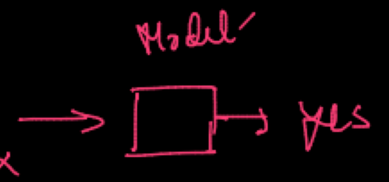
Opera File Edit View History Bookmarks Developer Window Help

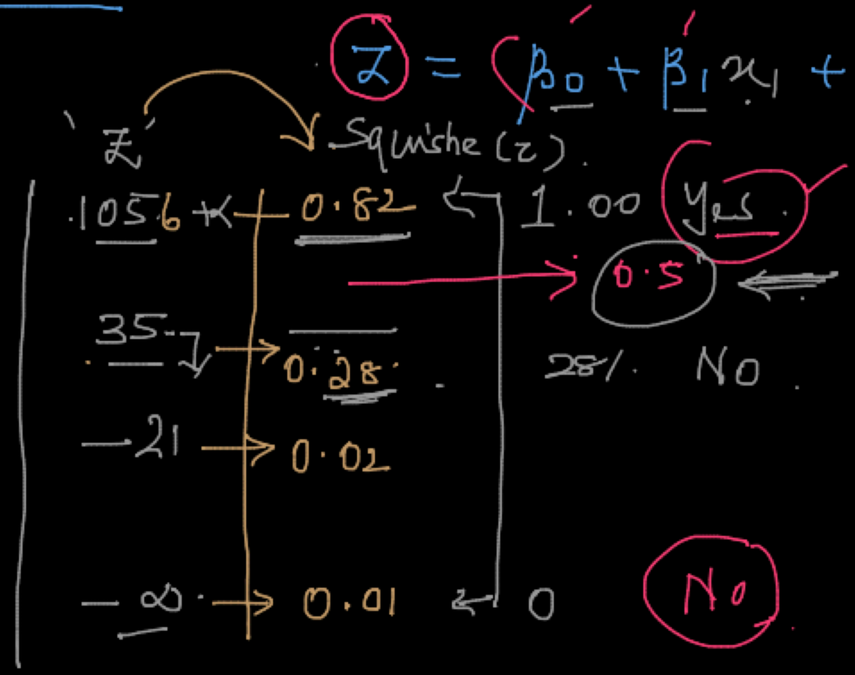
Logistic Regression

- classification Model
- 'y' is categorical
- linear Model



x_1	x_2	x_3	
BP	Cholestrol	WC	y
200	75	125	Yes
80	23	64	No

$$z = (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$$



fn(z) ⇒ 0.0 1.0

↳ Sigmoid Fu.

x_1	x_2	x_3	z	p	$\hat{y}$
200	95	123	1200	0.8	Yes
80	23	64	-56	0.1	No

$z = -\infty + 0 + \infty$

$g(z)$ 0.0 0.5 1.0

$z = +\infty$

$g(z) = \frac{e^{\infty}}{1+e^{\infty}} = 1.0$

$z = -\infty$

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

or

$= \frac{0}{1+0} = 0$

$z = 0$

$g(z) = \frac{e^{0}}{1+e^{0}} = \frac{1}{2} = 0.5$

$z = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + \beta_3 x_3$

$-\infty$ to $+\infty$

0.5 Threshold.

$z \rightarrow$ Fn. $\rightarrow$ 0.0 / 1.0 } probability.

$-\infty$ to $+\infty$

$g(z)$.

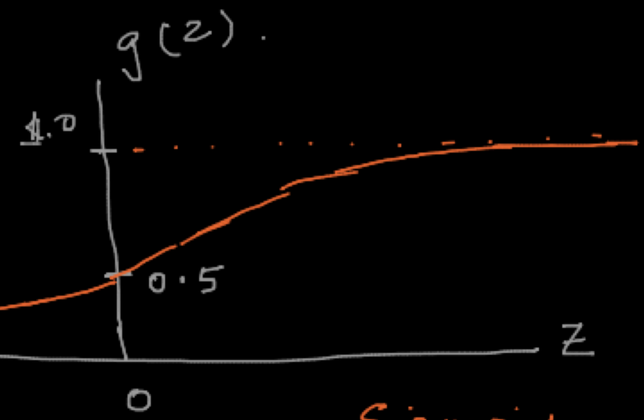
Sigmoid.

Sigmoid$(z) = g(z) = \frac{1}{1+e^{-z}}$

$g(z) = \frac{1}{1+\frac{1}{e^{z}}} = \frac{e^{z}}{e^{z}+1}$

$-\infty$ & $+\infty$.

