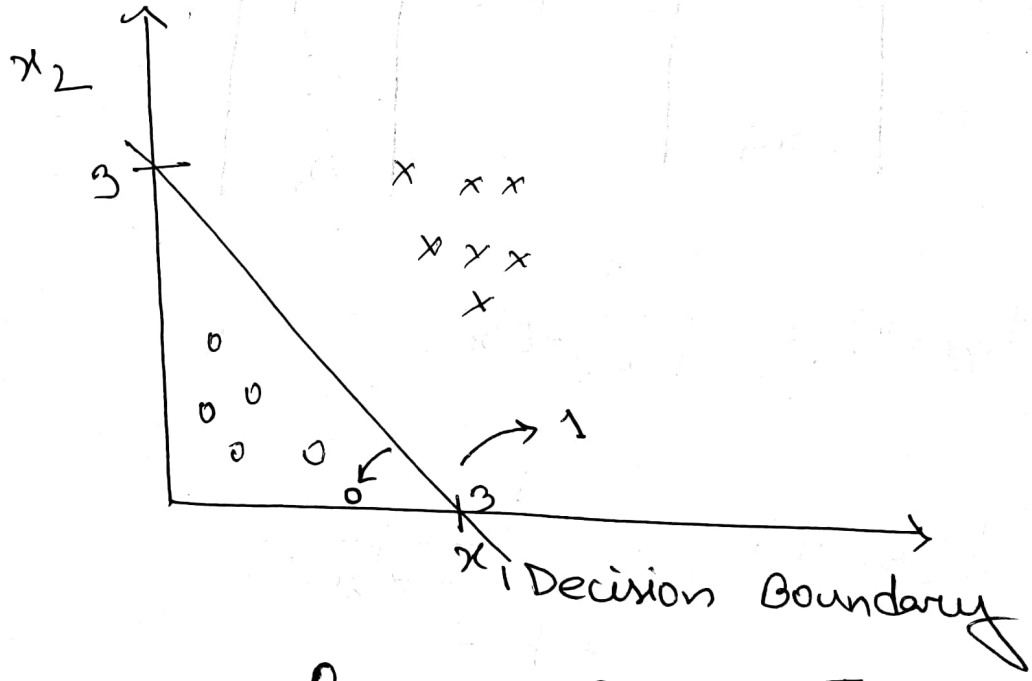
$P(y = 1 \mid x, \theta) = 1 - P(y = 0 \mid x, \theta)$

$P(y = 0 \mid x, \theta) = 1 - P(y = 1 \mid x, \theta)$

Predict, $y = 1 \longrightarrow h_\theta(x) \geq 0.5 \rightarrow \theta^T x \geq 0$

$y = 0 \longrightarrow h_\theta(x) < 0.5 \rightarrow \theta^T x < 0$

Decision Boundary: is the line that separates the area where $y = 0$ and $y = 1$. It is created by hypothesis function.



$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 = \theta^T x$$

$$\begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

$$-3 + 1x_1 + 1x_2$$

$$-3 + x_1 + x_2 \geq 0$$

$$-3 + x_1 + x_2 = 0$$

$$x_1 + x_2 = 3$$

— x —

29.05.18.

Logistic Regression:

Training Examples: $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})\}$

Examples: m

Features: n

$$x \in \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \theta \in \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

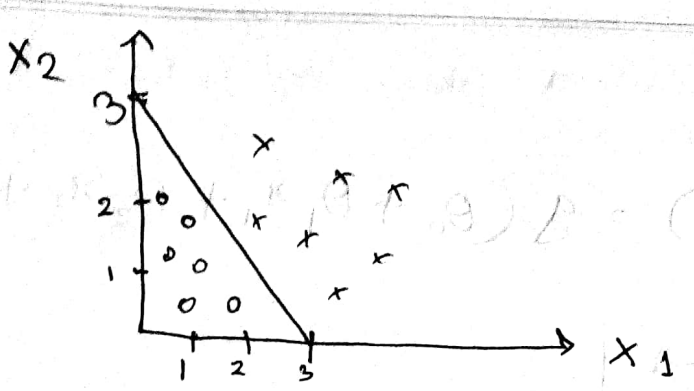
$$g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

$Y = 1$, $g(z) \geq 0.5$

$\theta^T x = 0$

if $\theta^T x \geq 0$, $Y = 1$

else $Y = 0$



$g(z) \longrightarrow h_\theta(x)$

$\theta^T x = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

$$\begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = -3 + x_1 + x_2$$

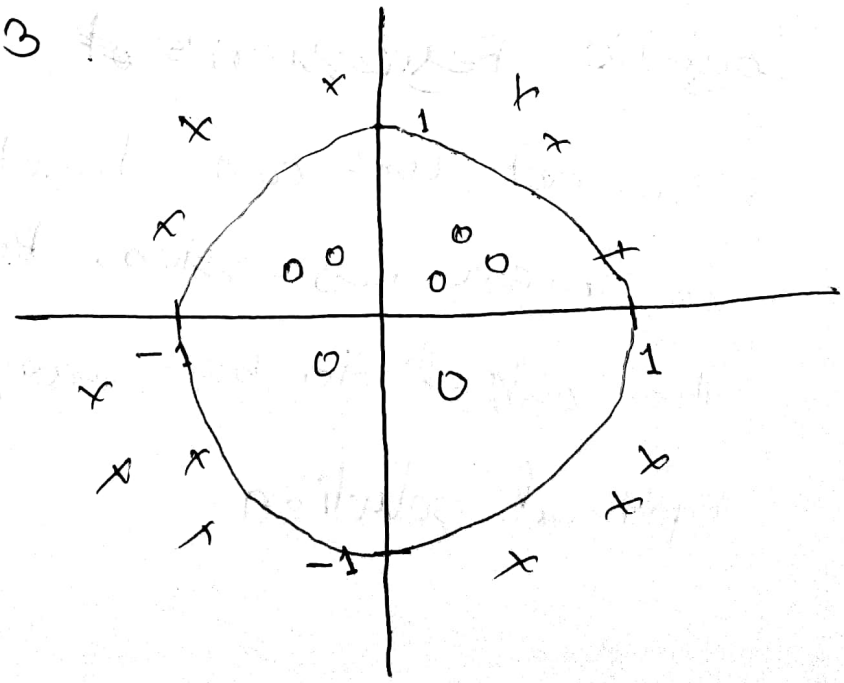
$Y = 1$ when $\theta^T x \geq 0$

$Y = 0$ when $\theta^T x < 0$

$-3 + x_1 + x_2 \geq 0$

$x_1 + x_2 \geq 3$

$x_1 + x_2 = 3$



$h_\theta(x)$

$$g(\theta^T x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

$$\theta = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = -1 + x_1^2 + x_2^2$$

$$-1 + x_1^2 + x_2^2 \geq 0$$

$$x_1^2 + x_2^2 = 1 \quad [\text{equ. for circle}]$$

* Decision boundary is a property of hypothesis, not of training set.

Logistic Regression's ~~of~~ cost function:

→ Can not use cost function that we've used in linear regression. Because, it will cause the output to be wavy, causing many local optimal solution.

→ it will not be ~~cost~~ convex function.

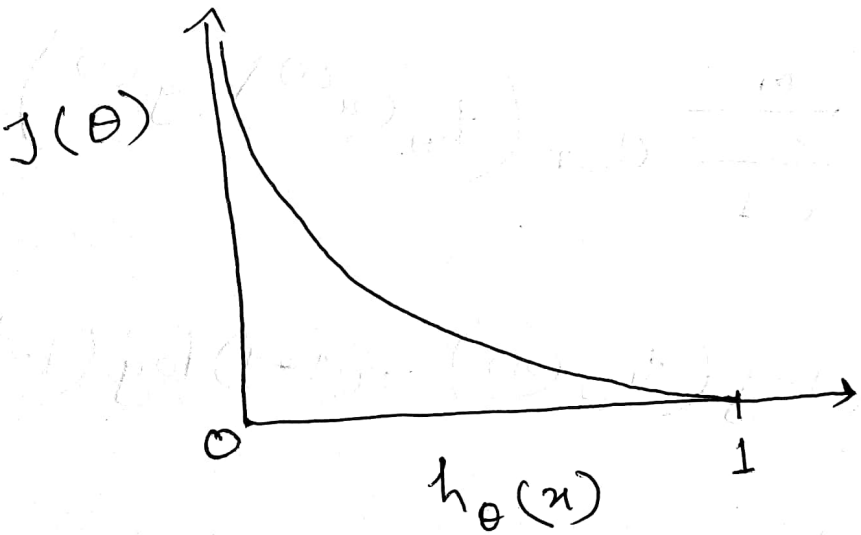
* Cost function:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \text{Cost}\left(h_\theta(x^{(i)}), y^{(i)}\right)$$

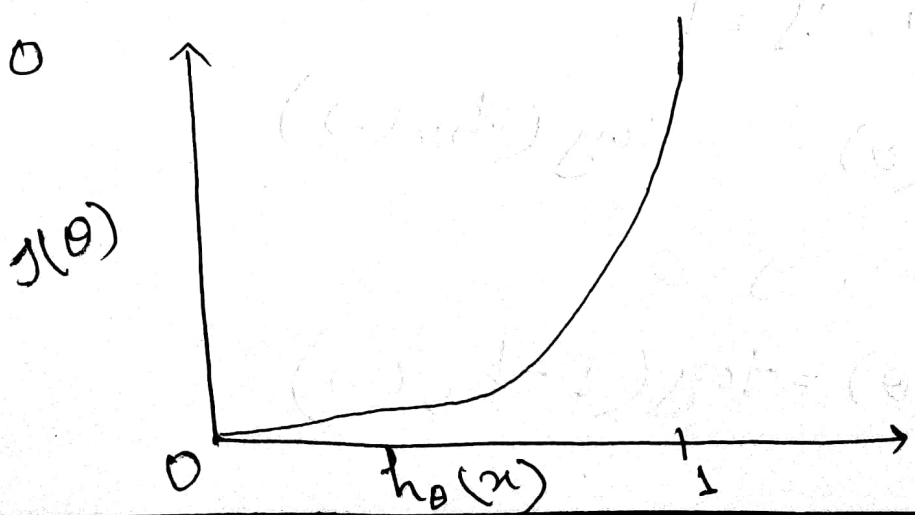
$$\text{Cost}\left(g\left(\sum_{i=0}^{n} \theta^T x^{(i)}\right), y^{(i)}\right)$$

$Y = 1$ or 0

when $y = 1$



$y = 0$



$$\text{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

$-\log(1) = 0$

$J(\theta) = 0, \quad y = h_\theta(x)$

$J(\theta) = \infty, \quad y = 1, \quad h_\theta(x) = 0$

$J(\theta) = \infty, \quad y = 0, \quad h_\theta(x) = 1$

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \text{Cost}\left(h_\theta(x^{(i)}), y^{(i)}\right)$$

$$J(\theta) = -y \log(h_\theta(x)) - (1-y) \log(1 - h_\theta(x))$$

when, $y = 1$

$$J(\theta) = -\log(h_\theta(x))$$

when, $y = 0$

$$J(\theta) = -\log(1 - h_\theta(x))$$

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log\left(h_\theta(x^{(i)})\right) + \left(1-y^{(i)}\right) \log\left(1-h_\theta(x^{(i)})\right)\right]$$

$$J(\theta) = \frac{1}{m}\left(-y^T \log(h) - (1-y)^T \log(1-h)\right)$$

Gradient Descent:

Repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

// update simultaneously

}

Repeat {

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)} - y^{(i)}\right) x_j^{(i)}$$

}

Vector form: $\theta := \theta - \frac{\alpha}{m} x^T (g(x\theta) - y)$

- BFGS
- L-BFGS

03.06.18.

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

↓

$g(z)$

0.5

$Y = 1 \quad \theta^T x \geq 0$

$$\begin{cases} 1 & -\log(h_\theta(x)) \\ 0 & -\log(1 - h_\theta(x)) \end{cases}$$

↓

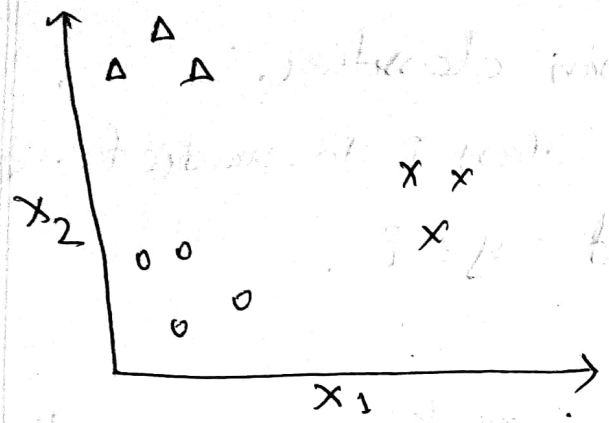
$g(z)$

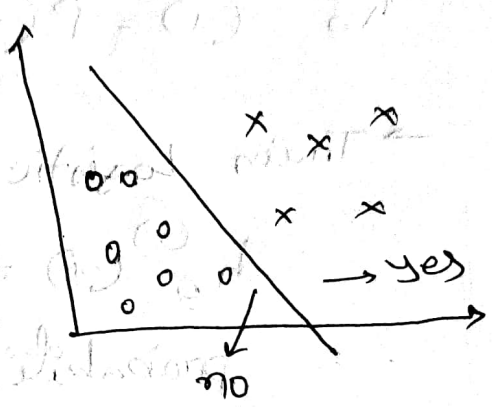
$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left(y \cdot \log(h_\theta(x)) + (1-y) \log (1 - h_\theta(x)) \right)$$

