Linear Regression → Target data → Continous $(-\infty, +\infty)$

What if the data is discrete ?

Categorical

classify something

Will it rain or not ?
pass or fail ?
QA pass or fail ?
email spam or not ?
Credit tran fraud or not ?
Loan give or not ?
Sales Buy or not ?
Health Cancer or not ?

} 1 or 0 Binary Classification

Classification Problem

Logistic Regression

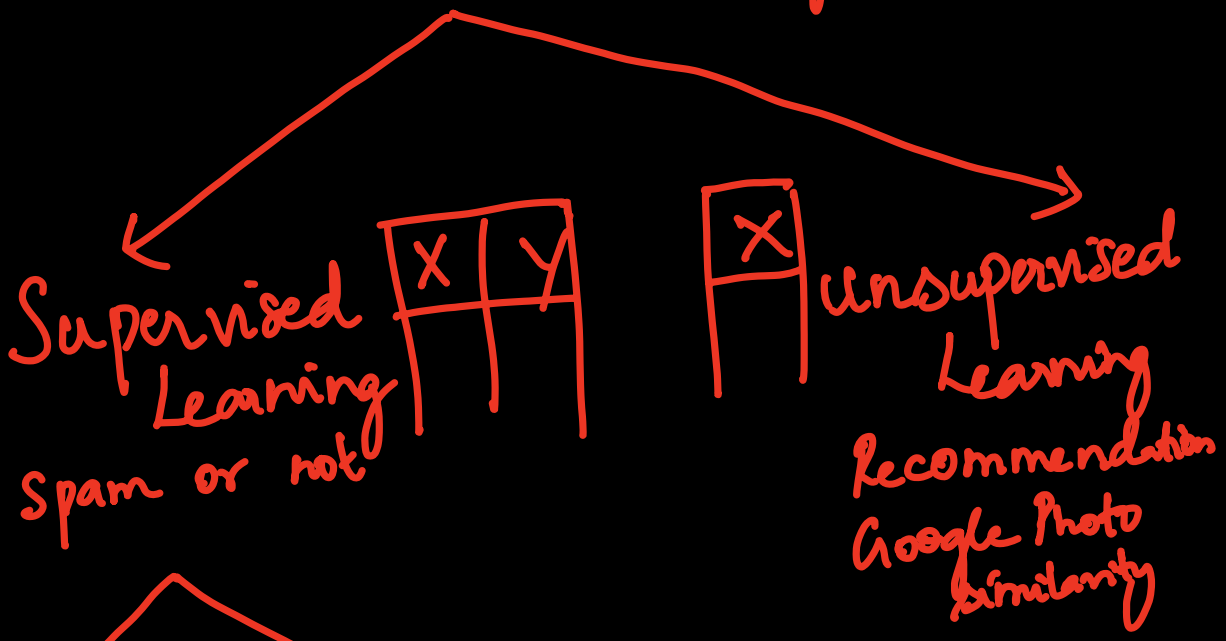
Classification

Linear Regression → Classification

↓

Logistic Regression

Machine Learning Problem



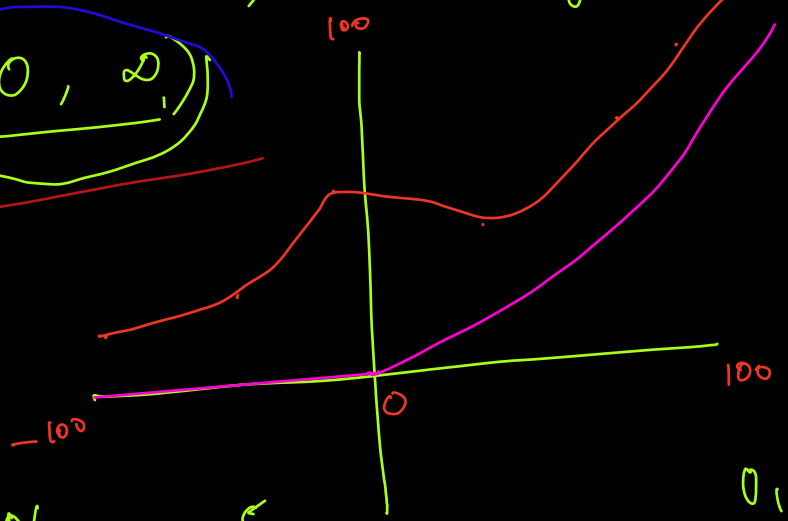
Transform Linear to Logistic

1. Linear → output → $(-\infty, +\infty)$

Agenda → Convert $(-\infty, +\infty)$ ⇒ $(0, 1)$ Probability

2. Exponential $(e^{\wedge})$ ⇔ Anti log

$(-\infty, \infty)$ ⇒ $(0, \infty)$



3. $N \Rightarrow \frac{N}{N+1}$ $\quad \frac{5}{6} =$ $\quad \frac{10}{10} = 1$

$(0, 1)$ $\quad 0 \to 1$, $\quad 0.\overline{0} \to 0$

$g(y) = \beta_0 + \beta(X)$ => Linear Regression

Converting Linear Reg to Logistic

$$p = \frac{\exp(\beta_0 + \beta(X))}{1 + \exp(\beta_0 + \beta(X))}$$

