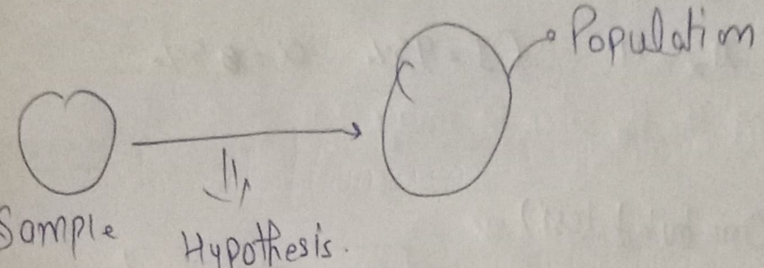
Inferential Statistics

Hypothesis Testing Mechanism



1. Null Hypothesis (H_0) :→ Assumption you are begning with.
2. Alternate Hypothesis (H_1) :→ Opposite of null hypothesis
3. Experiments → Statistical Analysis.
4. Accept the H_0 or reject H_0.

P-value

→ It is a no. calculated from a statistical test, that describes how likely you are to have found a particular set of observations if the null hypothesis were true.

→ if p is lie b/w ~~Signific~~ Confidence Interval then H_0 accepted.

Hypothesis Testing and Statistical Analysis

1. Z-Test } ⇒ Avergge ⇒ Z test → Z score & P-value
2. t Test } t test
3. CHI SQUARE } Categorical data
4. ANNOVA } Variance.

1. Z-test

(i) Population std (σ)

(ii) $n \geq 30$

* $\mu = 10$ $\sigma = 4$

$n = 100$ $\mu_{\bar{x}} = 11$ $CI = 95\%$ $\alpha = 5\%$

H_0 $\mu = 10$

H_1 $\mu > 10$ (One tailed test)

$\boxed{Z_{critical} = 1.645}$

~~Zscore = $11 \pm (1.645)\frac{(4)}{\sqrt{100}} = 11 \pm (1.645)(0.4)$~~

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{100}} = 0.4$$

$$Z = \frac{x - \mu}{\sigma_{\bar{x}}} = \frac{11 - 10}{0.4} = 2.5$$

$\boxed{Z_{critical} < Z}$

2. T- test

→ When we don't know population σ.

$$\boxed{t = \frac{\bar{x} - \mu}{s/\sqrt{n}}}$$

s = sample Std.

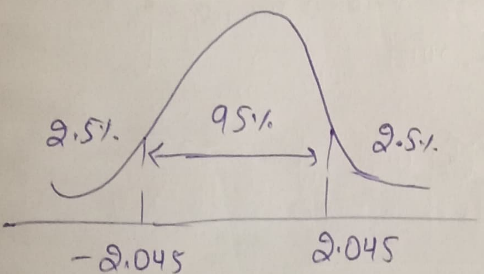
⇒ Degree of freedom : $\boxed{dof = n-1}$

* $\mu = 100$ $s = 30$ $\bar{x} = \mu_{\bar{x}} = 140$ $\sigma_{\bar{x}} = 20$

CI = 95% $\alpha = 5\%$

1. H_0 $\mu = 100$
2. H_1 $\mu \neq 100$ {2 tailed test}
3. $\alpha = 0.05$

   DOF = S-1 = 29.
4. Decision Rule : I



$t_{critical} = \pm 2.045$

$$t = \frac{140 - 100}{30/\sqrt{30}} = 10.96$$

$\boxed{t > t_{critical}}$

H_0 ✗

H_1 ✓

Type 1 and Type 2 Errors

Reality :→ Null hypothesis is True or Null hypothesis is false

Decision :→ H_0 is True or H_0 is false

→ Outcome :

Outcome 1 :→ We reject the H_0 when in reality it is false
↓
Good case

Outcome 2 :→ We reject the H_0 when in reality it is True
↓
Type 1 Error

Outcome 3 :→ We retain H_0, when in reality it is false
↓
Type 2 Error

Outcome 4 :→ We retain H_0, when in reality it is True
↓
Good case

Bayes Statistics (Baye's Theorem)

1. Independent Events
* Rolling a dice
* Tossing a coin

2 Dependent Event
* 2 color balls in a bag if a Red ball is picked, prob. of yellow ball next.

$P(R \text{ and } Y) = P(R) + P(Y/R)$

$$\Rightarrow \boxed{P(B/A) = \frac{P(B) * P(A/B)}{P(A)}}$$ Bayes Theorem

$P(A) = P(B) =$ events independent probabilities of A, B.

$P(A/B) = P(A)$ when B is occurred.

Cost Function

$w^T x = 0$ → line equation pass through origin.