$Y = 1, \quad h_\theta(x) \to 0 \quad ;$ cost = infinity

$Y = h_\theta(x) \quad ;$ cost = zero

Multiclass classification:





binary classification

Multiclass classification

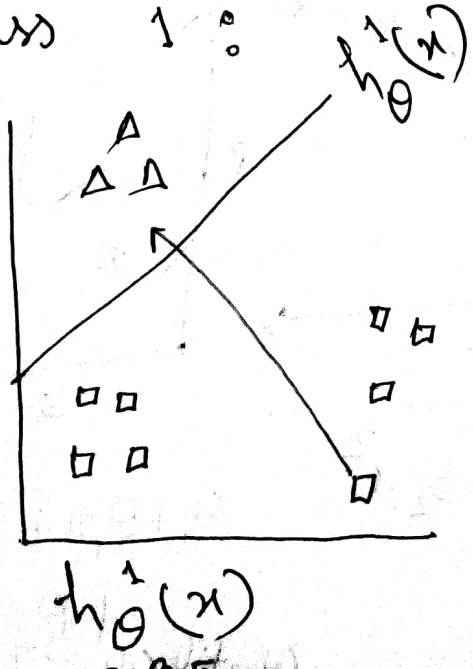
* One Vs all (one Vs rest):

class 1: Δ

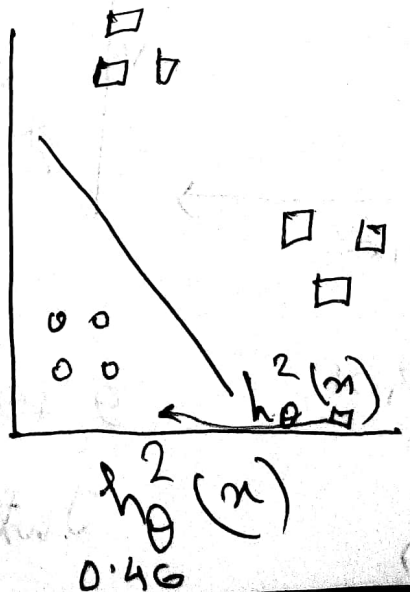
class 2: □

class 3: X

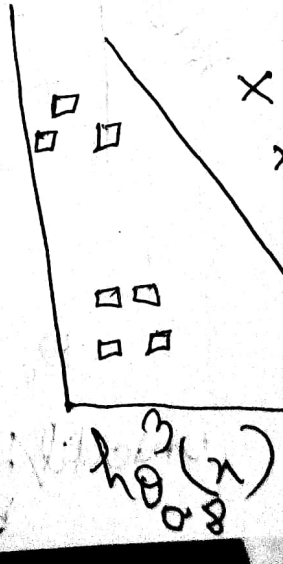
class 1:



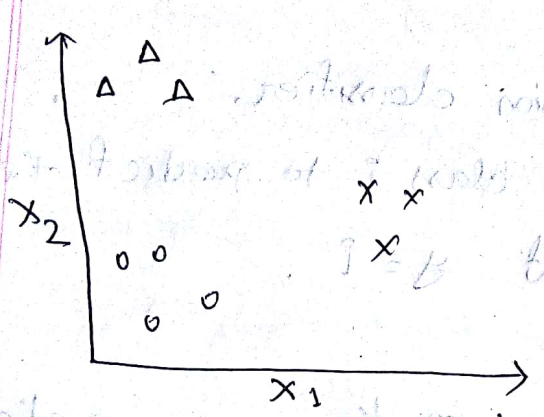
class 2:

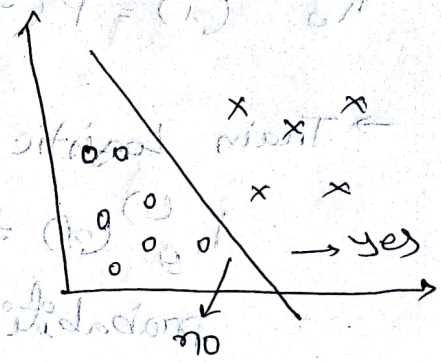


class 3:



Multiclass classification:





binary classification

Multiclass classification

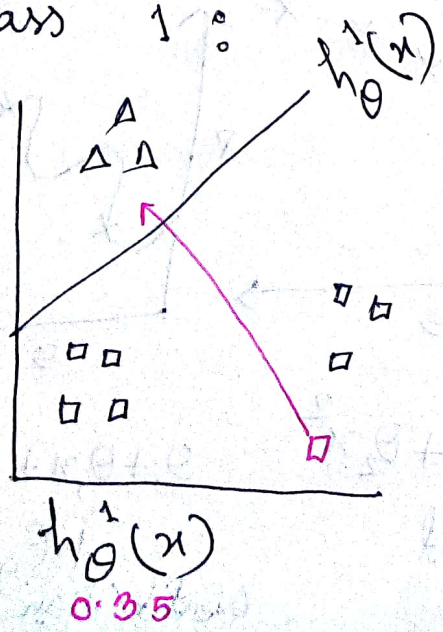
* One Vs all (one Vs rest):

class 1: Δ

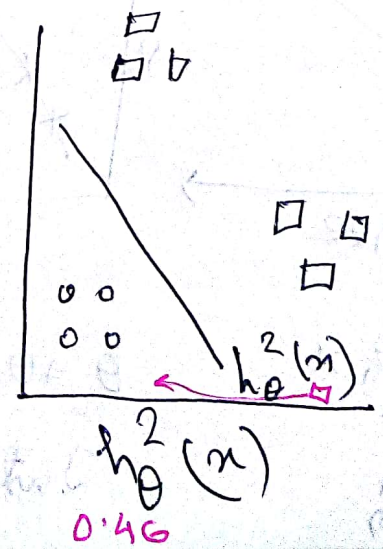
class 2: □

class 3: x

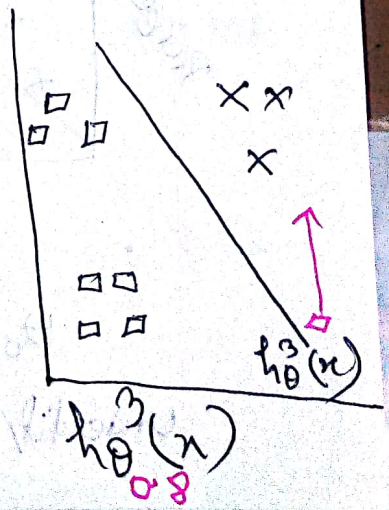
Class 1:



Class 2:



Class 3:



$h_\theta^{(i)}(x) = P(y = i | x, \theta) \quad (i = 1, 2, 3)$

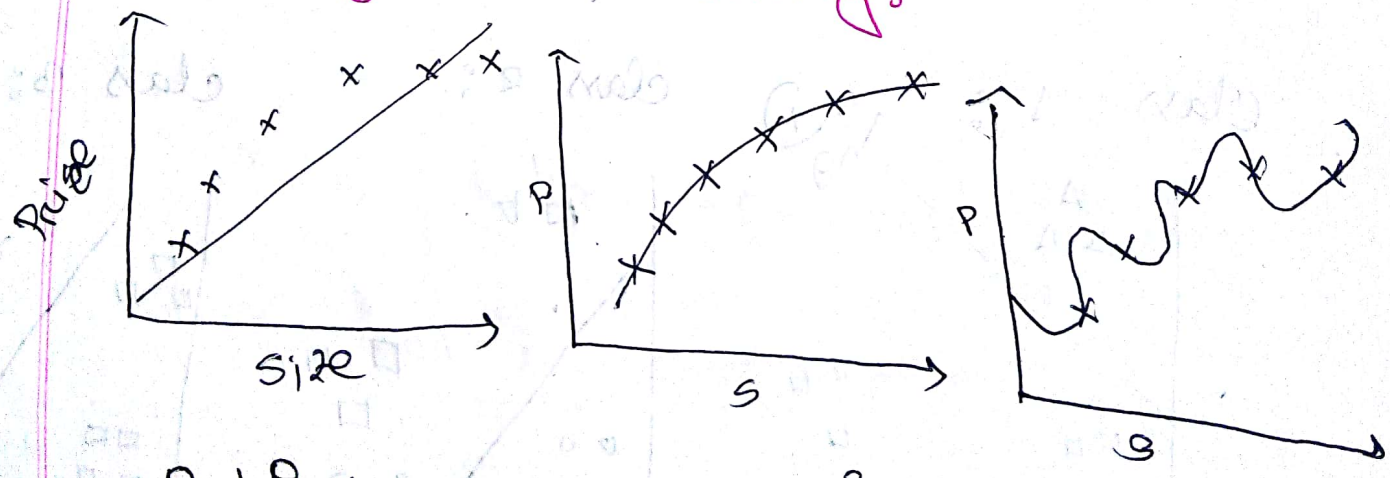
→ Train Logistic Regression classifier, $h_\theta^{(i)}(x)$ for each class i to predict the probability that $y = i$.

→ One new input x, to make a prediction pick the class i^{th} maximize

$$\max_i \; h_\theta^{(i)}(x)$$

—x—

Overfiting Vs Underfiting:



$\theta_0 + \theta_1 x$

Underfit/high bias

$\theta_0 + \theta_1 x + \theta_2 x^2$

Just fit

$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

Overfit/High variance

* Underfit:

Structure not captured by model.

$h_\theta(x)$ maps poorly.

* Overfit:

Fits training data set very well ($J(\theta) \approx 0$),
but fails to generalize to new input

Addressing overfiting problem:

1. Reduce the # of features
   - → manually select which feature to keep.
   - → Use a model selection algorithm

2. Regularization
   - → keep all features, but reduce the magnitude of θs.
   - → works well when we have a lot of slightly useful features.

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 + 1000\,\theta_3^2 + 1000\,\theta_4^2$$

θ_3^2 ↓ tends to $\theta \to 0$; θ_4^2 ↓ 0

$x_0 \quad x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_n$

$\theta_0 \quad \theta_1 \quad \theta_2 \quad \theta_3 \quad \ldots \quad \theta_n$

} don't know which one to reduce.

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 + \lambda \sum_{j=1}^{n} \theta_j^2$$

; λ = regularization constant.

—x—

06.05.18.

Four pillers of Management:

1. Planning → Min. resource, max. output.
2. Organizing → Chain of command, work break down structure (WBDS)
3. Leading
4. Controlling → feedback, Conformance (As per specification)

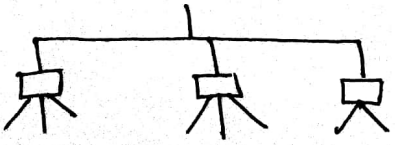
[1] Feasible point → Optimal Solution

Capacity → Fixed time duration

1. Design Capacity → Theoritical max. capacity.
2. Effective Capacity → Actual output < Design capacity.

Due to machine maintenance, fatigue etc.

[2] Project



Project
↓
Task
↓
Subtask
↓
Work package

Total Quality Management → TQM

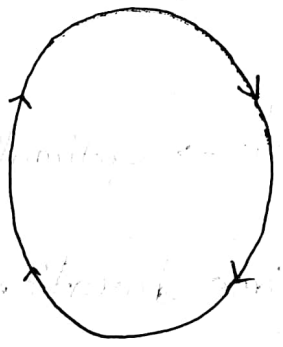
Tool of TQM

↳ PDCA cycle / Shewart cycle / Deming wheel

↳ Plan Do Check Act.

Continual process of betterment

BPR



→ Detect Detection

→ Planning for improvement

↳ small experimental change

large scale change for full facility.

Controlling

(A never ending process)

—x—

08.05.18.

Definition of Management:

Process of designing and maintaining an environment in which individuals, working together in groups.

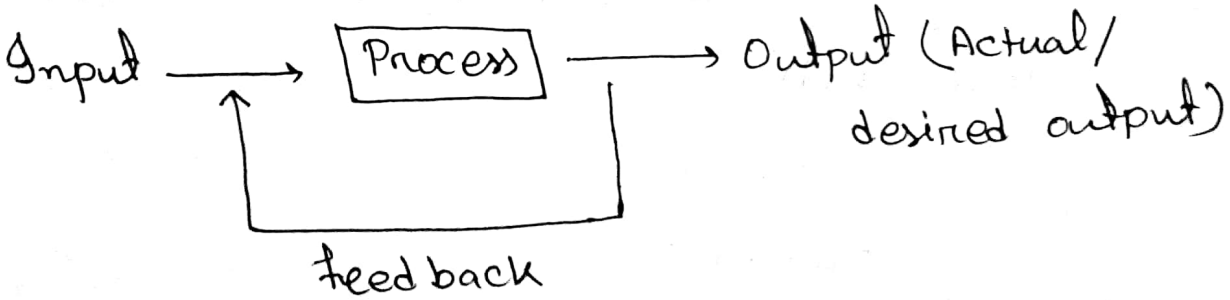
* Organizational level:

Top line

Middle line

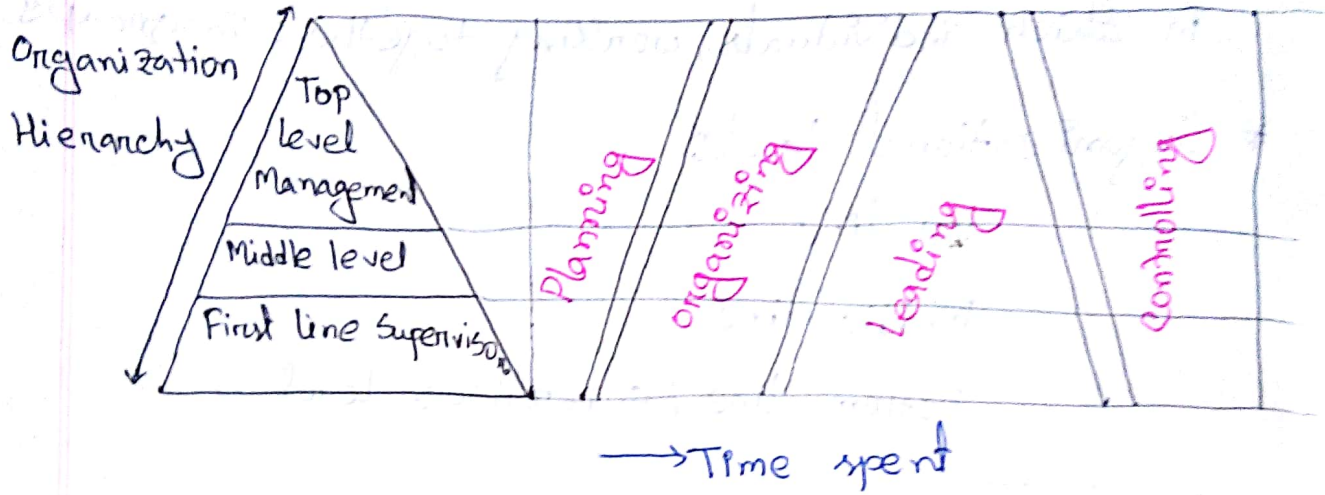
bottom line / Supervisory level

Staffing



The process will continue, Unless actual output = desired output

* Managerial Functions at Different Organizational levels:



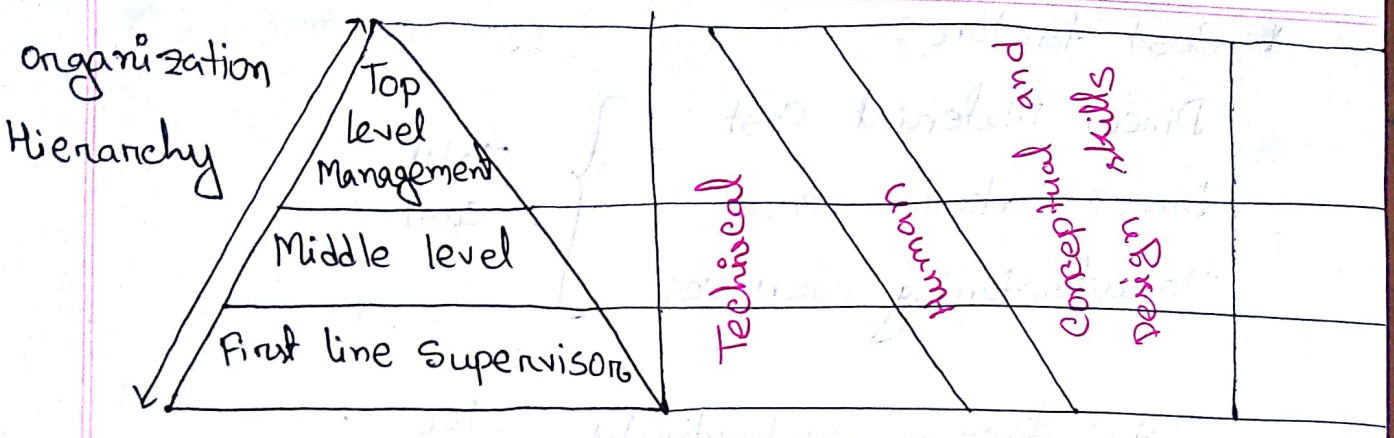
* Managerial Skills and organizational Hierarchy:
(Four skills)

Technical

Human

Conceptual

Design skills



* The goals of all managers and organizations:
  * Surplus of:

Business organization ⟶ Profit

Nonprofit " ⟶ Satisfaction of needs

* Key Terms:
  * Productivity

$$= \frac{\text{output}}{\text{input}}$$ (time period & quality is fixed)

(time/cost)

↓ Unit time ⟶ Unit cost

  * How to increase
  * Total-factor productivity
  * Partial productivity
  * Efficiency
  * Effectiveness.

