* Logistic Regression

Supervised.
- Regression → y → continuous
- Classification → y - discrete.

* Classification

→ Predict pass or fail based on no of hours.

X (no of hours)	Pass / fail
2	1
3	0
4	1
5	1

Sinc

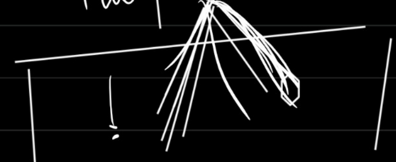
* Predict diabetes or not.

Cholestrol level	diabetes. (y)
100	1
250	0
80	1
–	0
–	–
–	–
–	–

*

Salary	Age	Fraud
20K	19	1
22.5K	20	1

Two possibility



* Spam | ham.
* Cancer / not cancer.

Classification → binary classificat → bi → 2 category

Classification → multi-class classification → >2 Categories

Age	Salary	Credit Score
18	25	Good — 1
		Fair — 2
		bad — 3

Iris

Setosa Viginica Versi

50%

Soumya, Shubhaker

⇓

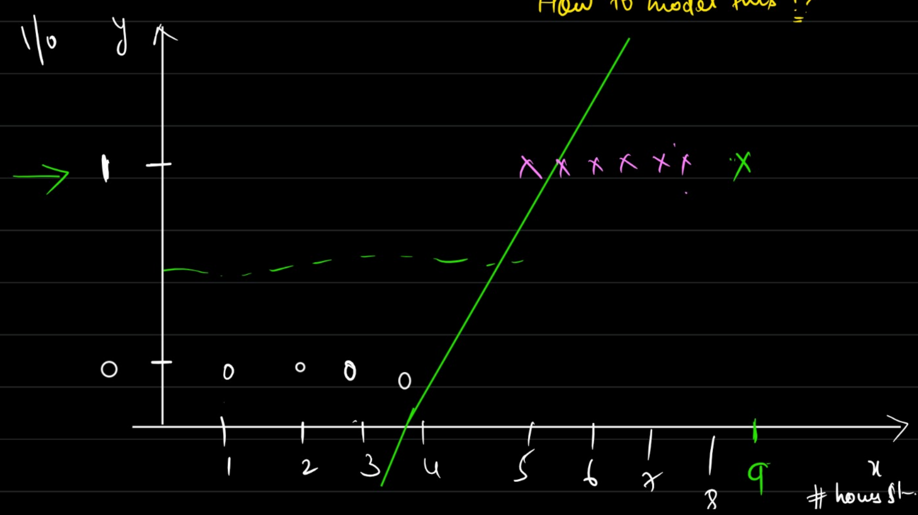
To reward with 50% of fee

* There are diffeent Algos to solve classification problem.

→ Logistic Regression :

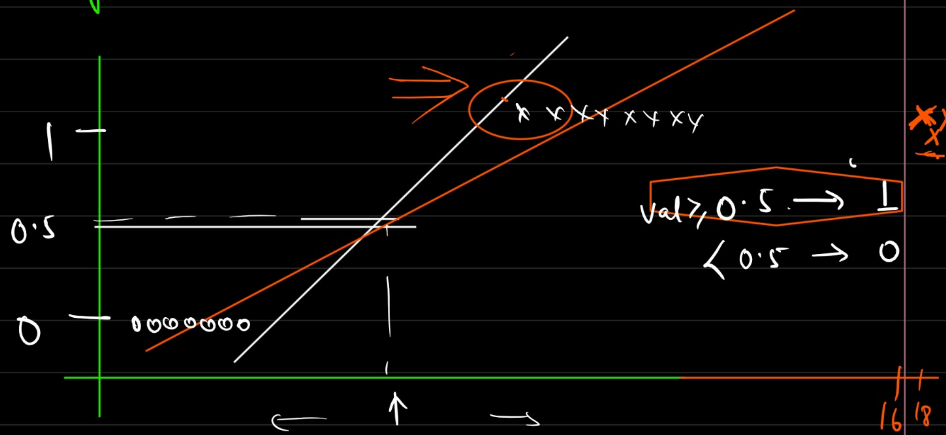
X #hours	Pass/fail
2	0
8	1
3	0
4	0
6	1

How to model this ??



Q Why we can not use regression line?

① Best fit line will change due to presence of outliers.