$= \exp(\beta_0 + \beta x)$ in numerator and denominator

$$p = \frac{1}{\frac{1}{\exp(\beta_0 + \beta x)} + 1}$$

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta x)}}$$ → Linear Regression

Sigmoid Function

$Y = \beta_0 + \beta x$

$$p = \frac{1}{1 + e^{-Y}}$$

$p(1 + e^{-y}) = 1$

$p + pe^{-y} = 1$

$$p e^{-y} = 1 - p$$

$$e^{-y} = \frac{1-p}{p}$$

$$e^{y} = \frac{p}{1-p}$$

$p \Rightarrow$ probability

$1-p = q$

$$\frac{p}{q} \Rightarrow \text{Odds } (0, \infty)$$

$$e^{y} = \text{odds} \quad (0, \infty)$$

$$e^{y} = \frac{p}{1-p}$$

apply log on both side

$$\log(e^{y}) = \log\left(\frac{p}{1-p}\right)$$

$$y = \log\left(\frac{p}{1-p}\right)$$

$$y = \log(\text{odds})$$

logits

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta(X)$$

Linear Regression => Logistic Regression

find p

$$\log\left(\frac{p}{1-p}\right) \Rightarrow \frac{p}{1-p} \Rightarrow \frac{p}{1-p} + 1$$

$$\frac{p}{1-p} = \frac{1}{1+e^{-(\beta_0+\beta x)}}$$

$$\frac{x}{x+1}$$

$$\frac{\frac{p}{1-p}}{\frac{p}{1-p}+\frac{1-p}{1-p}} = \frac{p}{\frac{1-p}{1-p}}$$