* Cost factor:

Direct Material Cost
Direct Labor cost
Manufacturing over cost
} Total cost

Total factor productivity = $\frac{100}{\text{Total cost}}$

$100+50+150$

If output = 100 unit

DM cost = 100 TK

DL cost = 50 TK

MO Cost = 100 TK

* Productivity in terms of direct material

$= \frac{\text{output}}{\text{Direct material cost}} = \frac{100}{100}$

Efficiency = To do the things in the right way.

Effectiveness = To do right things.

Ref. Book → Management (a global perspective)

-Heinz weihrich, Harold Koontz.

09.05.18.

* Key Terms:

Capacity

Design capacity (Theoritically max. output)

Effective capacity (Actual output)

Efficiency

Utilization.

* Capacity:

Output of a system in a particular amount of time.

Design capacity > Effective capacity

↓

due to loss

$$\text{Efficiency} = \frac{\text{output}}{\text{Effective capacity}}$$

$$\text{Utilization} = \frac{\text{output}}{\text{Design capacity}}$$

* Efficiency > Utilization at constant output.

* Principle of Management:

1. Fredrick Taylor's Scientific management:

   Experience → Rules of thumb.

2. Henri Fayol's modern operational management theory:

   Father of modern management theory.

   14 principles → Search

   Division of work
   Authority and responsibility
   Unity of command
   Unity of Direction
   Subordination of Individual interest
   Remuneration
   The degree of centralization

Scalar chain

Order

Equity

Stability of Tenure of Personnel

Initiative

Esprit de Corps (Unit of employees).

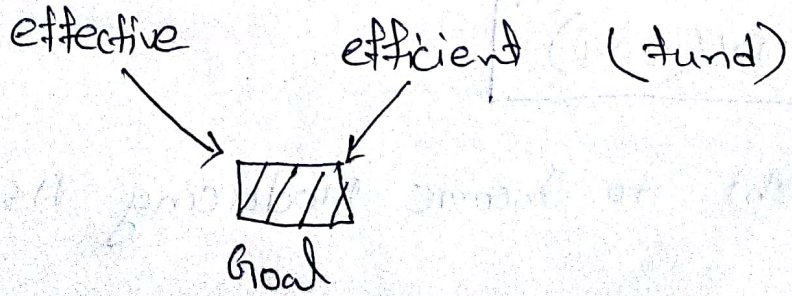
3. Elton Mayo and F.J. Roethlis bergen's Hawthorne studies:

—x—

13.05.18

* Financial Management:

Refers to the efficient and effective management of money (funds) in such a manner as to accomplish the objectives of the organization.



Planning
↑
fund

* 10 basic Principles of Financial Management:

→ Organize your finances.

organize — planning

what you have? — what you need to do?

→ Spend Less Than You Earn.

→ Put your money to work.

"Time value of money"

Future value (FV) = Present value (PV) $(1+i)^n$

n → Time

i → interest

$$FV = PV(1+i)^n$$

→ Limit Debt to Income Producing Assets.

→ Continuously Educate Yourself.

→ Understand Risk.

→ Diversification is Not Just for Investments.

→ Maximize Your Employment Benefits.

→ Pay Attention to Taxes.

→ Plan for the Unexpected.

——x——

CT-1

Next Sunday

Class Time

Till Here Syllabus

——x——

⇒ Ref. Book → Quality control and management.

—Dr. Ahsan Akter Nasir.

(TQM)

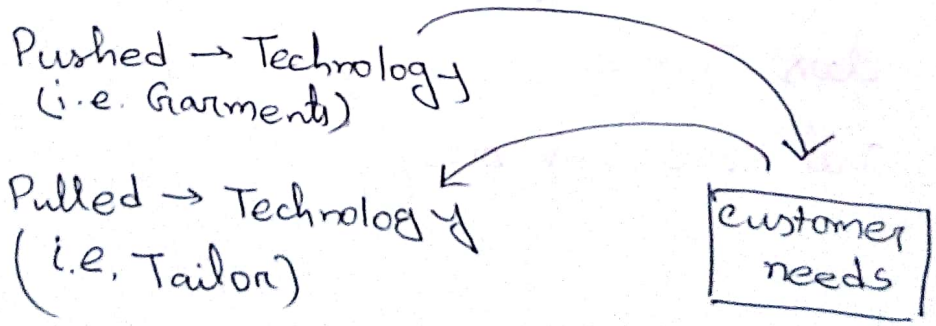
15.05.18.

Management of Innovation:

R&D
↳ Research & Development.

Common Tools:
Brainstorming
Virtual prototyping
⋮

* Innovation process may be either pushed or pulled process.



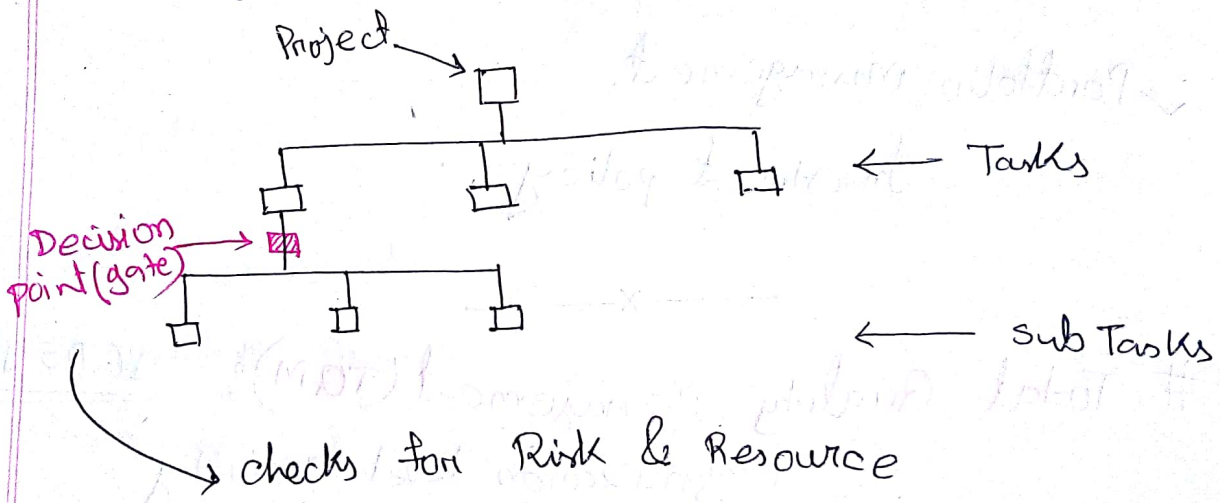
Common Tools:
✓Brainstorming
✓Virtual prototyping
✓Product lifecycle management.

Trade-off
- Competition → ↓ Product lifecycle
- Quality → ↑ (")

✓ Idea management.

✓ TRIZ (Theory of the Resolution of Invention-related tasks)

✓ Phase gate model.



✓ Project Management.

Goal

Specific Time.

* Primary constraints → scope, time, quality and budget.

✓ Product line planning.

Marketing
↓
Product
↓
Product line

{ Minimize marketing risk.
Maximize difference among items.

✓ Portfolio management.

Investment policy

———x———

Total Quality Management (TQM): 16.05.18

Organization level activity.

Quality ⟶ Customers Satisfaction.

* TQM in different stages in product development.

Quality Function Deployment ⟶ QFD.

Production Activity Control → PAC

Shop floor Control → SFC

Characteristics of TQM:

* Continual improvement:
* Customer focus:
* Organization-wide activity.
* Employee empowerment.

Power dedication.

gives birth to Quality circle (QC).

↓

Volunteer work

* Team approach.
* Competitive benchmarking.
* Knowledge of tools.
* Internal and external customers.
* Long-term relationship with the suppliers.

—x—

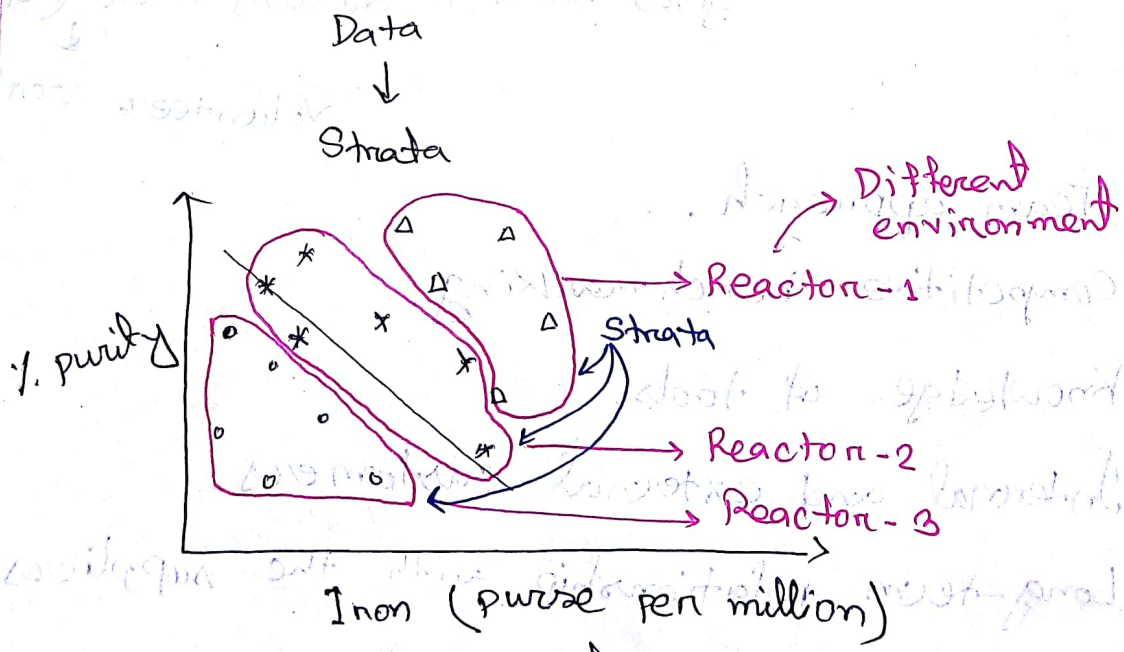
20.05.18.

Tools of TQM:

1. Check sheet.

Type	Tally	Frequency
Distortion	HHT II	7
Crack	HHT	5
Spatter	III	3

2. Stratification analysis.

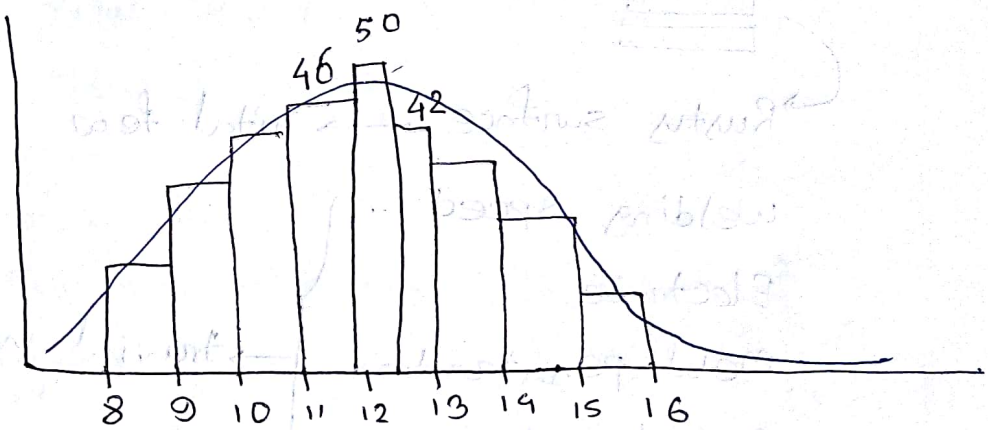


3. Histogram.

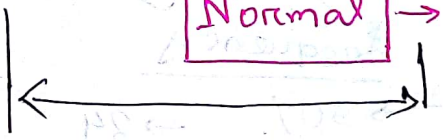
Frequency Distribution → Graphical view.

12 cm → Desired value

9-15 cm → tolerance



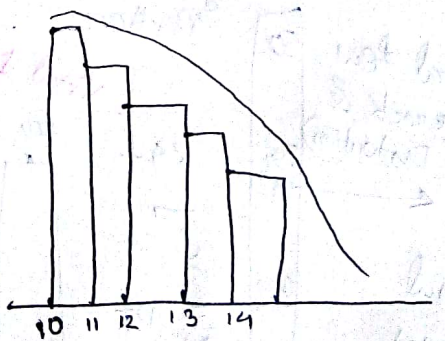
Normal → Distribution.