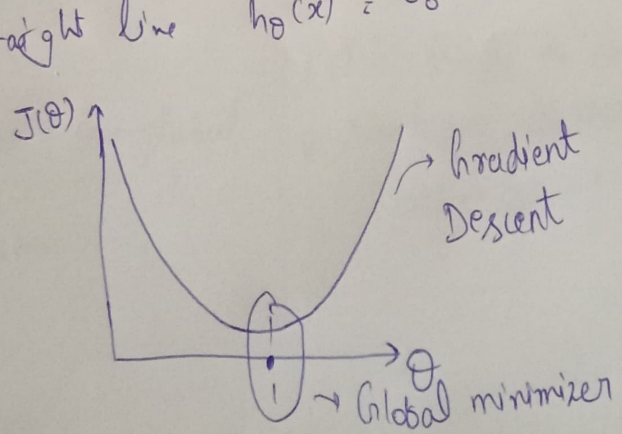
Cost function :-

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

$h_\theta(x)$ ↓ predicted, $y^{(i)}$ ↓ True O/P

Error to minimize

1. Eqn of straight line $h_\theta(x) = \theta_0 + \theta_1 x$

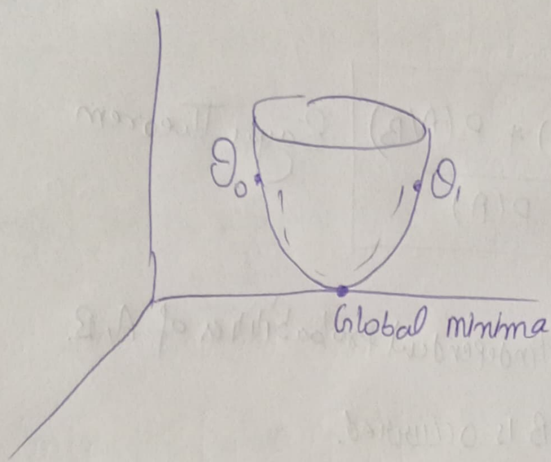


Convergence Algorithm

$$\theta_j = \theta_j - \alpha\left(\frac{\partial J(\theta_j)}{\partial \theta_j}\right)$$

$\alpha \rightarrow$ learning Rate
≈ 0.001

Conclusions :

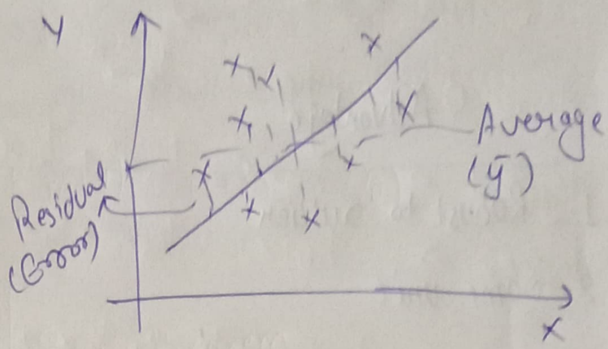


$\Rightarrow$ $\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$; $h_\theta(x) = \theta_0 + \theta_1 x$

$\Rightarrow$
$$\begin{cases} \theta_0 = \theta_0 - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x)^i - y^i)\right] \\ \theta_1 = \theta_1 - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x)^i - y^{(i)})x^i\right] \end{cases}$$

Performance Matrix

① R-squared



$$R_{squared} = 1 - \frac{SS_{Res}}{SS_{Total}}$$

$$= 1 - \frac{\Sigma (y_i - \hat{y}_i)^2}{\Sigma (y_i - \bar{y}_i)^2}$$

≈ 1 (more accurate model)

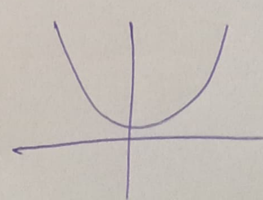
② Adjusted R-squared

→ If no. of features inc then Rsquared ↑ even if all are not related directly to final dependent feature

$$\text{Adjusted Rsquared} = 1 - \frac{(1-R^2)(N-1)}{N-p-1}$$

$N \rightarrow$ no. of data points
$p =$ no. of independent feature

Mean Squared Error

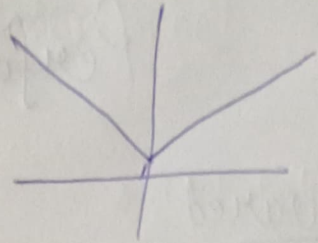


$$1\ MSE = \sum_{i=1}^{n} \frac{(y-\hat{y})^2}{n} = \text{Cost fun} \downarrow\downarrow$$

~~A~~ Advantage	Disadvantage
1. Differentiable	1. Not Robust to outlier
2. It has one local & one global minima	2. No longer ~~[illegible]~~ for same unit. (lakh → $lakh^2$)
3. Converges faster	

2. Mean Absolute Error (MAE)

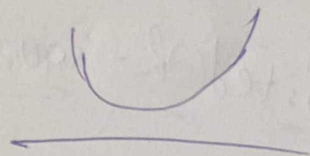
$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y-\hat{y}|$$



Advantage	Disadvantage
1. Robust to outlier	1. Convergance take more time as Optimization is a complex task as on some point differenitation not possible.
2. Same unit	

3. RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i-\hat{y})^2}$$



Advantage	Disadvantage
1. Same unit	1. Not Robust to outlier
2. Differentiable	