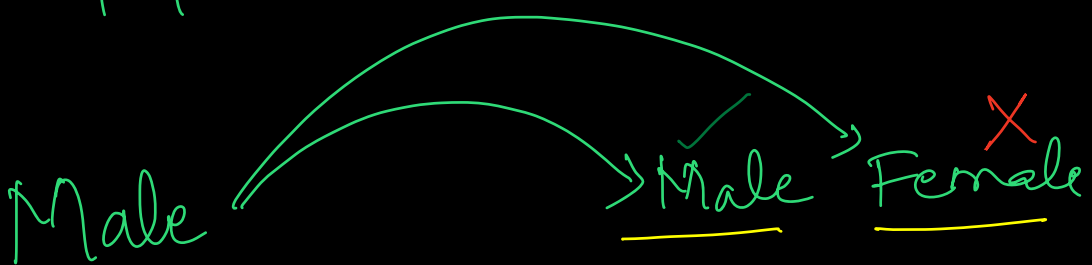
$$\frac{p}{1-p} + \frac{1-p}{(1)} = p$$

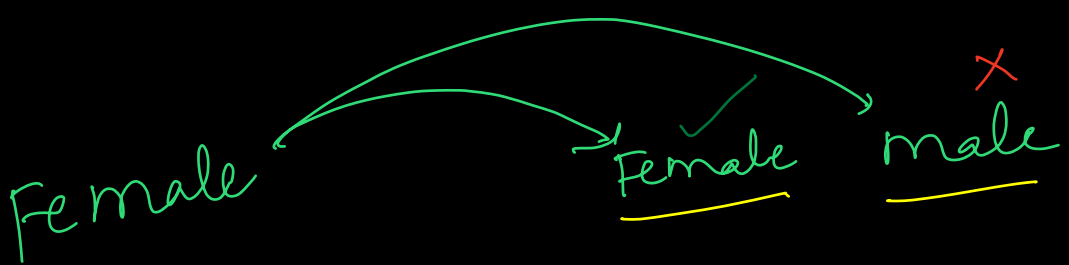
Actual Value

Male Female

Prediction Value

Male Female





4 possibilities

Confusion Matrix

Predicted Value	Actual Value: positive Male	Actual Value: neg Female
Male, positive	True Positive	Fale Positive (wrongly Predicting Negative as Positive)
Female Negative	Fale Negative	True Negative

Confusion Matrix

Predicted Value	Actual Value: positive Male (60)	Actual Value: neg Female (40)	
Male, positive (50)	True Positive 40	Fale Positive 10	50
Female Negative (50)	Fale Negative 20	True Negative 30	50
	60	40	100

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{40 + 30}{40 + 30 + 10 + 20} = \frac{70}{100}$$

$$= 0.7$$

$$= 70\%$$

Unbalanced

Confusion Matrix

		Actual Value: Positive — Cancer (10)	Actual Value: Neg — Not 9990	
Predicted Value	Cancer — positive	True Positive ①	Fake Positive 0	1
	Not — Negative	Fake Negative 9	True Negative 9990	9999
		10	9990	10000

10000 { 10 cancer
9990 not cancer

$$\text{Accuracy} \Rightarrow \frac{1 + 9990}{10000} = \frac{9991}{10000}$$

$$= 99.1\%$$

Accuracy is not always the right metric

1/10 = (10%)

Confusion Matrix

Predicted Value \ Actual Value	Male (positive)	Female (neg)
Male (positive)	True Positive	Fale Positive
Female (Negative)	Fale Negative	True Negative

Recall 100%.

Covid-19 Test ✓ Fever.

(Sensitivity) TPR
Recall
$\frac{TP}{TP+FN} = \frac{TP}{\text{Actual Positive}}$

$\boxed{\text{F1 score}} = \frac{2\,PR}{P+R}$

Precision
$\frac{TP}{TP+FP}$

Specificity = FPR
$= \frac{TN}{TN+FP}$

(1 - Specificity)