Acceptance Range

① Normal Distribution

② Skewed ,,

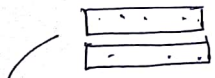


Chapter-3
page-39

4. Pareto Analysis.

Vital few
Trivial many

Welding



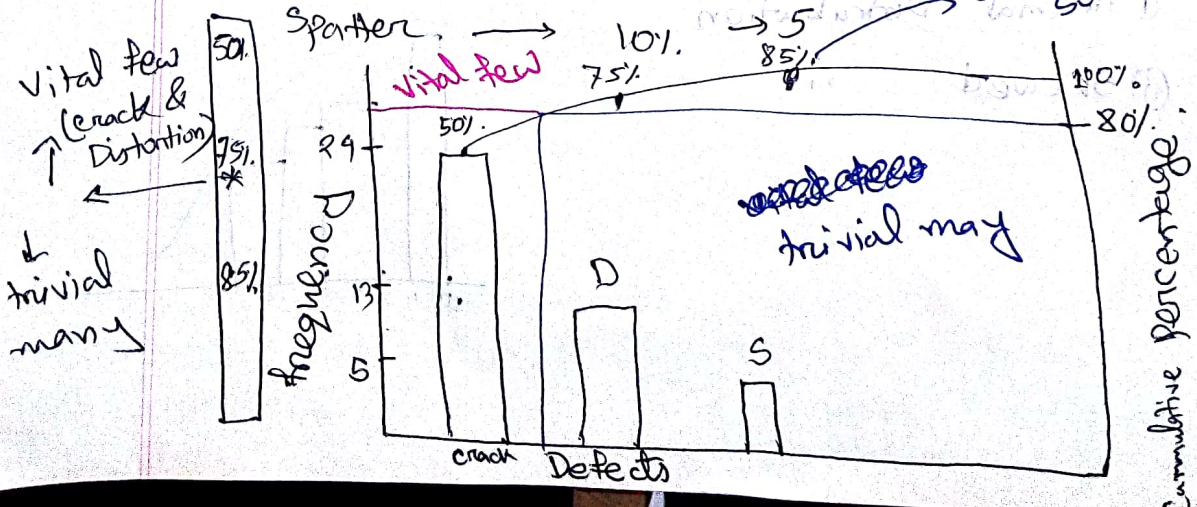
→ Rusty surface → vital few

- Welding speed
- Electrode
- Metal penatration
- Arc length.

→ trivial many

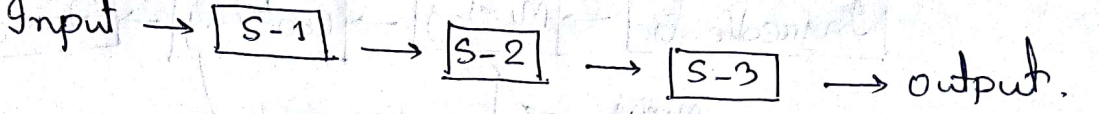
	Frequency	$\frac{24}{\text{Total sum of frequency}} \times 100\%$
Crack → 50%.	→ 24	
Distortion → 25%.	→ 13	
Spatter → 10%.	→ 5	

Cummulative percentage sum.

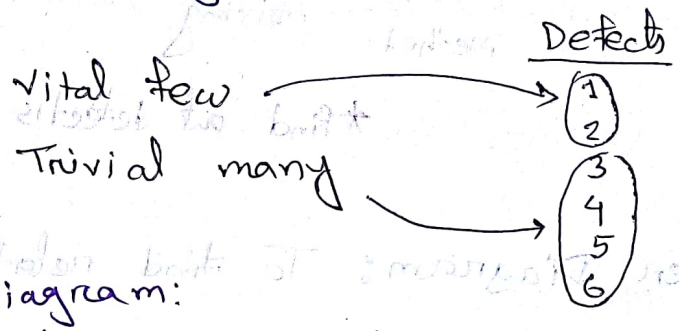


22.05.18.

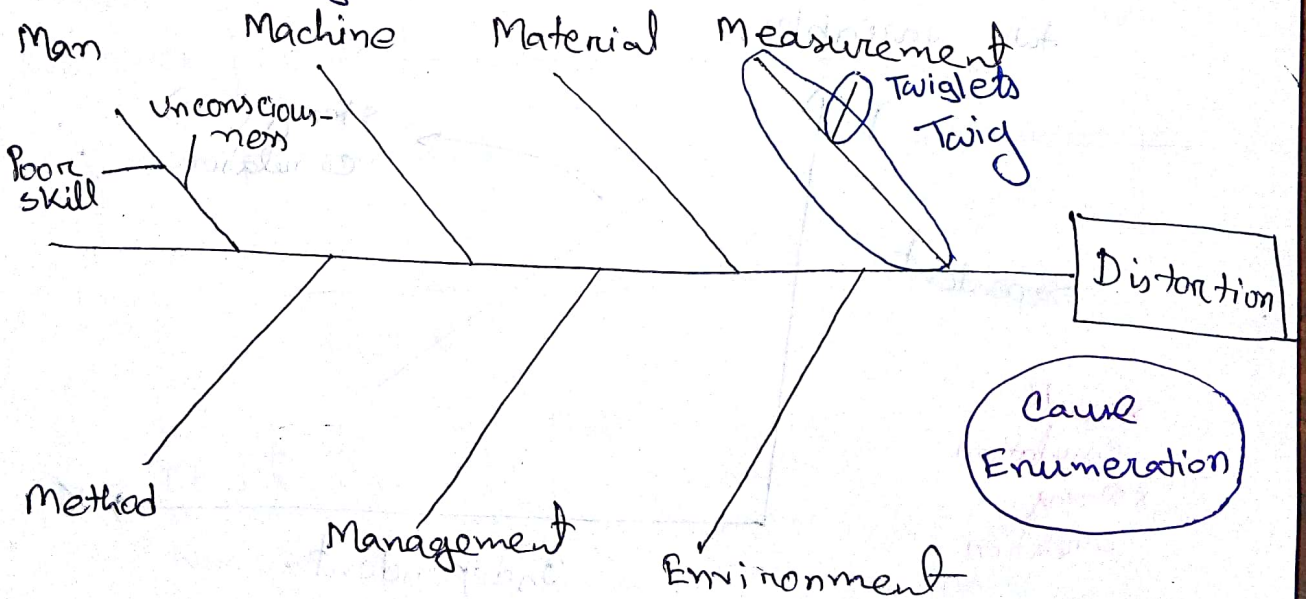
5. Process flow chart:



6. Cause-effect Diagram:



Fishbone Diagram:



Check sheet → Frequency

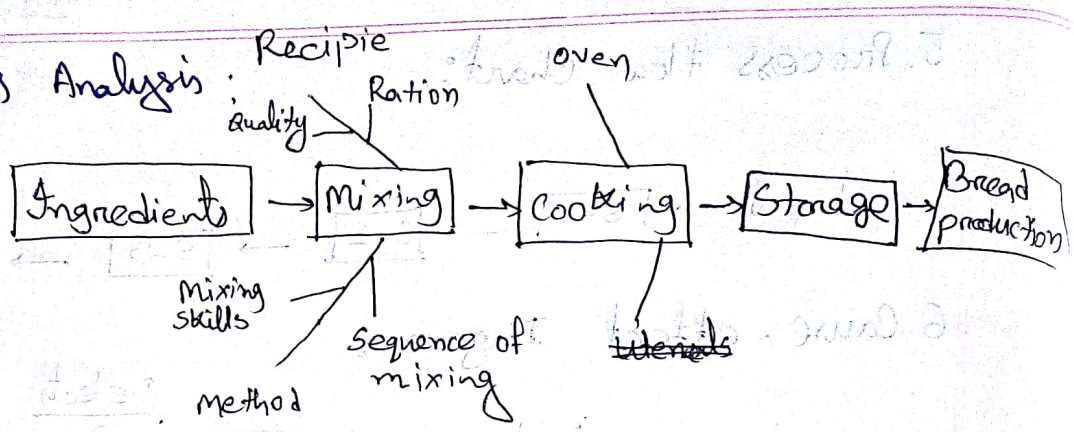
↓

Pareto Analysis → vital few

↓

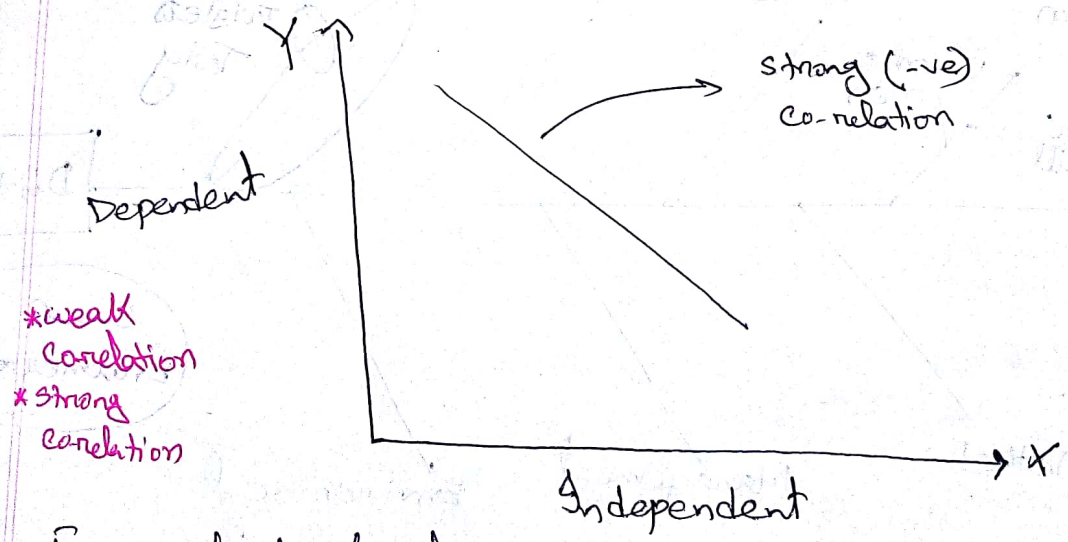
Cause-effect Diagram → Reason/cause finding

Process Analysis:



*find out defect is exact location.

7. Scatter Diagram: To find relation between two variables.



*weak correlation
*strong correlation

From check sheet:

days (x)	Defect (Y)
1	20
2	15
3	18
4	19

temparature(x)	Defect(Y)

8. Control Chart: To decide the process in control/not.

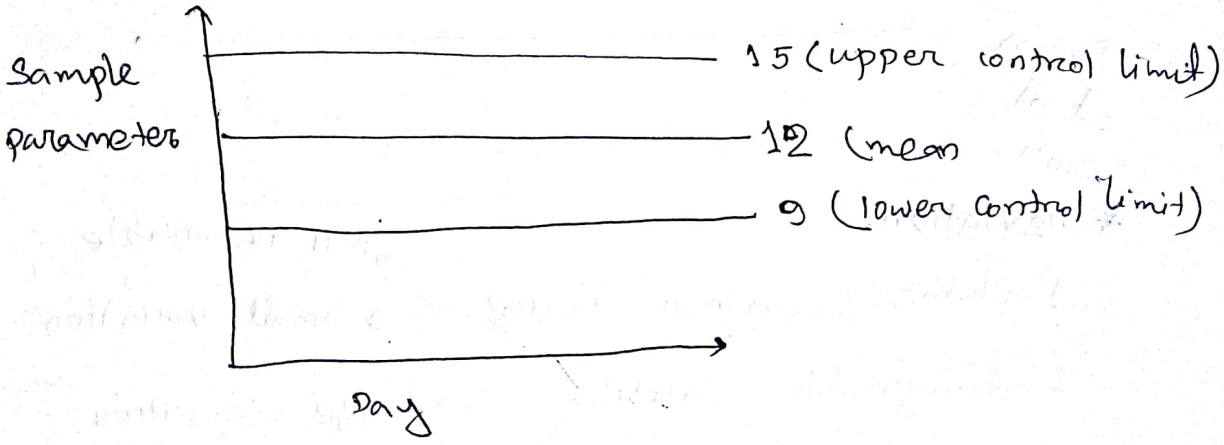
expected rate → targed value (12 cm)

Tolerance → $\pm 3\sigma$ (→ 1)

Acceptable range → $\mu \pm 3\sigma$

$= 12 \pm (3 \times 1)$

$= 9 - 15$



Variable type

1 → 1.1 → 1.9 → 2 → length

Attribute type.

Yes/No

(no fraction type).

———X———

8. Control Chart: To decide the process in control/not.

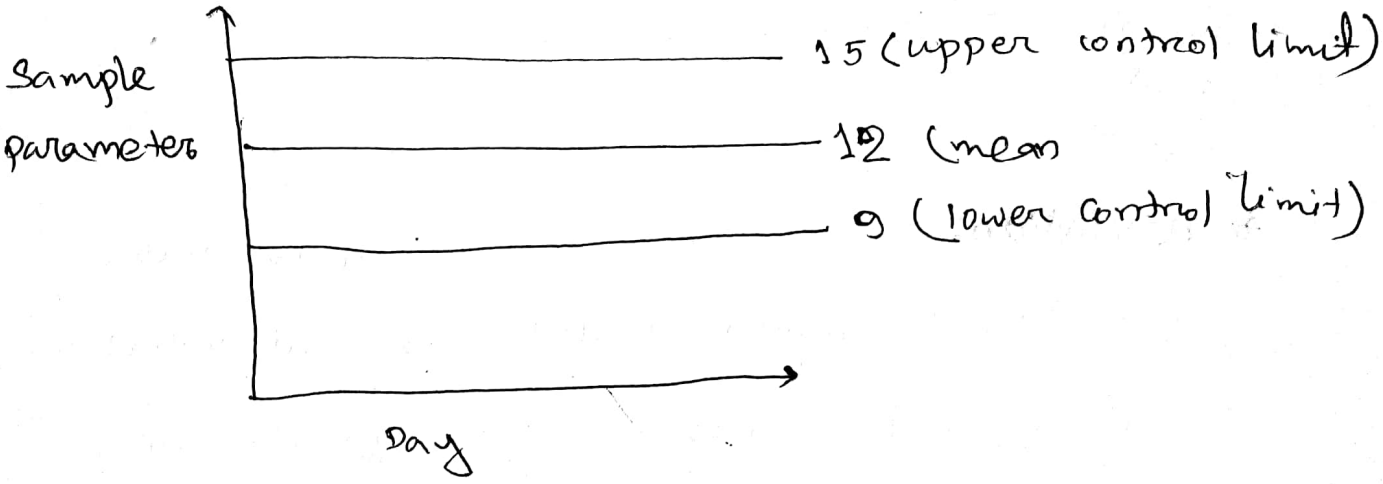
expected rate → target value (12 cm)

Tolerance → $\pm 3\sigma$ ($\sigma \rightarrow 1$)