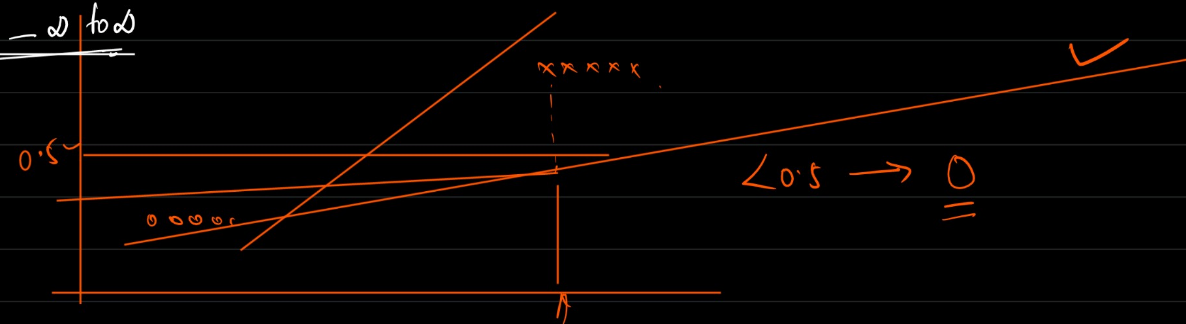
val ≥ 0.5 → 1

< 0.5 → 0

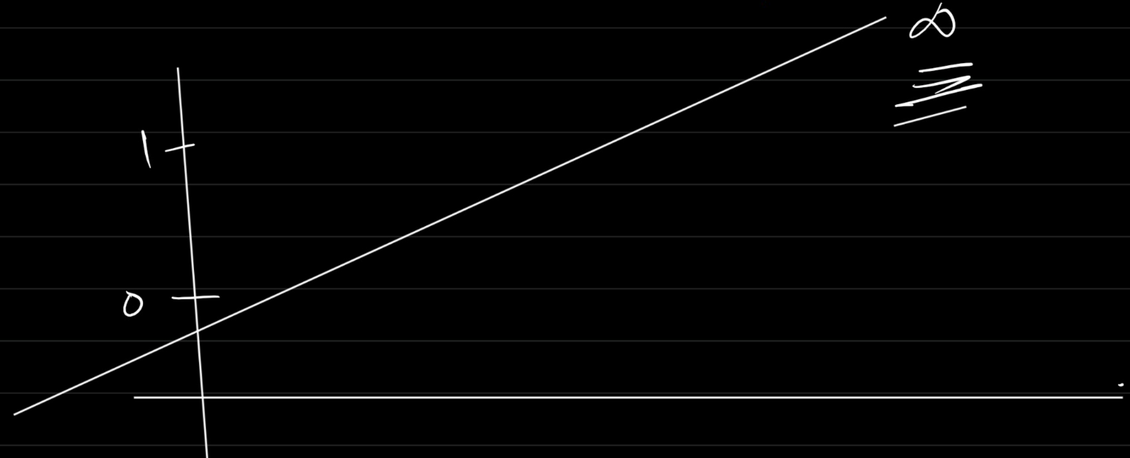
② range of output - y (classification)