$\frac{TP}{\text{Predicted Positive}}$

Cancer Screening → Precision → No Cancer 100% accurate

Spam → Primary ✓

Primary → Spam ✓ ☹

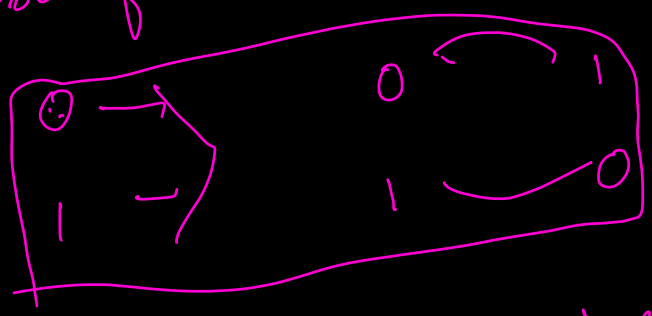
ROC

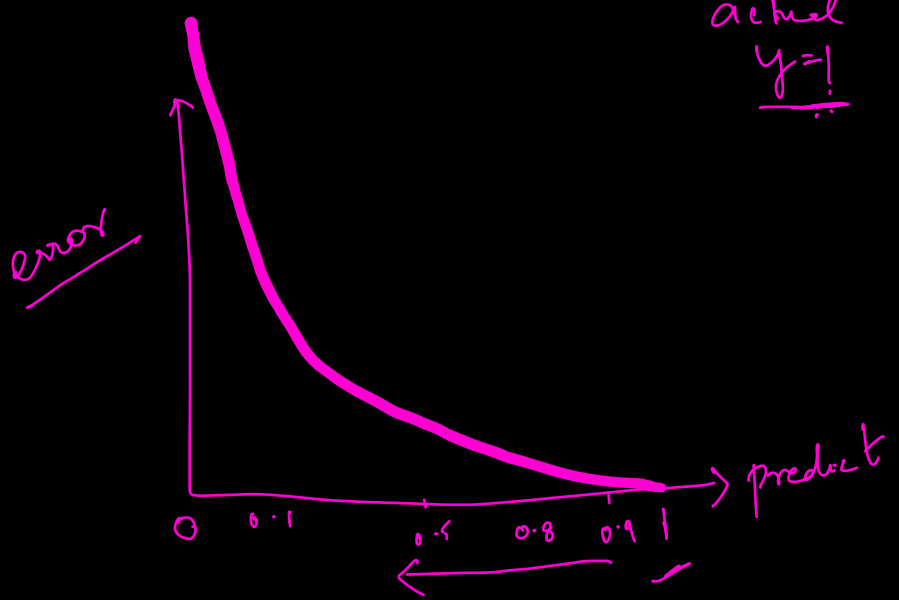
Probability → 0.5

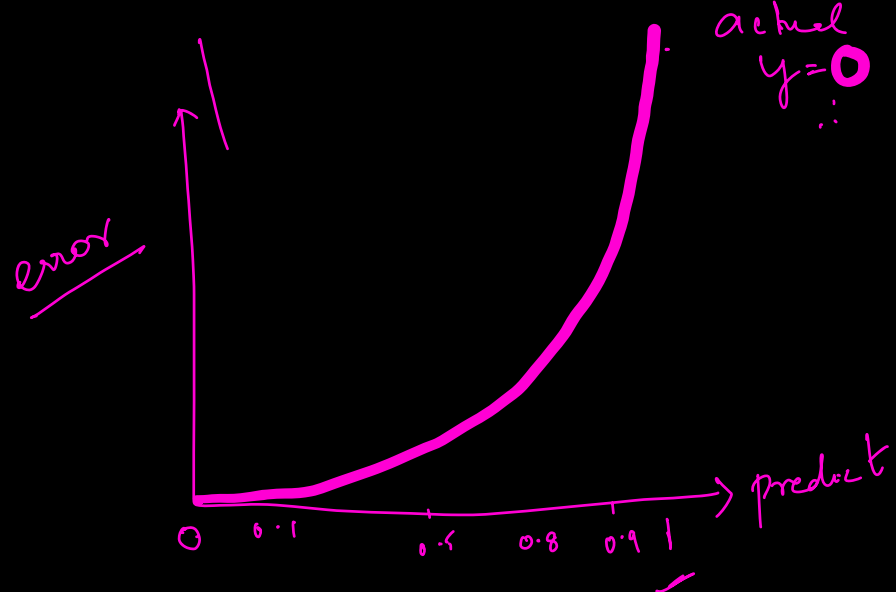
0.5

Gradient Descent

Cost function for Logistic







$y=1$

$\Rightarrow \hat{y}$

$1 \Rightarrow 0$

$0 \Rightarrow \infty$

$-\log(0) \Rightarrow \infty$

$-\log(1) = 0$