Acceptable range → $\mu \pm 3\sigma$

$= 12 \pm (3 \times 1)$

$= 9 - 15$



length

1 → 1·1 → 1·9 → 2 →

Variable type

Yes/NO

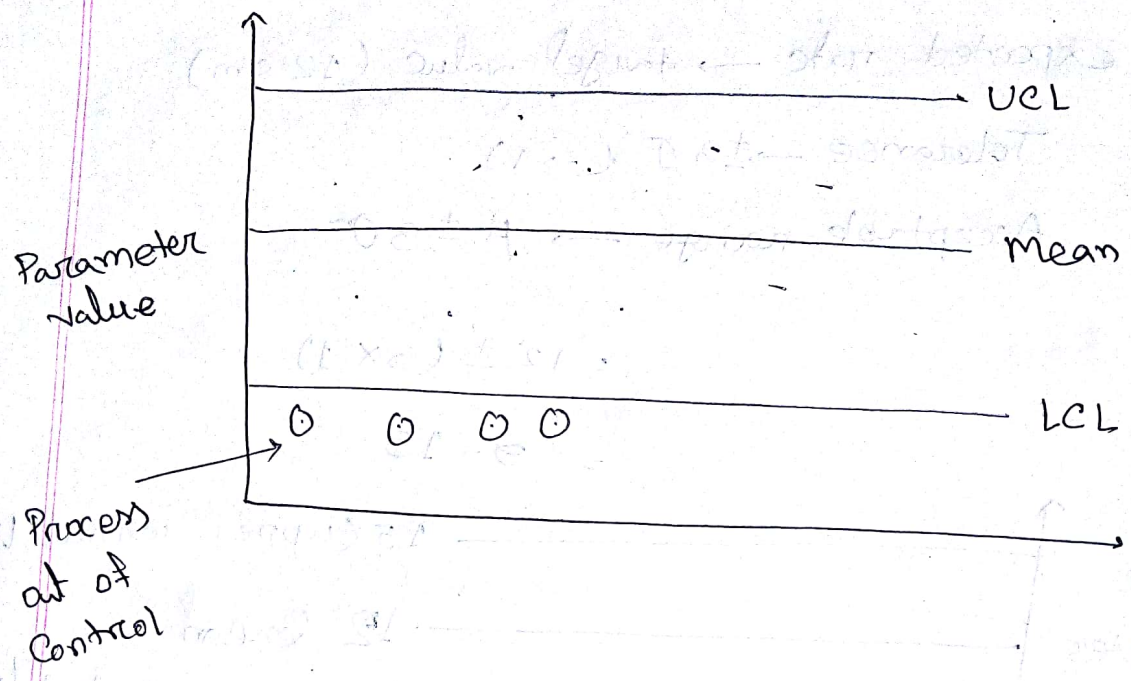
Attribute type

(no fraction type).

—X—

24.05.18.

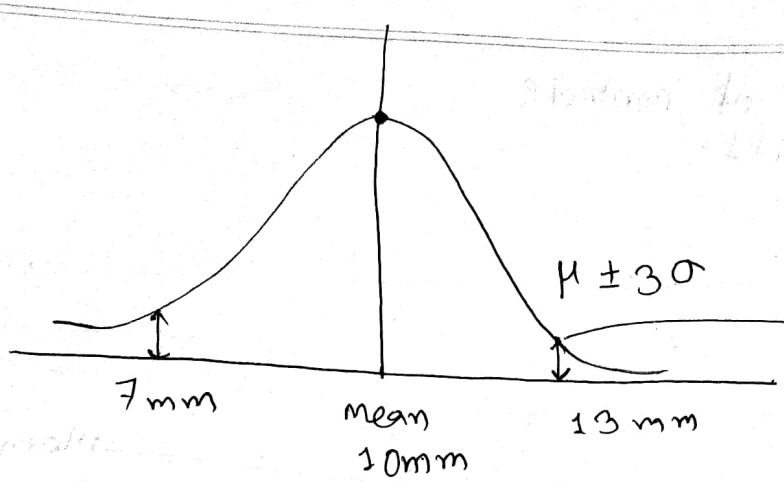
Control Chart:

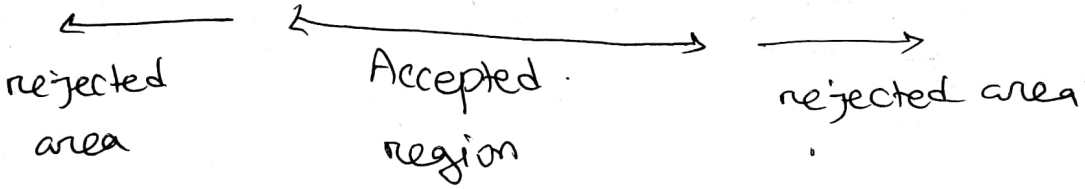


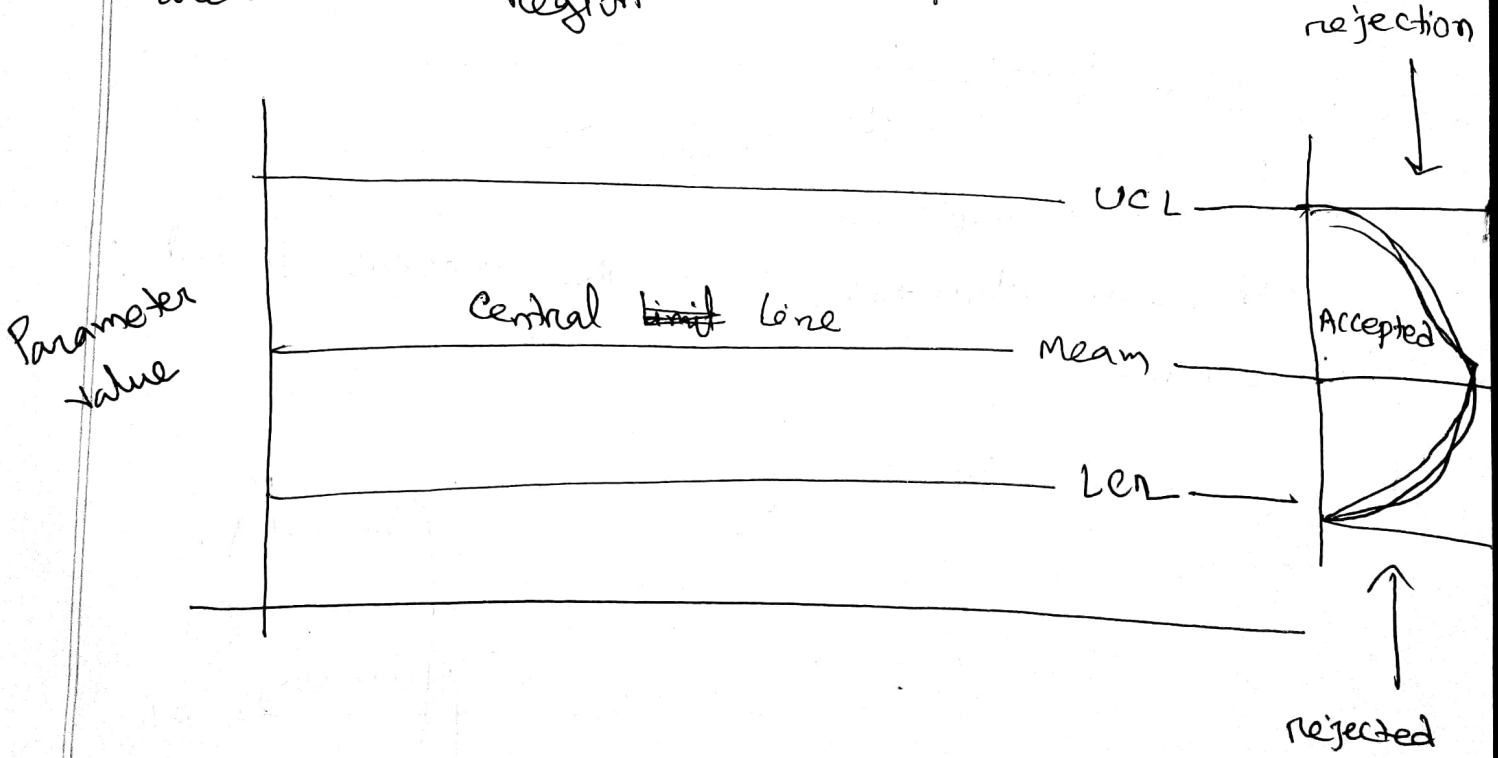
* Deviation:

1. Chance / Common Cause → Not removable
   → Small variation
2. Assignable cause → large deviation
   → Too much → Process out of control.

* Normal distribution for discrete value

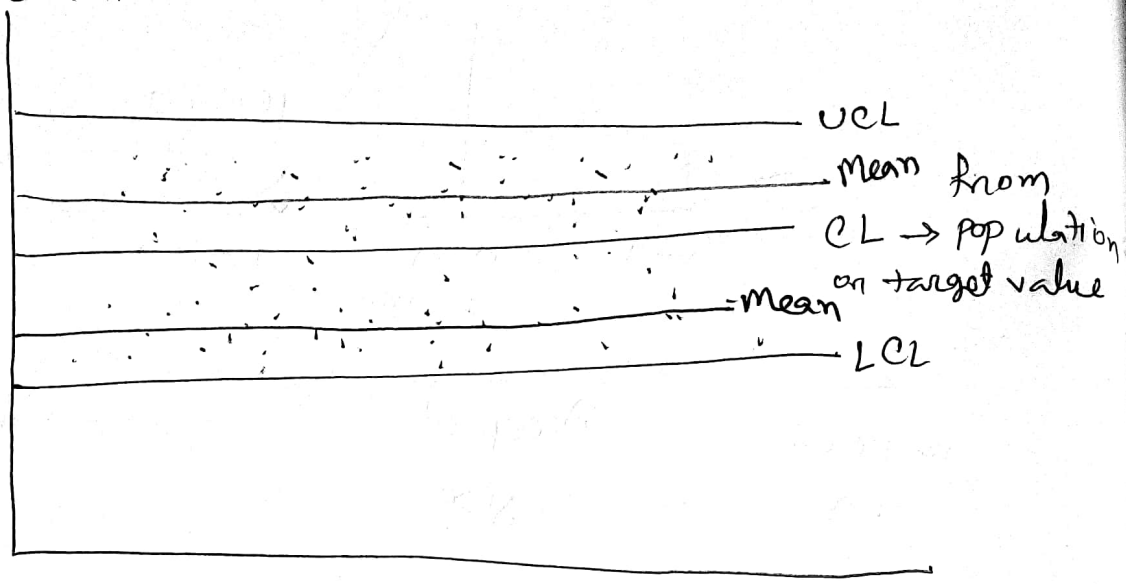






* Process out of control:

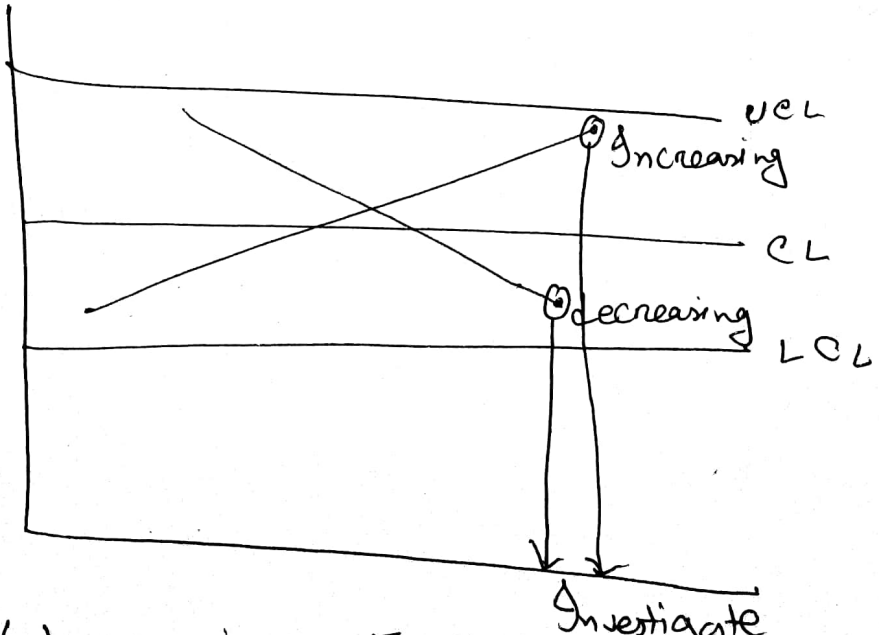
(1) mean shift



↙
5 times each → sample size

*If mean changes, then the process is out of control.

(2)



(2) Increasing / decreasing Trend

* Alarming situation, needed to be investigated.

* Decision:

1. Mean Shift ?
2. Increasing / Decreasing rate.
3. Points which are out of UCL & LCL.
4. Points which are very close to UCL & LCL.
5. If there are successive points out of UCL & LCL, then the process is out of control; else need to be investigated..

P-Chart : (P = Probability).

Attribute type ⟶ yes or no ans.

$$P = \frac{\text{defective Unit}}{\text{Total Unit}}$$

$$\text{Mean} = P = \left(\frac{\text{No. of defective unit, d}}{n} = \frac{np}{n} = P \right)$$

$$\text{variance} = np(1-P)$$

$$\text{Standard deviation} = \sqrt{\frac{\text{variance}}{n^2}}$$

n = sample size.