0 & 1

regression line -∞ to ∞

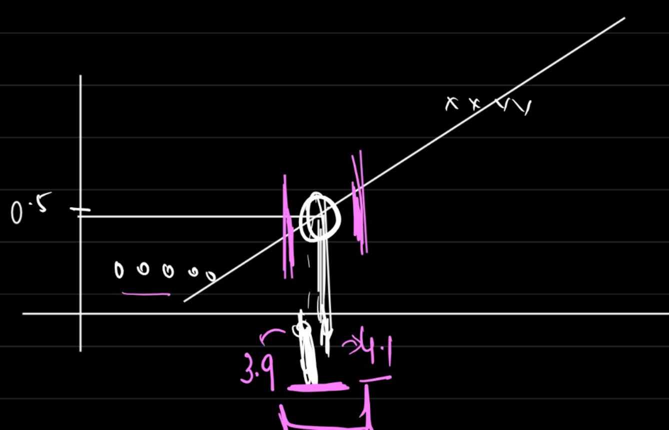


< 0.5 → 0



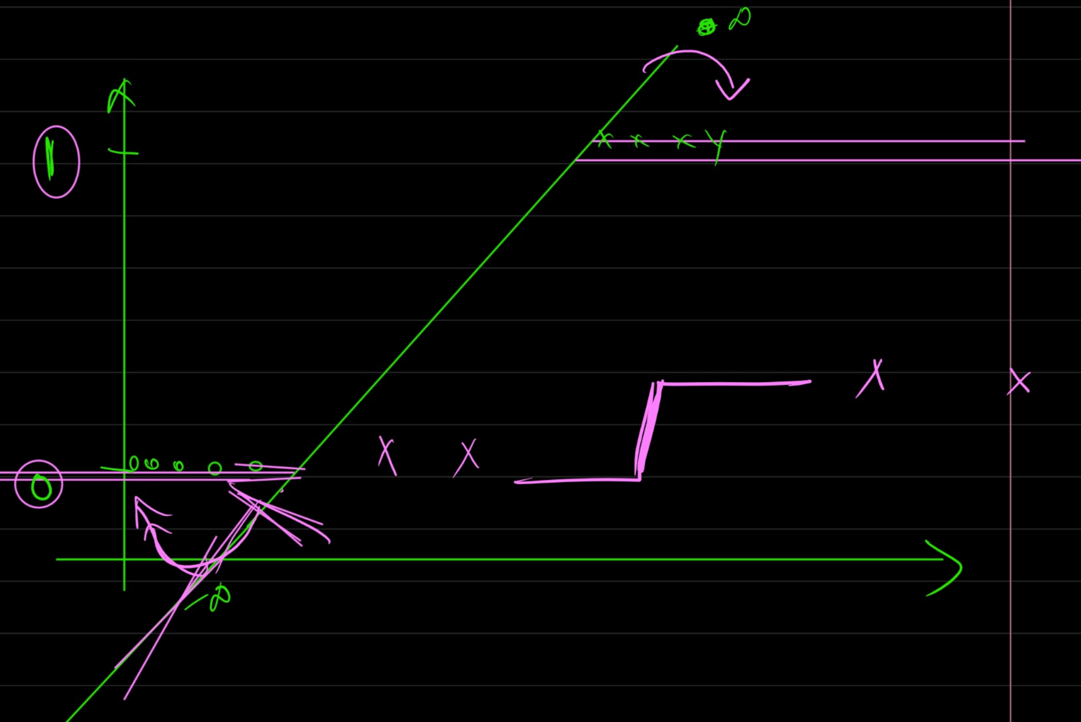
③ cut off

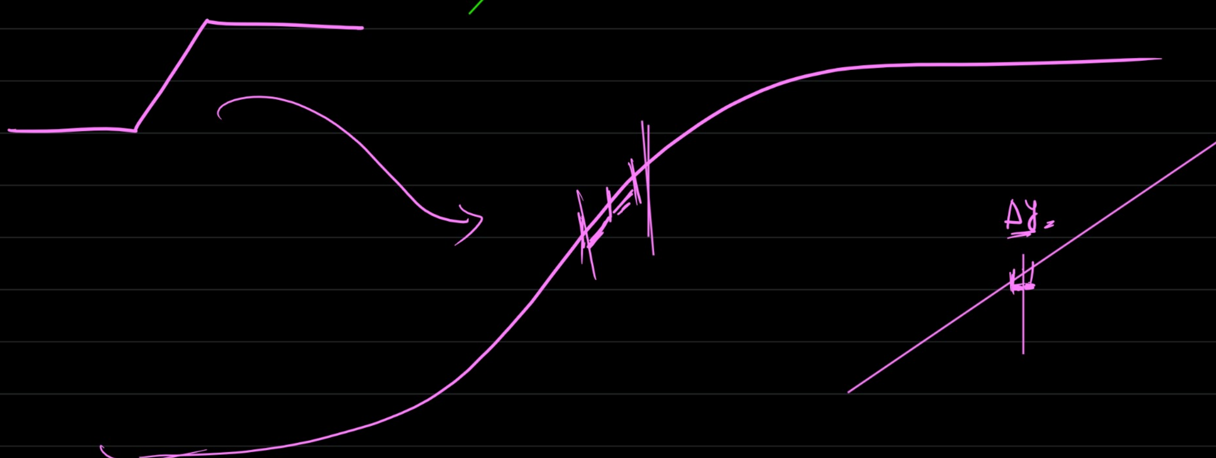
↳ Prediction changes suddenly

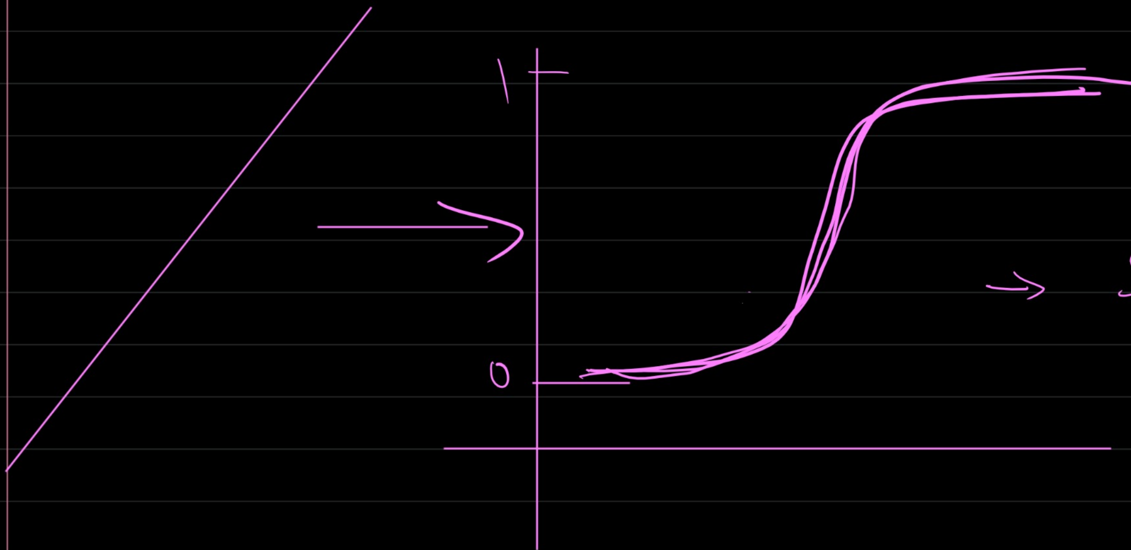


① → outlier
② range −∞ to ∞

① −∞ to ∞
② outlier
③ cutoff







→ S shaped curved

↓

Sigmoid

⇓

logistic fun

⇓

σ (sigmoid) $= \dfrac{1}{1+e^{-z}}$

e – euler no
2.371......

logistic Regression

Outlier curve →

$-\infty$ to ∞

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

$-\infty$ to ∞ → 0 → 1

/ → 0 to 1

