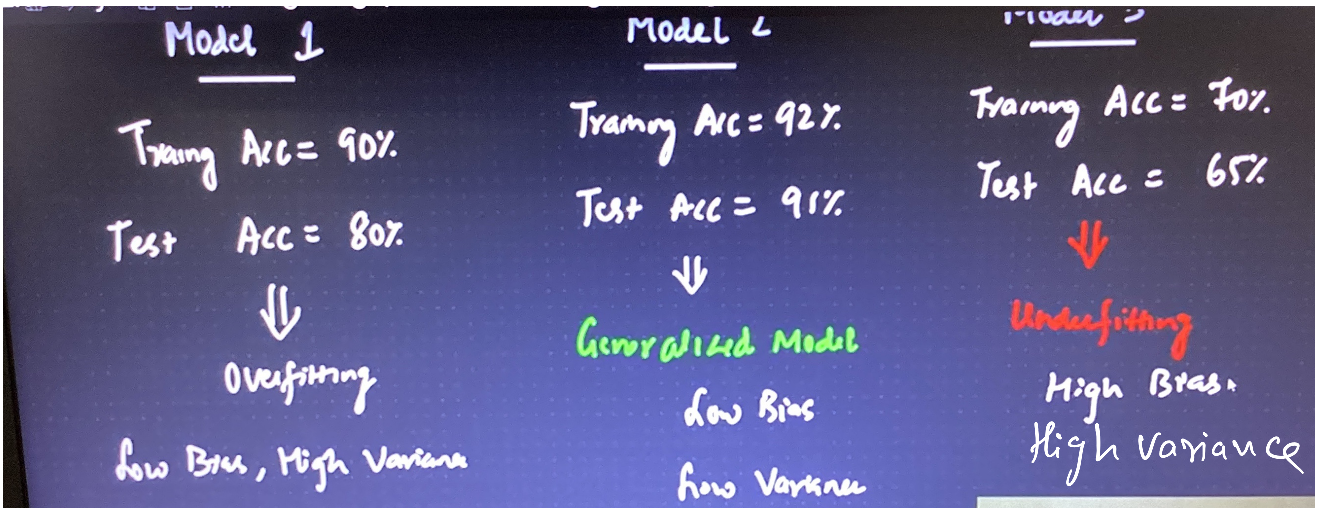
# Ridge And Lasso Regression:

→ To deal with overfitting ( performance well → Train data → Low Bias
bad → Test data ) → high Variance

note: underfitting
1. model accuracy is bad with training data → high bias
2. Mod " " also bad with test data → high Variance

Ex:




## Ridge (L2 regularization) :-

→ it basically adds a unique paramtere for the Linear regression Cost function
$\lambda$ = a hyperparamtre.

Considering the Cost fucn :-

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

Let $\boxed{h_\theta(x) = \hat{y}_i}$

we add $\lambda(\text{slope})^2$

$\Rightarrow$ $(\hat{y}_{(i)} - y_{(i)})^2 + \lambda(\text{slope})^2$

↳ (small value) ↳ (L2 regression)

→ it helps us to control overfitting and gives us a best fit line in a way which generalizes the model.

note: slope should not be steep, it causes overfitting

→ to train the model to reach this state, we need iterations and we call it hyperparametre. $(\lambda)$

notes for value of $J(\theta_0, \theta_1)$ can't be 'zero' if it is then it is over fitting.

## Lasso Regression ($L_1$ regression):

$$= (\hat{y} - y)^2 + \lambda \, |slope|$$

for feature selection

note: apart from avoiding overfitting its helpful in feature selection.

$$|slope| = |\theta_0 + \theta_1 + \theta_2 + \ldots + \theta_n|$$

⇒ for the values which are not affecting the model much, their respective slopes ($\theta_0$) will be minimal and we can actually neglect them using $\boxed{mod}$

⇒ hence, it is helping in feature selection.

note:

1. In $L_2$ we are squaring
2. In $L_1$ we are using mod

→ we also use $\lambda$ for cross validation.

note: Try both $L_1$ & $L_2$ and use whose performance matrix is good.