Standard deviation $= \sqrt{\frac{nP(1-P)}{n^2}}$

$= \sqrt{\frac{P(1-P)}{n}}$

* Sampling distribution of sample mean

variance $= \frac{nP(1-P)}{n^2}$

Standard deviation $= \sqrt{\text{variance}}$

$= \sqrt{\frac{nP(1-P)}{n^2}}$

$= \sqrt{\frac{P(1-P)}{n}}$

* General formula:

$UCL = E(a) + k\sqrt{\text{variance}(a)}$; k = factor = 3

$CL = E(a)$

$LCL = E(a) - k\sqrt{\text{variance}(a)}$

σ = Standard deviation $= \sqrt{\text{variance}(a)}$

(30)

* For P-chart:

$$UCL = P + 3\sqrt{\frac{P(1-P)}{n}}$$

$$CL = P$$

$$LCL = P - 3\sqrt{\frac{P(1-P)}{n}}$$

Chapter - 10

page → 160 - 162

——×——

29.05.18

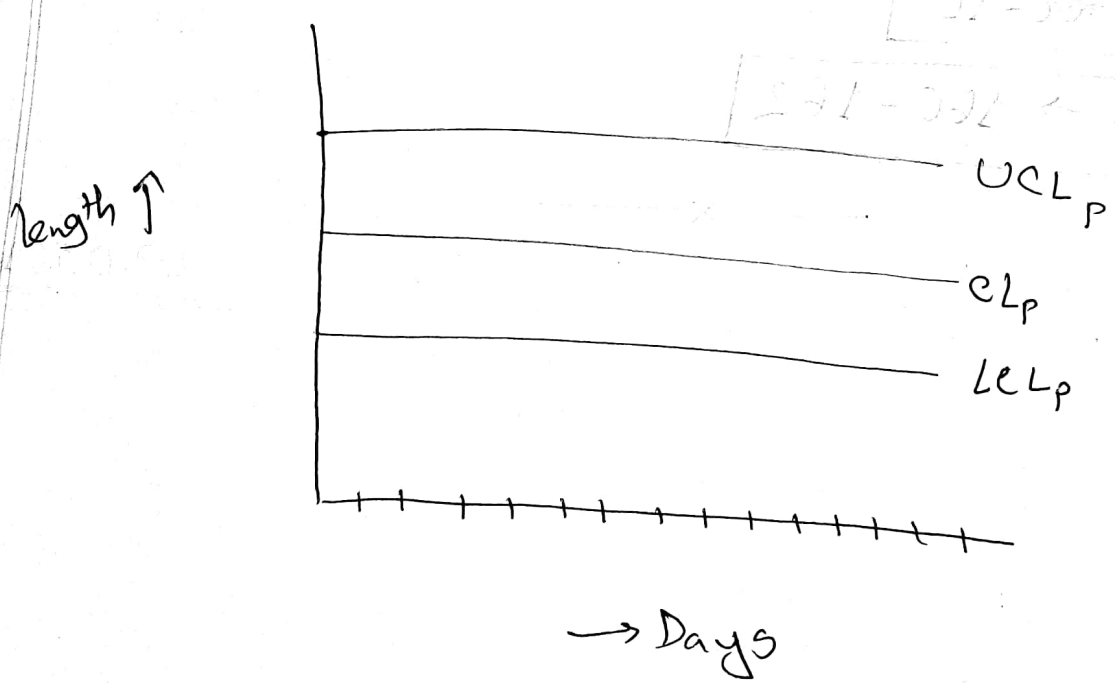
$$UCL_p = P + 3\sqrt{\frac{P(1-P)}{n}}$$

$P \rightarrow$ Fraction nonconforming (population)

$n \rightarrow$ Sample size (given)

$$CL_p = P$$

$$LCL_p = P - 3\sqrt{\frac{P(1-P)}{n}}$$



If P is not given,

$$\bar{P} = \frac{\sum_{i=1}^{m} d_i}{mn}$$

d = defective units in the ith day

$(\bar{P} \rightarrow P)$

m = No. of days

n = sample size

Problem: (Attribute type → yes/no decision)

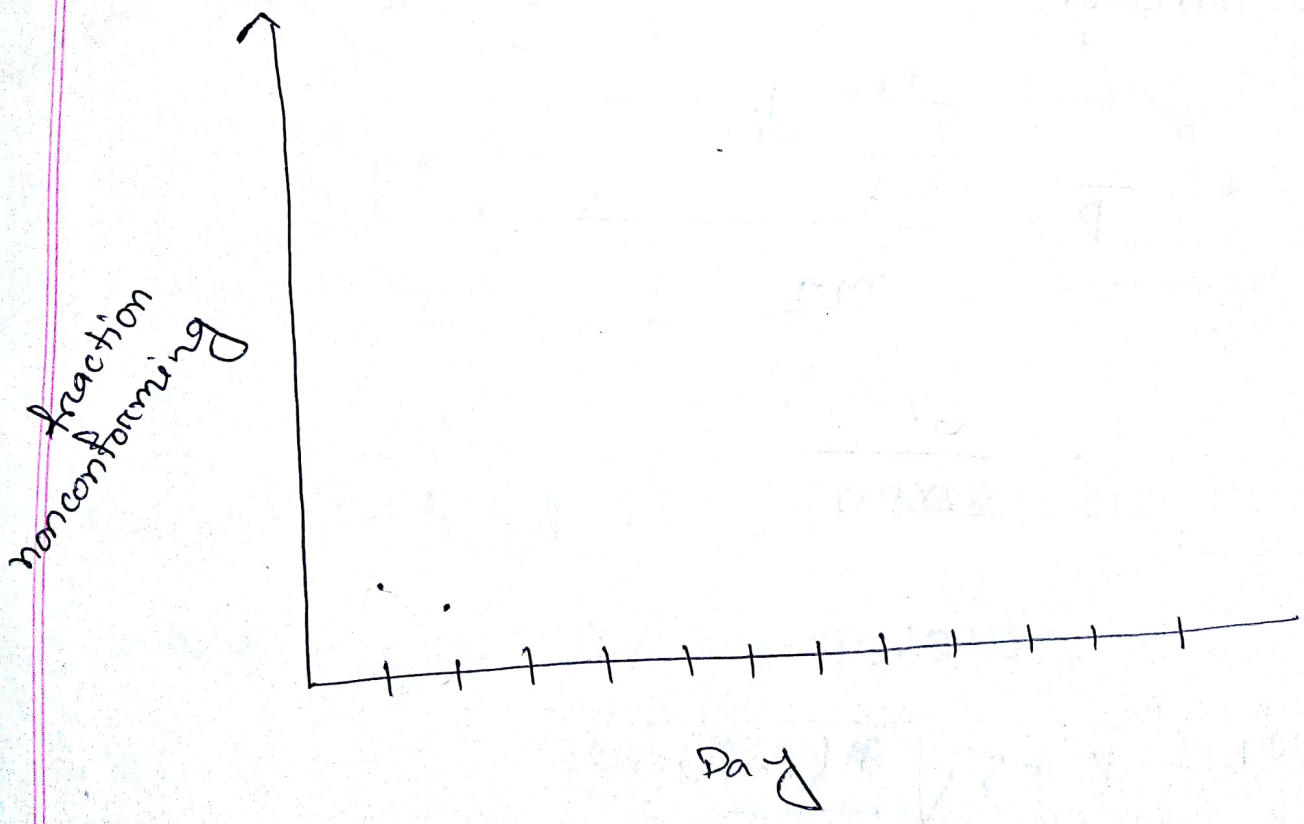
$n = 50$

$m = 22$

Fraction nonconforming (point plotting) $= \frac{\text{No. of failure at a particular day.}}{n}$

$= \frac{3}{50}$ (for day 1)

$= \frac{2}{50}$ (for day 2)



Steps:

1. Find Fraction nonconforming
2. Find $\bar{p}$
3. Find UCL
   CL
   LCL
4. Graph plot
5. Decision
6. Discard

$$\bar{p} = \frac{\sum_{i=1}^{m} di}{mn}$$

$$= \frac{67}{22 \times 50}$$

$$= 0.0609$$

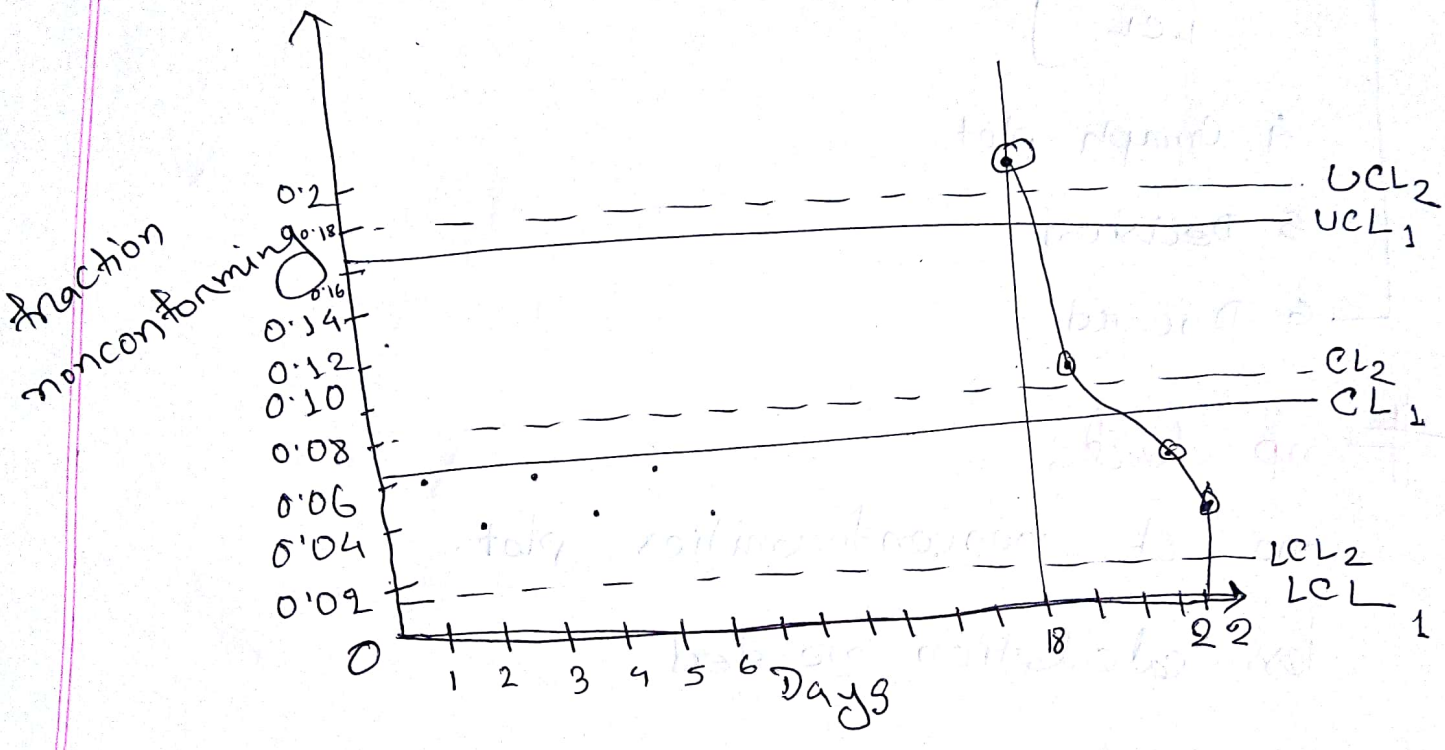
$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$= 0.1624$$

$$CL = \bar{P} = 0.0609$$

$$LCL = \bar{P} - 3\sqrt{\frac{\bar{P}(1-\bar{P})}{n}}$$

$$= -0.04 \text{ (consider as } 0)$$



Discard points out of UCL and LCL.

∴ Now,

$$\bar{P} = \frac{(67-9)}{(22-1)\times 50}$$

*Discard until the process is in control.

———×———

31.05.18.

* Steps:

1. Fractional nonconforming find
2. $\bar{p}$ find
3. UCL, CL, LCL } find
4. Graph plot
5. Decision
6. Discard

np chart:

no. of nonconformities plot.

less calculation needed.

C-chart:

When sample forms an inspection unit.

Poission distribution

Mean
Variance
Standard deviation.

Expected / mean → $\lambda = c = np$; n = sample size, p = fraction nonconformities

variance → $c = np$

Standard deviation → $\sqrt{c} = \sqrt{np}$ = $\sqrt{\text{variance}}$

$UCL = E(a) + k\sqrt{\text{variance}(a)}$

$= c + 3\sqrt{c}$

$CL = E(a) = c$

$LCL = c - 3\sqrt{c}$

→ c → population's nonconformities

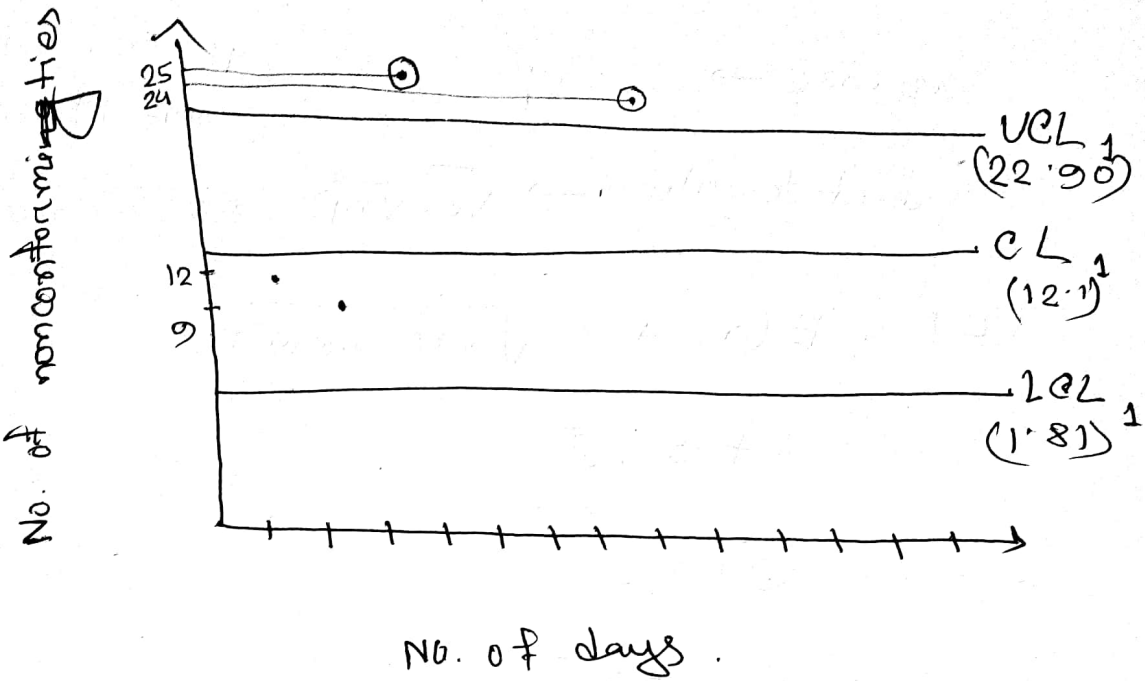
$\bar{c} = \frac{\text{Total defect}}{\text{Total day}}$

$\bar{c} \rightarrow c$

* Algorithm:

1. $\bar{c}$ find
2. UCL, CL, LCL find
3. Graph plot
4. Decision
5. Discard

Problem:



Discard 3rd day & 17th day

↓ 25 ↓ 24

(new) $\bar{c} = \frac{272 - 25 - 24}{22 - 2}$

$= 11.15$

(new) UCL, CL, LCL find & repeat algorithm.