$$z=-\infty \Rightarrow \frac{1}{1+e^{-(-\infty)}} = \frac{1}{1+e^{\infty}} = \frac{1}{\infty} = 0$$

$$z=\infty = \frac{1}{1+e^{-\infty}} = \frac{1}{1+\frac{1}{e^{\infty}}} = \frac{1}{1+\frac{1}{\infty}} = \frac{1}{1+0} = 1$$

Proved :

$h_\theta(x) = \theta_0+\theta_1 x$ → Best fit line

$h_\theta(x) = \sigma(\underbrace{\theta_0+\theta_1 x}_{z})$

→ $\dfrac{1}{1+e^{-z}}$

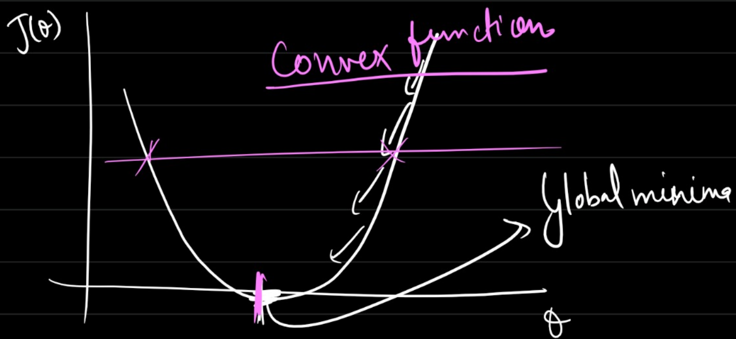
$$\boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}} \rightarrow \text{Logistic regression model.}$$

To get optimal θ_0 & θ_1, minimise the Cost function.

$$CF = \frac{1}{n}\sum (y_{act} - y_{pred})^2$$

* Linear Regression model

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



Logistic regression model

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

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

CF = Loss fun

Log loss function

$$J(\theta_0,\theta_0) = -y_i\log(h_\theta(x)_i) - (1-y_i)\log(1-h_\theta(x_i))$$

$$\text{Prd} \rightarrow h_\theta(x)_i = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}$$

$$J(\theta_0,\theta_1) = \begin{cases} -\log h_\theta(x)_i & \text{if } y_i = 1 \\ -\log(1-h_\theta(x)_i) & y_i = 0 \end{cases}$$

y_i – actual value

$$\frac{1}{1+e^{-(\theta_0+\theta_1 x)}}$$

To minimise the CF, $\underline{J}(\theta, \theta_1)$

Chang θ_0 & θ_1

Convergence Algorithm

Repeat unil convergence

$$\left\{ \left\{ \underline{\theta_j : \theta_j - \alpha \frac{\partial}{\partial \theta} J(\theta_0, \theta_1)} \right\} \right\}$$

to get optimal θ_0, θ_1

for multiple variable

$$h_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1 + \theta_2 x_2 \quad \ldots \quad \theta_n x_n)}}$$

* logistic regression with Regularisation

$$CF = J(\theta_0, \theta_1) = -y \log(h_\theta(x)) - (1 - y_i) \log(1 - h_\theta(x))$$

Rigdey = CF + L_2 regularization (reduce overfitting)

$\hat{\lambda}(\text{slope})^2$

Lasso = CF + L_1 " (feature select)

$\hat{\lambda}|\text{slope}|$

elastic net = CF + L_1 + l_2 (Both)

$CF + \lambda_1 (\text{slope})^2 + \lambda_2 |\text{slope}|$

$C = \frac{1}{\lambda}$

Logistic Regression

* Logistic Regression

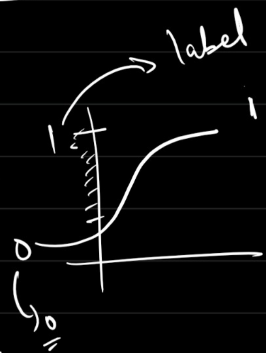
$\frac{1}{1+e^{-z}}$ → built a model ?

No of hours Studied	Pass / fail.
5	1
2	0
3	0
6	1

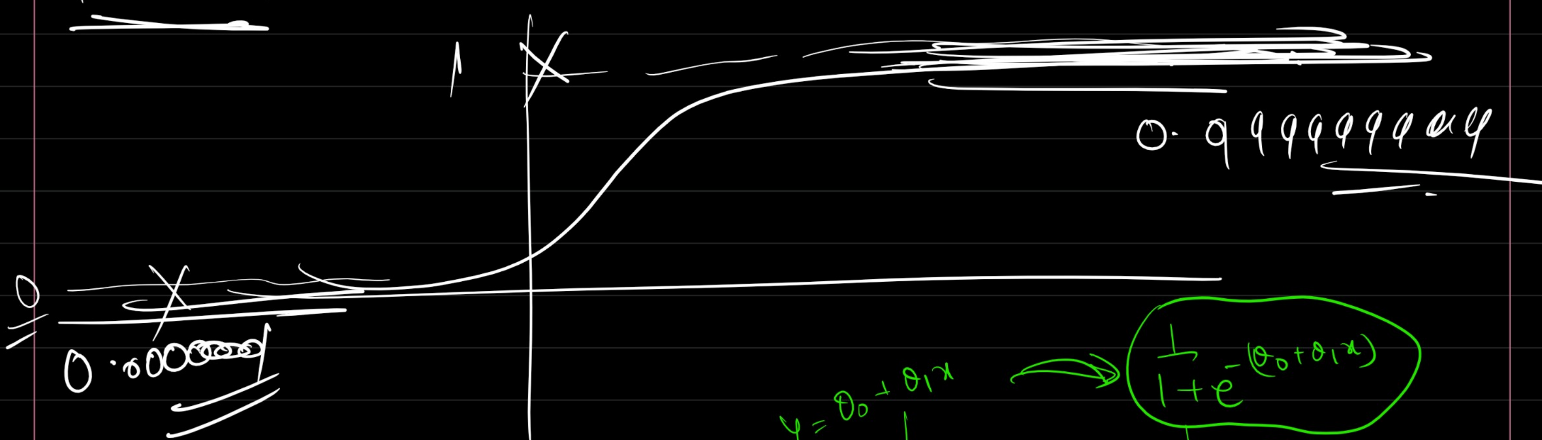
0 - 1

⇓

Probability



Sig → 0 to 1 (tend)