Sigmoid



Opera File Edit View History Bookmarks Developer Window Help

	x_1	x_2	x_3	y	$\hat{y}$	
1	–	–	–	0	0	TN
2				0	0	→ FP
3				0	1	→ FP
4				0	0	TN
5				1	0	→ FN
6				1	1	→ TP

TN = 2 ✓ TP = 1 ←
FP = 0 ✓ 1 → FN

$$\text{Accuracy} = \frac{4}{6} = 66.66\%$$

50%

Highly Imbalanced Dataset -

0 ≈ 50% } Balanced.
1 ≈ 50% }

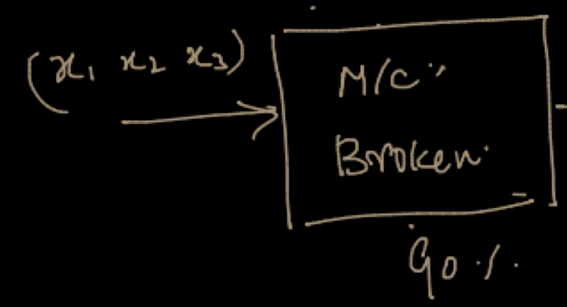
4/6

0 ≈ 90% } 'Imbalanced Dataset'
1 ≈ 10% }

0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
1	0

50%

→ 90%



100

90 NO, 100 NO.

Opera File Edit View History Bookmarks Developer Window Help

Confusion Matrix

predicted \ Ground Truth	True	False
True	TP ✓	FP
False	FN	TN ✓

$y \quad \hat{y}$

$- \quad -$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{\text{Actually true cases}}{\text{Cases we predicted as true}}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

(TPR, Sensitivity)

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

$$\frac{\text{We predicted as true}}{\text{Cases who actually have the diseases}}$$

Actual → 1 1 1 1 1 1 1 1 0 1 → No disease 3, disease 7

We predict → 1 1 1 1 1 1 1 1 0 1

$$\text{Precision} = \frac{4}{5} \rightarrow \text{Precision}$$

$$\text{Recall} = \frac{4}{7}$$

Opera File Edit View History Bookmarks Developer Window Help

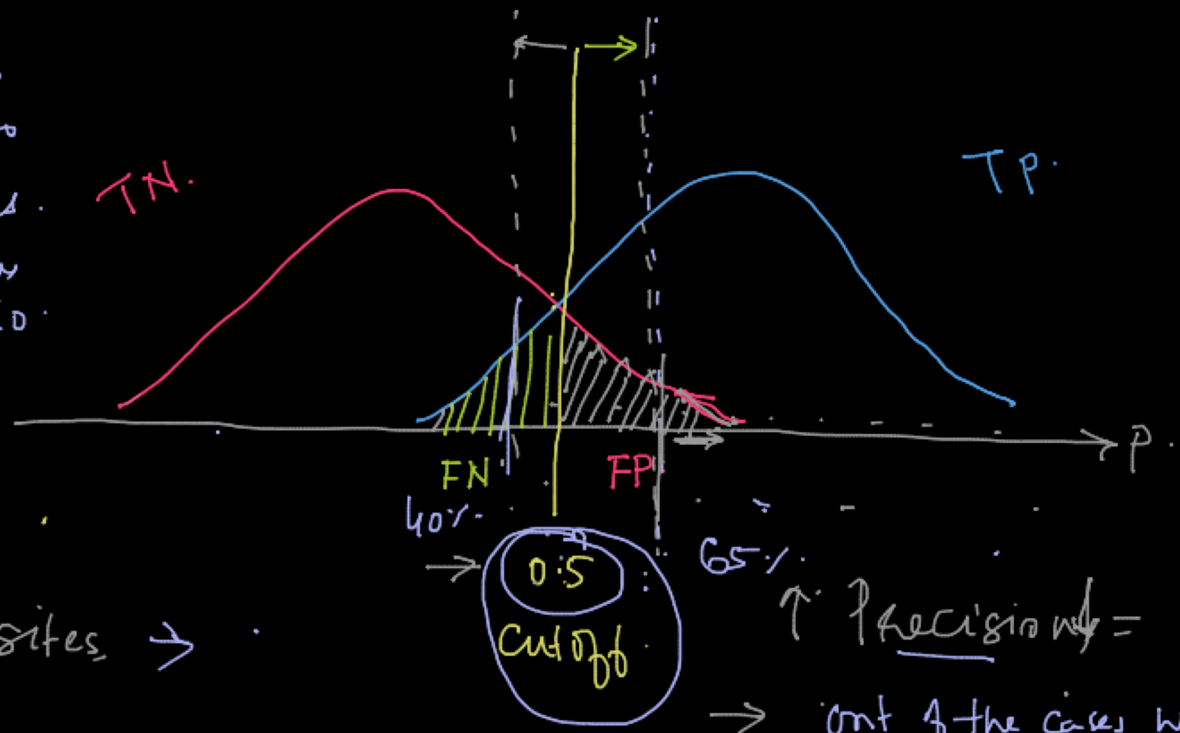
y ŷ
Yes ← Yes
No ← No
Yes → Yes
No — Yes
No — No

TN.

TP.

→ No

2 → ⇒ large
→ Curved



FP ↓ FP ↑
FN ↑ FN ↓

1. Safe Websites →

Precision → Out of Sites we predicted as safe how many were actually Safe.

↑ Precision = $\frac{TP}{TP + FP}$ ↑ ↓

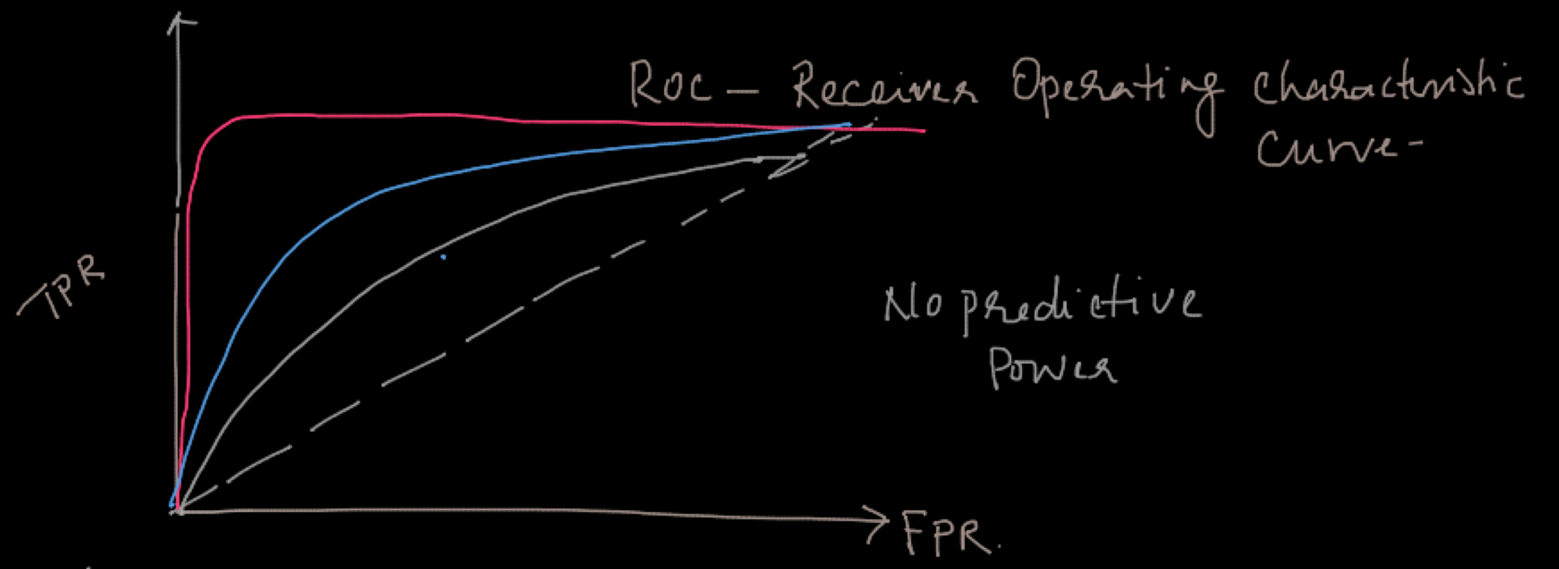
→ out of the cases we predicted as true how many Were actually true.

↓ Recall ↑ = $\frac{TP}{TP + FN}$ ↓ ↑

— out of cases that were actually how many did we correctly predict.

(Recall) → Out of Sites that were actually Safe. how many did we Correctly Say Safe

5 / 6



AUC → ROC

Regression

R^2

MSE / RMSE

MAE / MAPE

Classification

Accuracy / F1. score.

Recall

Precision

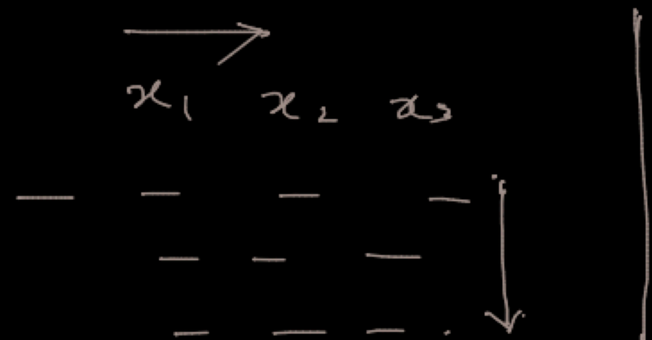
AUC – ROC

1. Predicting Spam ✓ →

2. Risky Cases →

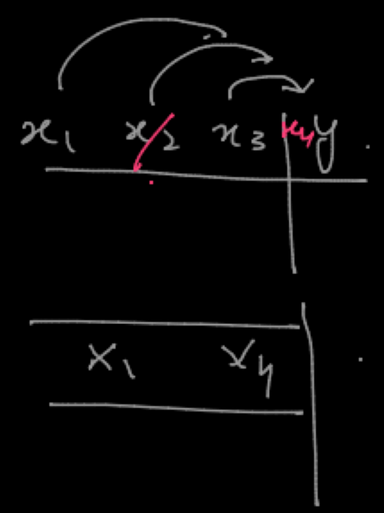
Imbalanced Dataset → Both Acc / F1-score

$$\text{F1-Score} = \frac{2\,(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$



Final Model

→ ?



1. What model to build.

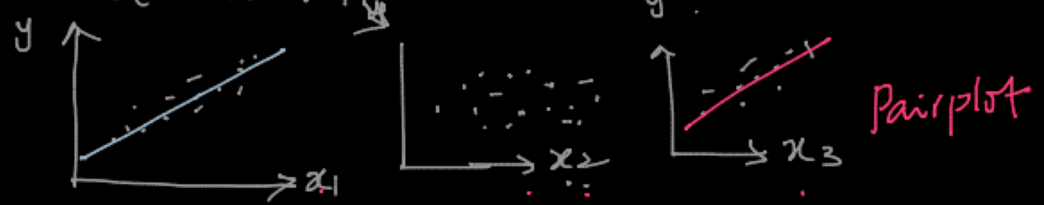
→ Regression / classification

y - continuous y - categorical.

Regression.

1. Can we build linear model ?

a. Scatter plot



b. find the correlation coefficient

$r \geq .85$

2. Is there a relation between X-values.

(a) x_1 x_4 — $y \leftarrow x_1$ R^2
$y \leftarrow x_4$

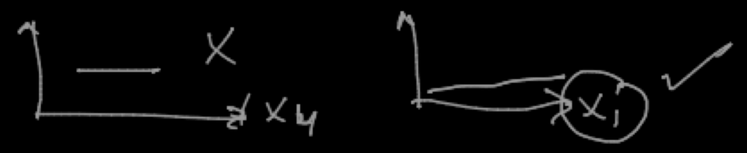
(b) VIF = x_1 x_3 x_4.

$x_1 \rightarrow x_3$ x_4. $-R^2 = \frac{1}{1-R^2} = VIF_{-x_1}$. Smallest VIF →.

$x_3 \rightarrow$

Feature Engineering

(c) Partial Regression

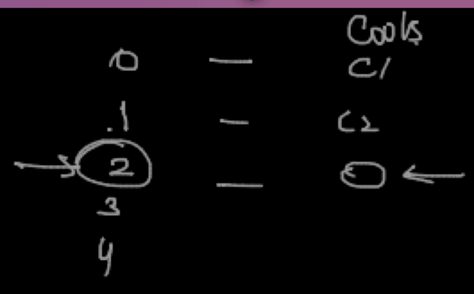


zoom.us Meeting View Edit Window Help

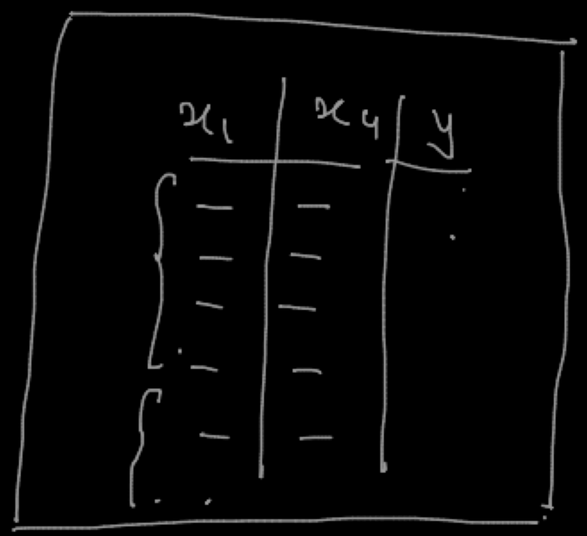
Any Records / datapoints - influential?

→ cook's distance ..
→ Inflation

} Remove data points.



Outliers, High leverage → Influential.



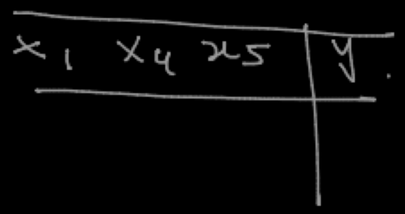
train-test-split

→ . low Training Error
High Testing Error
overfitting.

High Training Error
Testing Error high.

R^2
MSE, RMSE
MAE.

} Metric.



→ ⋮ → deployment.

Accuracy / F1
Prec
Recall
AUC-ROC.