—x—

Problem:

* U - chart: (Full Inspection)

$$\bar{u} = \frac{\text{Total No. of defective units}}{\text{Total sample size}}$$

$$= \frac{72}{635}$$

$$= 0.1134$$

Mean $\rightarrow \bar{u}$

Variance $= \frac{\bar{u}}{n}$

Standard deviation $= \sqrt{\frac{\bar{u}}{n}}$

$$UCL \longrightarrow \bar{u} + 3\sqrt{\frac{\bar{u}}{n}}$$

$$CL \longrightarrow \bar{u}$$

$$LCL \longrightarrow \bar{u} - 3\sqrt{\frac{\bar{u}}{n}}$$

——x——

03.06.18

U-chart:

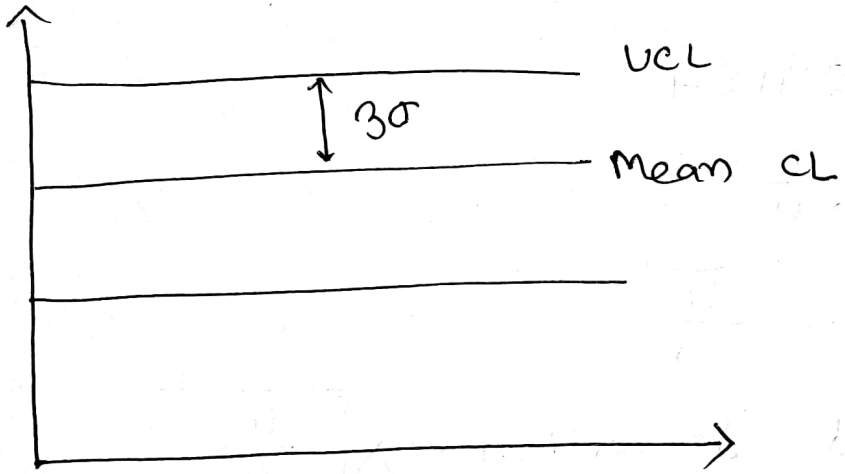
$$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n_i}}$$

$\bar{u}$ = mean

$$CL = \bar{u}$$

$$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n_i}}$$

$\sqrt{\frac{\bar{u}}{n_i}}$ = Standard deviation.



UCL = Mean + 3x Standard deviation.

$$\bar{u} = \frac{\text{Total defective Units}}{\text{Total Sample}}$$

$$= \frac{72}{635} = 0.1134$$

$n_i \rightarrow n$ varies

$25 \rightarrow n_{25} = n_1$

$30 \rightarrow n_{30} = n_2$

$40 \rightarrow n_{40} = n_3$

$n_1 = 25$

$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{25}}$

$CL = \bar{u}$

$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{25}}$

$n_2 = 30$

$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{30}}$

$CL = \bar{u}$

$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{30}}$

$n_3 = 40$

$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{40}}$

$CL = \bar{u}$

$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{40}}$

5^{th} day $\rightarrow 40 \neq t^2$

(4 Decision)

↳ Describe

Control Chart for variables:

* 2 parameters are needed to express any dataset.

2 parameters:

i) Central tendancy. → $\bar{x}$ (sample mean) or μ (population mean)

ii) Dispersion on variability. → R | S | σ

R → Range

S → Sample standard deviation

σ → population standard deviation

2
4
6
8

→ Mean = 5

Range = (2-8) variability

4
5
6
5

→ Mean = 5

Range (4-6) variability

* Control charts:

1. $\bar{X}$ - R chart
2. $\bar{x}$ - s chart
3. Cusum chart (cumulative sum)

m = no. of days

n = sample size

Sampling distribution of sample mean

↓

Central limit theorem follows

$$\mu \approx \bar{\bar{x}}$$

1st day(x) → 10 10·2 10·5 11 11·2

$$\bar{x} = \sum_{i=1}^{n=5} x_i/5 \quad \text{or} \quad \sum_{i=1}^{n=5} x_i/n$$

$$\bar{\bar{x}} = \sum_{j=1}^{m} \bar{x}_j/m$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

n = sample size

$\sigma_{\bar{x}}$ → Sample Standard deviation

σ → Population Standard deviation

"Chapter-1" 07.05.18

Data & Service:

Confidentiality

Data Confidentiality

Privacy

Integrity

Data integrity

System integrity

Availability

Accountability

Active Attack

Passive Attack

Attack surfaces and attack trees:

Model for network Security.

Ref. Book → Cryptography and network security.
Principles & practice.
7th Edition
William Stallings.

"Chapter-3" 08.05.18

* Symmetric Cipher:
  1. Substitution cipher
     Transposition cipher
  2. Number of keys used
  3. The way in which the plaintext is processed.

* Cryptanalysis & Brute-Force Attack:
  Computationally secure | Processors speed
  Unconditionally secure

Monoalphabetic Substitution cipher.

* One Time Pad:
  Perfect security

```
  BAUST
+ CNAFL   ⟶ random password
-------
  E
```

```
  143          722
+ 689        - 689
-----        -----
  722          143
```

```
  A            D            B            D
+ C          - C          + B          - B
---          ---          ---          ---
  D            A            D            B
```

"Sessional"

09.05.18.

* Transposition ciphers:

fs[0] → fs[4]

—x—

* Steganography:

10.05.18

Characters marking

Invisible ink

Pin punctures

Typewriters correction ribbon.

—x—

"Chapters-4"

Diffusion

confusion.

* The data Encryption Standard.

Simplified DES. (S-DES).

* S-DES:

8-bit Plaintext
10-bit Password
} final

* Procedure: (Key generation)

Plaintext
password → 1 2 3 4 5 6 7 8 9 10
1 1 0 0 0 1 1 1 1 0

P10

P10
3 5 2 7 4 10 1 9 8 6

LS-1

1 2 3 4 5
0 0 1 1 0 | 1 2 3 4 5
0 1 1 1 1

1 2 3 4 5
0 1 1 0 0 | 6 7 8 9 10
1 1 1 1 0

P8

P8
6 3 7 4 8 5 10 9

1 1 1 0 1 0 0 1 → K_1

LS-2

1 2 3 4 5
1 0 0 0 1 | 6 7 8 9 10
1 1 0 1 1

P8 1 0 1 0 0 1 1 1 → K_2

Example:

1 2 3 4 5 6 7 8 9 10

1 0 1 0 0 0 0 0 1 0

P10 → 1 0 0 0 0 | 0 1 1 0 0

LS-1 → 1 0 0 0 1 | 1 1 0 0 0

P8 → 1 0 1 0 0 1 0 0 → K_1

LS-2 → 1 0 0 1 0 | 0 0 0 1 1

P8 → 0 1 0 0 0 0 1 1 → K_2

—x—

14.05.18.

Encryption Detail:

IP							
1	2	3	4	5	6	7	8
2	6	3	1	4	8	5	7

IP^{-1}							
1	2	3	4	5	6	7	8
4	1	3	5	7	2	8	6

1	2	3	4	5	6	7	8
A	B	C	D	E	F	G	H

IP → B F C A D H E G

IP^{-1} → A B C D E F G H.

E/P							
4	1	2	3	2	3	4	1

E/P → Expansion Permutation

IP → Initial Permutation.

1 2 3 4

1 0 1 0

E/P → 1 0 10 10 1 0

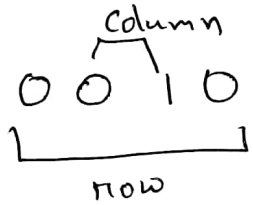
Substitution Box (S0, S1):

S0

	0	1	2	3
0	1	0	3	2
1	3	2	1	0
2	0	2	1	3
3	3	1	3	2

S1

	0	1	2	3
0	0	1	2	3
1	2	0	1	3
2	3	0	1	0
3	2	1	0	3

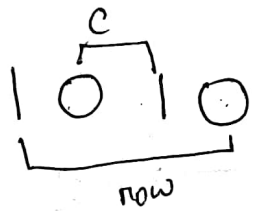


In = 0

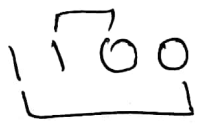
le = 1

00 → 0
01 → 1
10 → 2
11 → 3

0 → 0 0



2 → 1 0



1 → 0 1

P4
2 4 3 1

Next Lab
code
search

—— x ——

Ex: Key generation:

10-bit key: 0 1 1 1 0 1 0 0 0 1

↓

P10 → 1 0 1 0 1 | 1 0 0 0 1

LS-1 → 0 1 0 1 1 | 0 0 0 1 1

P8 → 0 0 0 1 0 1 1 1 → K_1

LS-2 → 0 1 1 0 1 | 0 1 1 0 0

P8 → 0 1 1 0 1 1 0 0 → K_2

Encryption:

8-bit plaintext → 1 1 0 1 0 1 0 1

IP → 1 1 0 1 | 1 1 0 0

01101001 ← E/P

⊕ 00010111 ← K1

0111 | 1110

S0 → 0 0 | 0 0 ← S1

0000 ← P4

⊕ 1101

1101

1100

11101011 ← E/P

⊕ 01101100 ← K2

1000 | 0111

S0 → 0 0 | 1 1 ← S1

0110 ← P4

⊕ 1100

1010

1 0 1 0 1 1 0 1

01110011 ← IP^{-1}

∴ Ciphertext → 01110011 (Ans.).

S-DES:

15.05.18

* In case of decryption, the total process will be the same, only if K_2 will be used for the first time, then the K_1 will be used.

DES:

64-bit → Plaintext

56-bit → Password

Key needed → 16, size → 48-bit.

16 times operation.

Left side will be encrypted 8 times

Right " " " " 8 " .

—— x ——

Decryption:

Ciphertext → 01110011

IP → 1010 | 1101

11101011 ← E/P

⊕ 01101100 ← K_2

10000111

S0 → 00 11 ← S1

0011

0110 ← P4

⊕ 1010

1100

01101001 ← E/P

⊕ 00010111 ← K_1

01111110

S0 → 00 00 ← S1

0000

0000 ← P4

⊕ 1101

1101 1100

11010101 ← IP^{-1}

Plaintext → 11010101 (Am).

"Sessional" 16.05.18.

S. initialize (key);

↓

```
Void initialize (string key)
{
    Keys_Generation();
}
```

↓

```
void Keys_Generation()
{ Permutation (P10, KEY);
```

↓

String LShift (string input, int n)

out = out + inp[array[0]-1]

= out + inp[1-1]

= out + inp[0] ✓

key ~~input~~ = 0 1 1 1 0 1 0 0 0 1

array → P10 → 3 5 2 7 4 10 1 9 8 6

array

10

!0!

0 1 1 1 0 → input

LS-1 0 1 1 0 0

output = 0 0 1 1 0

firstbit = 0

0 1 1 0

inp[array[0]-1] output = output + firstbit

inp[3-1]

inp[2]

0 1 1 0 0

0

11000

```
void DES_Encryption()
{  IP();
     ↓
   Function_F (LIP, RIP, 1);
        ↓
    E/P
```

```
in [ 6, 7, 8 ]
out [          ]
ip [ 3, 1, 2 ]
for (i=0; i<3; i++)
   { out[i] = in[ip[i]-1]
   }
```

Next Lab → input key, plaintext → output

(code)

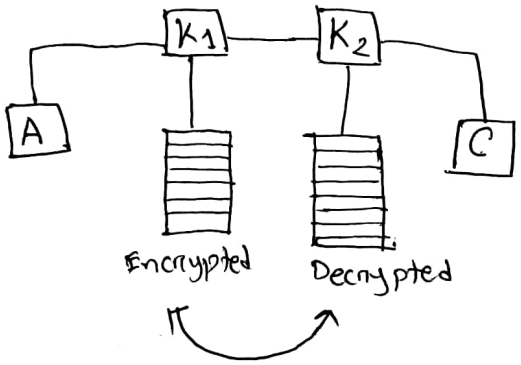
21.05.18

DES:

64-bit password is given and used all bits without 8, 16, 24, 32, 40, 48, 56, 64 th bit.

And key will be generated as 48 bits.

Max possible combination of ciphertext = 2^{56}



if match,
if can break

$2^{56} \times 2^{56}$

$= 2^{112}$; but actual combination needed is 2^{57}.

Meet in the Middle Attack

22.05.18.

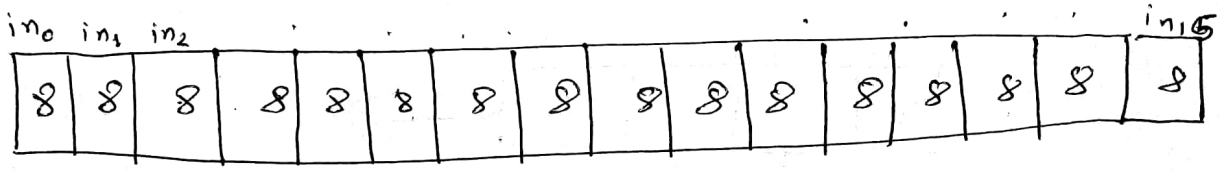
AES Structure:

Plaintext ⟶ 128 bit.

Password ⟶ 128 bit ⟶ 10 round

Output ⟶ 128 bit

Password ⟶ 44 no. of 32 bits key.



128 bits

in_0	in_4	.	.
in_1	.	.	.
in_2	.	.	.
in_3	.	.	in_{15}

Bytewise shift

—x—

"Sessional" 23.05.18.

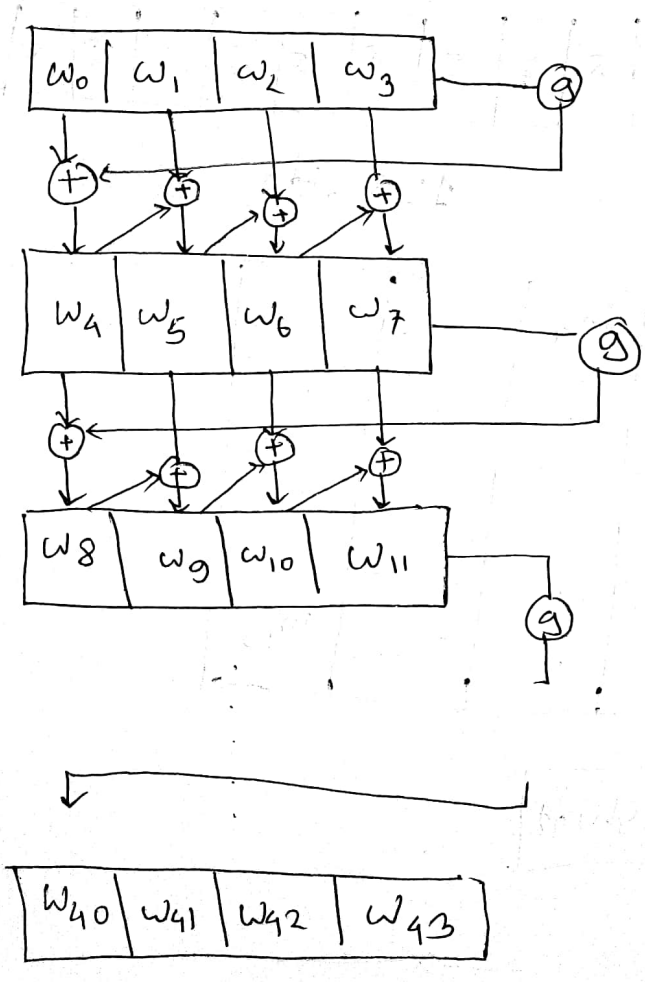
DES

Next Lab → AES

—x—

24.05.18

AES Key generation:



Chapter

RC → Round constant

Explanation of steps.