$y = \theta_0 + \theta_1 x$ ⇒ $\frac{1}{1+e^{-(\theta_0+\theta_1 x)}}$

N Samples $< \begin{matrix} 0 \\ 1 \end{matrix}$

→ For samples labelled 1 : Estimate θ, such that $\hat{p}(x)$ is as close as 1 possible

→ For sample labelled ; Estimate θ, such that $1-\hat{p}(x)$ is as close to 1 possible

or

$\hat{p}(x)$ is close to 0 possible

$$\rightarrow \prod_{s \text{ in } y_i = 1} p(x)_i \cdot \prod_{s \text{ in } y_i = 0} 1 - p(x_i))$$

$$\boxed{L(\theta) = \prod \left(p(x_i)^{y_i} \times (1-p(x)_i)^{1-y_i} \right)}$$

$a^x \Rightarrow x \log a$

Likelihood

$$\log(L(\theta)) = \sum_{i=1}^{n} y_i \log p(x_i) + (1-y_i) \log (1-p(x_i))$$

log likelihood esti.

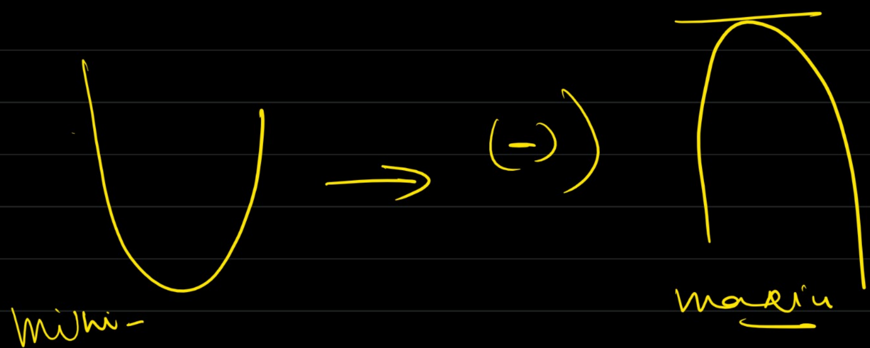
Maximum Likelihood estimation $-$

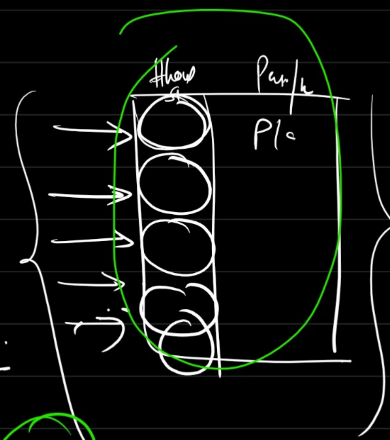
$$-\log(L(\theta)) = -y_i \log p(x) - (1-y_i) \log (1-p x_i)$$

$$-\ell(\theta)$$

LR → minimi

$-$logloss





→ Binomial distribution

n – bernoulli trial

⇓

$$= p(x)^{y}\,(1-p(x))^{1-y} \quad \to 0 \quad \to 1$$

$$y = 1 \Rightarrow p^{1}(1-p)^{1-1} = p^{1}(1-p)^{0} = p$$

$$y = 0 \Rightarrow p^{0}(1-p)^{1-0} = 1-p$$

$$p + 1 - p = 1$$

99.99

✓ CF = LR

Closed form soln

⇓

taylor series,
Gradient, Hessian Matrix

Gradient descent

m.c

Lot $\theta_0 \theta_1$

$\theta_0 \theta_1$

approx

Newton Rhapson

Sag

Saga

10 class 0

10 class 1

80 | 20

validation

CV.

20% | 80%

m.c

m.c

m.c

m.c

S CV

m.c → performing model

diabetes

Regression

MLR

DT Regressor

MSE

R square

95 : 5

0 1

99 : 1

0 1

1 x 10

1 x n

Randomised Search CV

Scaler $\frac{x-\mu}{\sigma}$

2
3
4
5
1
2
3

$\mu = 2$

$\sigma = 3$

Scaler

fit, transform.

M.r.

train – fit_transform

test Scaler.transform()