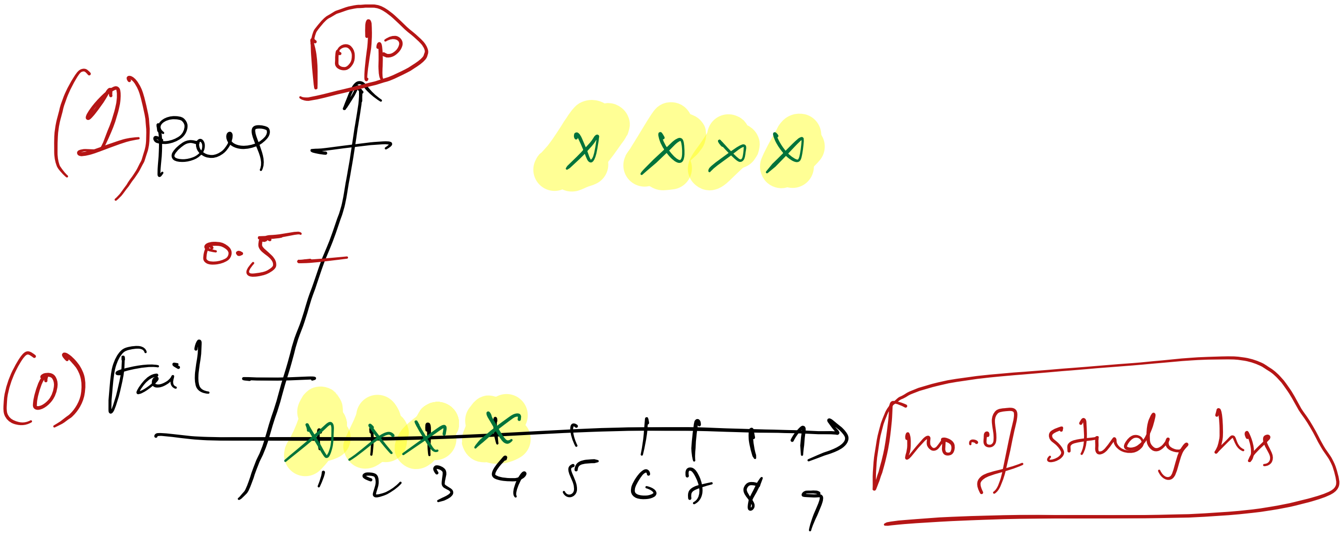
note:

~~xxx~~ google the Assumptions of Linear Regression

# Logistic Regression: (classification Algo)

→ Binary class classification

ex: we will plot points for study / pass data




note: can't we solve this using Linear?

⇒ we can solve, we can find the best fit line with our conditions.

Then why logistic?

↳ Because of outliers present in the data

(1.) [Even after treating outliers], the best fit line which we get from Linear Regression will be of no use

(2.) we might also get our outcomes greater than 1 and less than 0 as well

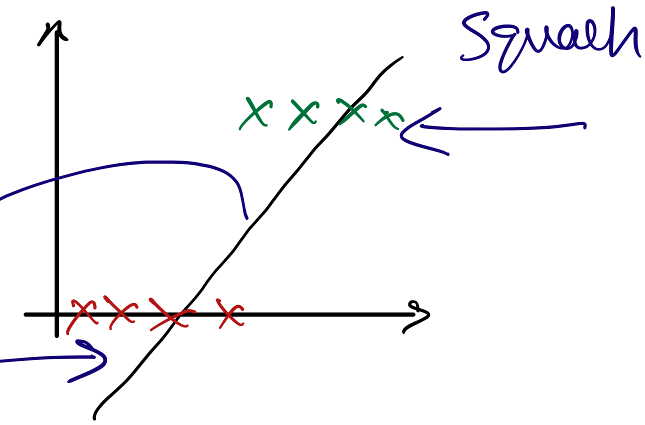
We squash those extra lines which are going out of boundaries using sigmoid function in this logistic Regression.

## Decision Boundary :-

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

$$h_\theta(x) = \theta^T(x)$$

$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x}$$




to squash the line, we need smtg called as sigmoid funcn.

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

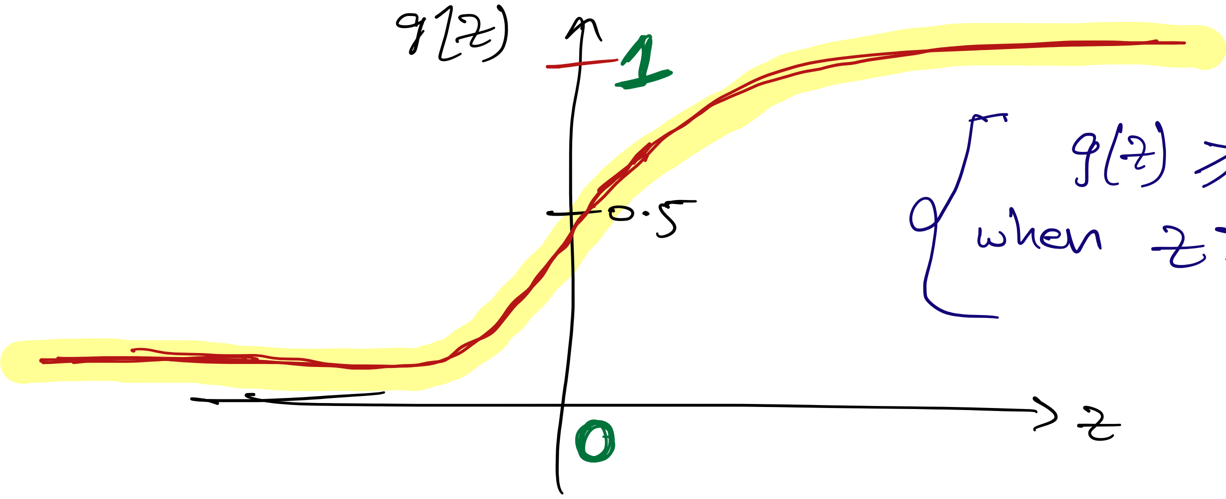
let, $z = \theta_0 + \theta_1 x$, $h_\theta(x) = g(z)$

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

$$\boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}}$$

sigmoid (or) Logistic function

graph of sigmoid function:




$g(z) \geq 0.5$ when $z \geq 0$

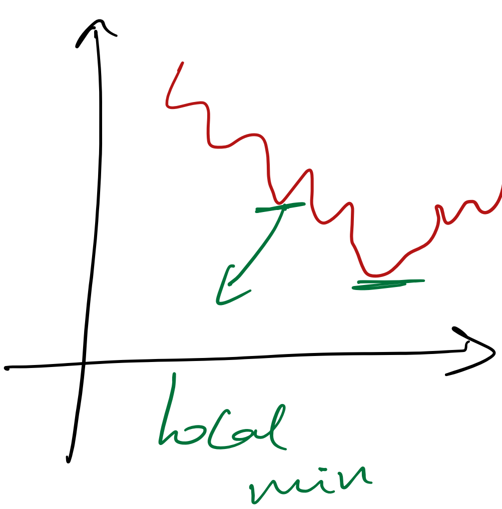
Cost func$^n$

$$h(\theta) = \frac{1}{1+e^{-(\theta,x)}}$$

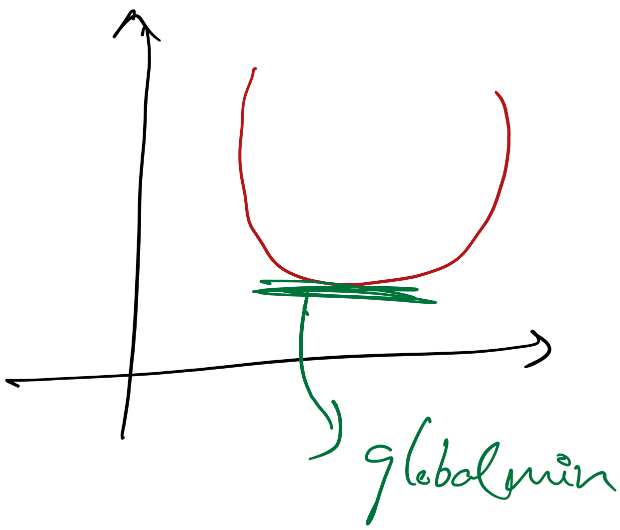
$$\frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x)^i - y^i\right)^2$$

this is a non-convex func$^n$

Non-Convex




Convex



note:

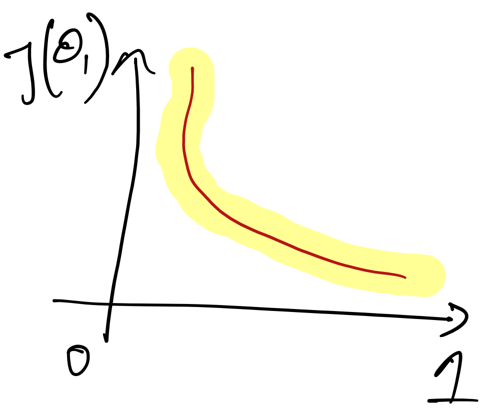
In case of logistic Regression, we can't find the global minima

So, the actual cost func$^n$ for logistic Reg:-

$$J(\theta_1) = \begin{cases} -\log(h_\theta(x^i)) & y=1 \\ -\log(1-h_\theta(x^i)) & y=0 \end{cases}$$

$$\boxed{h_\theta(x) = \frac{1}{1-e^{-(\theta_1 x)}}}$$

if $y=1$

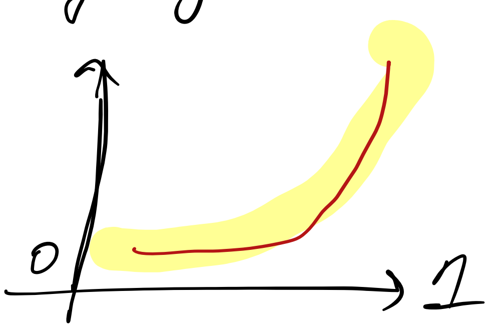



(classification)

⇒ assumptions from graph:

cost = 0 if $y=1$, $h_\theta(x)=1$

if $y=0$



Combining both the Costfun :-

$Cost(h_\theta(x)^i, y^i) =$

$$\boxed{-y\log(h_\theta(x)^i) - (1-y)\log(1-h_\theta(x))}$$

if $y=1$

$= -\log h_\theta(x)^i$

if $y=0$

$= -\log(1-h_\theta(x)^i)$

including this $h_\theta(x)$ value in the $J(\theta_0, \theta_1)$ formula

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

$$\boxed{h_\theta(x^i) = \frac{1}{1+e^{-\theta_1 x}}}$$

repeat until Convergence

$\{$

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

$\}$

## Performance Matrix: (Binary classification)

ex: $x_1$ $x_2$ $y$ → actual o/p   $\hat{y}$ → o/p of predicted model

| $y$ | $\hat{y}$ |
|---|---|
| 0 | 1 |
| 1 | 1 |
| 0 | 0 |
| 1 | 1 |
| 1 | 1 |
| 0 | 1 |
| 1 | 0 |

| (pred) \ (Actual) | 1 | 0 |
|---|---|---|
| 1 | III | II |
| 0 | I | I |

---

**Confusion Matrix:-**

| (predicted) \ (Actual) | 1 | 0 |
|---|---|---|
| 1 | TP | FP |
| 0 | FN | TN |

---

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

ex: $\frac{3+1}{3+2+1+1} = \frac{4}{7} = 0.57 = 57\%$

note:

we should try to reduce FP and FN

Case:

(1) $0 \rightarrow 900$, $1 \rightarrow 100$ } Imbalanced data

(2) 0 → 600
1 → 400 } balanced data

---

(1) precision (2) Recall (Sensitivity) (3) F-score

$$\frac{TP}{TP+FP} \qquad \frac{TP}{TP+FN}$$

ex: {spam classification} → Precision

{Has Cancer (or) NOT} → Recall

{Tommorow Stock market is going to crash} → F-score

(1) F-Beta

$$= (1+\beta^2)\frac{\text{Precision} \times \text{Recall}}{\beta^2 \times (\text{precision} + \text{Recall})}$$

$\beta = 1$ (F1 score) $\Rightarrow$

$$(1+1)\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$= \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

[Harmonic Mean]

$\beta = 0.5$ $\quad (1 + (0.5)^2) \dfrac{P \times R}{(0.25)\, P + R}$

($F_{0.5}$ score)

$\beta = 2$ $\quad$ [FN >> FP] priority

($F_2$ score)

1. z score
chi square
hypothesis
ROC curve
AOC
Cut-off

## Hypothesis Testing:

→ We evaluate '2' mutual Exclusive statements on a population using sample data.

Steps:

1. Make Initial Assumption (Ho) null hypothesis
2. Collect Data [evidence]

→ using this we check wether to take null hypothesis or not

3. Gather Evidence to reject (or) not reject null hypothesis

→ if we cant get info from '2' step to Ho we declare it H1 → Alternate hypothesis ex:- $H_0$ – defendant is innocent

$H_1$ – " " feeling guilty.

| | $H_0$ | $H_1$ (Truth) |
|---|---|---|
| Do not Reject | OK | Type 2 Error |
| Reject | Type 1 Error | OK |

Type 2 Error → due to lack of evidence, we dont choose $\boxed{H_0}$ then we choose $\boxed{H_1}$ but in future $\boxed{H_0}$ might be true / not

Type 1 Error ⇓ we choosed $H$ coz we thought its true but it's not.

note! which is more dangerous? Type 2 (or) Type 1

⇒ it depends on the data we choose.

→ To check wether $H_0$ or $H_1$ is true we use T test, Chi Square test, Anova test ....

ex:- [for 'obj' dtype]

One Categorical

| Gender | Age Group | Weight (kg) | Height (cm) |
|---|---|---|---|
| M | Elderly | 70 | 1.4 |
| F | Adult | 65 | 1.2 |
| M | Adult | 65 | 1.4 |
| M | Child | 20 | 1 |
| F | Adult | 75 | 1.3 |
| M | Elderly | 80 | 1.3 |

$H_0$ – There is no diff

$H_1$ – There is a diff

Test - One Sample proportion Test

$P \leq 0.05$ → Significance value ($\alpha$)

## Two Categorical :-

Test – Chi Square test

if $p \leq 0.05$

then it takes $\boxed{H_1}$

## One Continuous Variable:

Test – T test

Note:

If we have categorical data with more than 2 variables

$\Rightarrow$ Anova test.