Chapter → ~~[redacted]~~ . 1, 3, 4.6.

—x—

"Chapter-7"

Table: 7.1

Electronic Codebook (ECB)
Cipher Block Chaining (CBC) → Solution

—x—

Syllabus for CT-1:

28.05.18

Chapter-1: 1.1, Definitions, 1.2, 1.3, 1.4, 1.6, 1.7, Network security model,

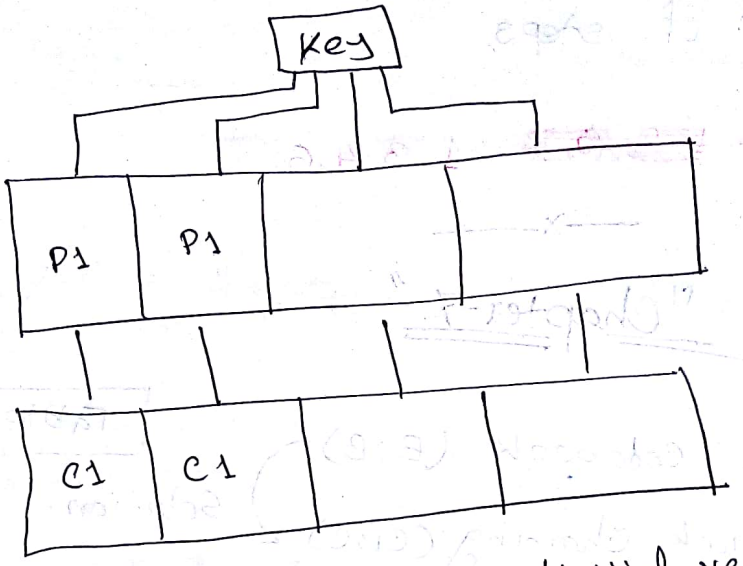
~~chapter 2~~

S-DES

↳ explanation, calculation

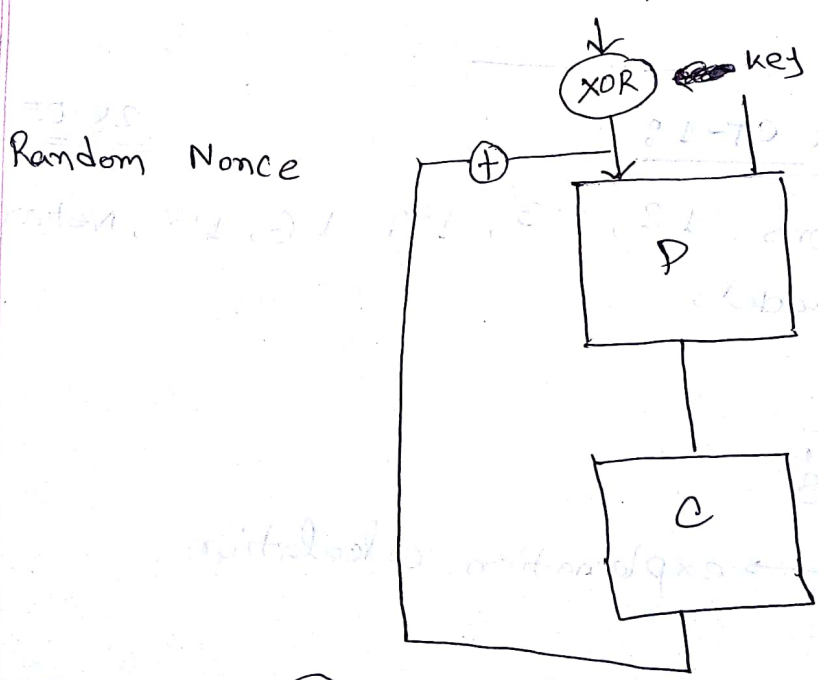
"Chapter-7"

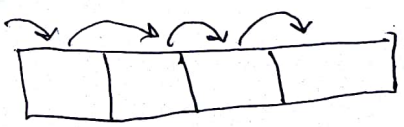
* Electronic Codebook (ECB):



256 bit

Nonce/Initial vector (IV)

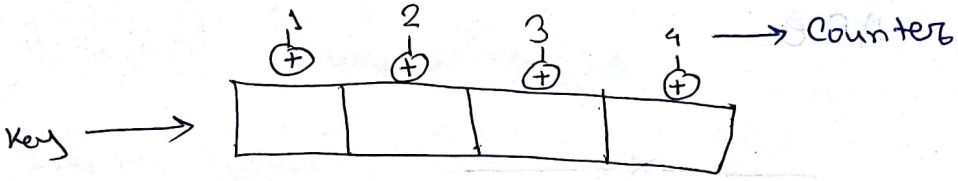




→ Cipher Block Chaining (CBC)

Cipher Feedback (CFB)

* Counter (CTR):

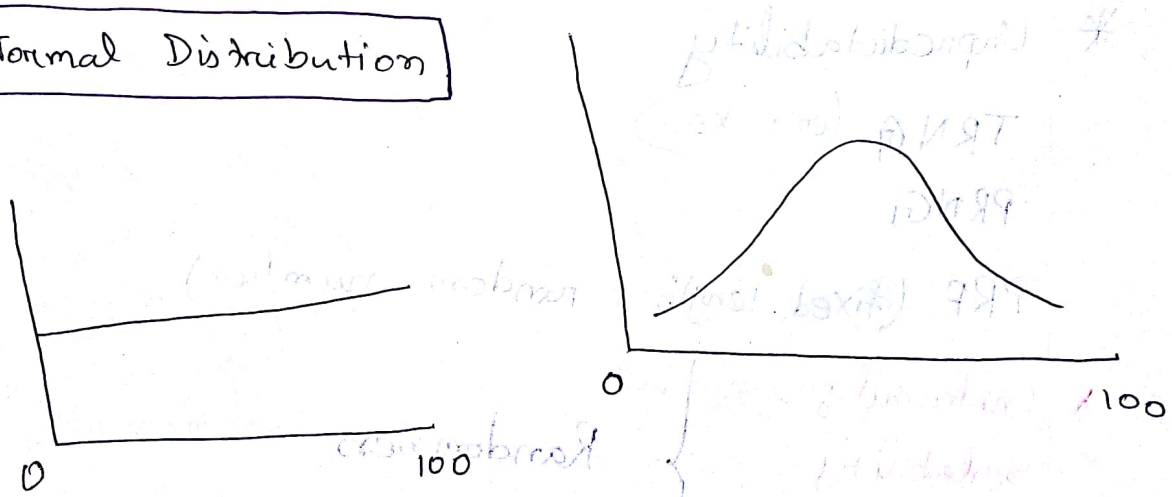


Format Preserving Encryption:

—x—

"Chapter-8"

Normal Distribution



random number generation

"Sessional" 30.05.18.

H.W. → plaintext → name

key → Roll

ABO

———×———

"Chapter-8" 31.05.18

* Randomness:
  - Uniform distribution
  - Independence.
* Unpredictability

TRNG (coin toss)

PRNG

PRF (fixed length's random number)

* Uniformity
* Scalability
* Consistency

} Randomness

Frequency Test

Run "

Maurer's universal statistical test.

* Unpredictability:

Forward unpredictability
Backward unpredictability

Stream cipher / Cryptography:

Key	Plaintext (PT)	ciphertext (CT)
0	0	(0+0) mod 2 = 0
1	0	(1+0) mod 2 = 1

(key + PT) mod m = CT

(CT - key) mod m = PT

Key = 1100
PT = 1010
⊕
0110

Encryption

0110
1100
1010

Decryption

Bit by Bit locker

—x—

04.06.18

Algorithm design:

$$X_{n+1} = (aX_n + c) \bmod m$$

10, 16, 18

$x_{10} = 10$

$x_{11} = 16$

$x_{12} = 18$

$x_{13} = ?$

$x_{10} = 10 = (ax_9 + c) \bmod m$

$16 = (a \cdot 10 + c) \bmod m$

$18 = (a \cdot 16 + c) \bmod m$

* Blum Blum Shub Generator:

$P \equiv q \equiv 3 \pmod 4$

relatively prime, S

$P \times q = n$

related prime

$x_0 = S^2 \bmod n$

for $i = 1$ to ∞

$x_i = (x_{i-1})^2 \bmod n$

$B_i = x_i \bmod 2$

i.e: 7, 11

$S = 19$

$n = 77$

$x_0 = (19)^2 \bmod 77$

=

$x_1 = (\quad)^2 \bmod 77$

=

$B_1 = (\quad) \bmod 2$

=

Using True random number generator, find prime S.

$x_0 = s^2 \bmod n$

for $i = 1$ to ∞

$x_i = (x_{i-1})^2 \bmod n$

$B_i = x_i \bmod 2$

i.e: 7, 11

$S = 19$

$n = 77$

$x_0 = (19)^2 \bmod 77$

=

$x_1 = (\quad)^2 \bmod 77$

=

$B_1 = (\quad) \bmod 2$

=

Using True random number generator, find related prime S,

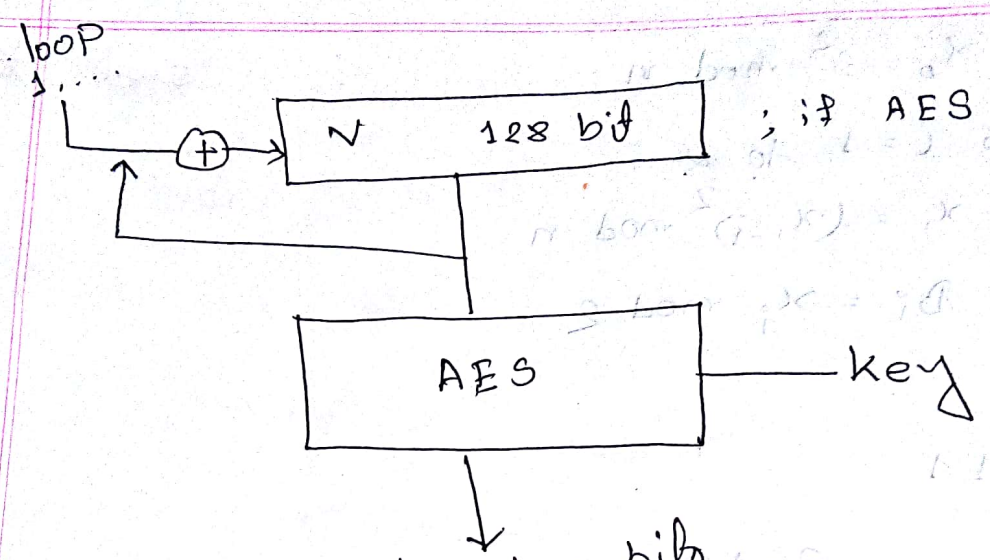


Fig: CTR mode

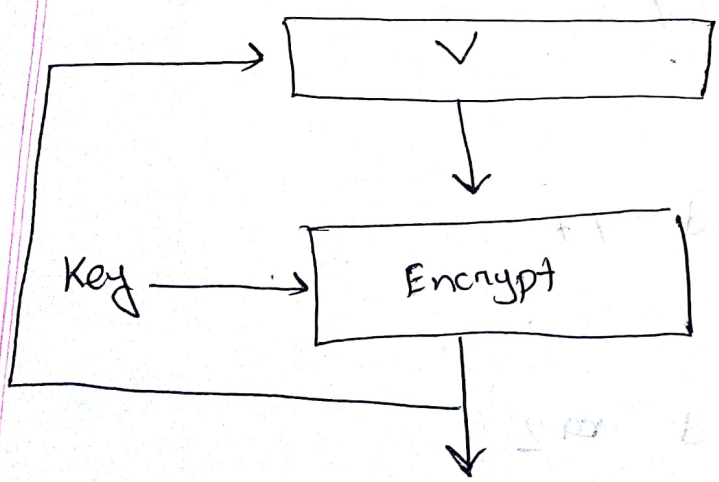


Fig: OFB mode

PRNG Mechanisms Based on Block Ciphers,

Stream Ciphers:

——— x ———

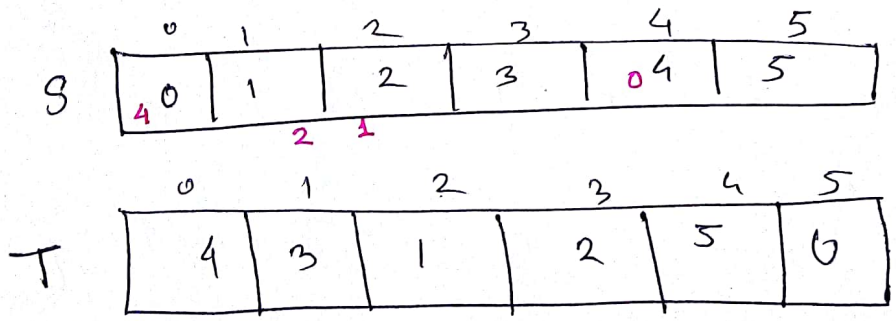
05.06.18

RC4: (Random bit generation)

bit Stream
// initialization
for i = 0 to 255 do

S[i] = i ;
T[i] = K[i mod keylen] ;

// T[i] = key



2nd Step:

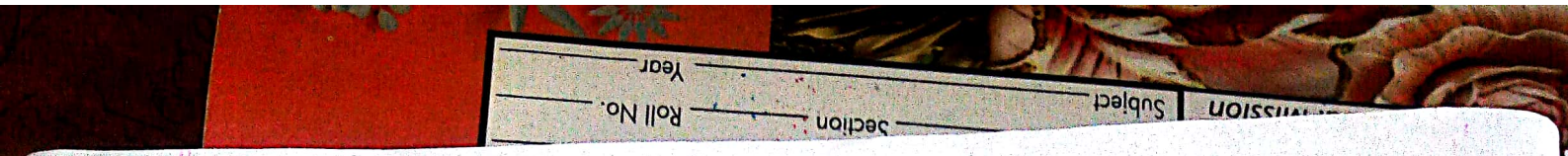
J = (4 + 1 + 3) % 6 = 2

Swap (S[1], S[2])

1st Step:

J = (0 + 0 + 4) % 6 = 4

Swap (S[0], S[4])



* True Random number generators: