Logistic Regression

* Supervised ML
* Classification

Classification

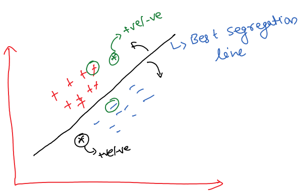
* we will categorize the o/p

E.g., Spam/Ham, Iris Dataset

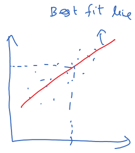
SL	SW	PL	PW	Species
–	–	–	–	Setosa
–	—	–	–	Versicolor
				Verginica

Classification

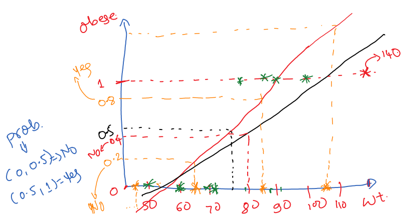
- Binary Classification
  - 2 Categories
  - E.g., Spam/Ham, Yes/No, 1/0, +ve/-ve
- Multiclass Classification
  - More than 2 category
  - E.g., Iris dataset.



Best fit line



Why we can't use Linear Regression for classification problem?

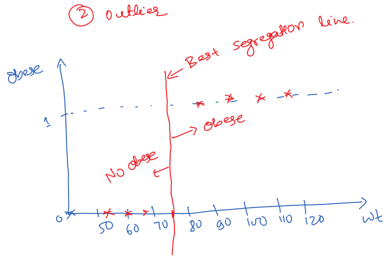


wt	obese
60	No
92	yes
72	No
78	yes
50	No
88	yes
70	No
100	yes
140	

1. Prob.
2. Outlier

yes = 1
No = 0

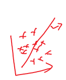
Best segregation line.



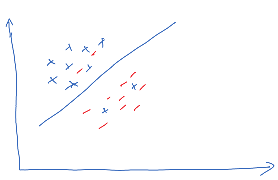
> 0.5 → 1
< 0.5 → 0

0.3 ⇒ 0
0.72 ⇒ 1

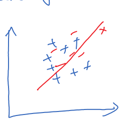
Continuous

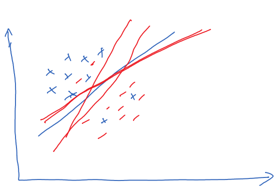


Logistic Regression



Log. Reg. → Features are linearly seperable

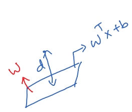


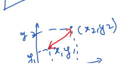


Linear Algebra

1. $d = \frac{w^T x + b}{\|w\|}$

Assumption – passing through origin





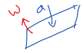
Linear Regression

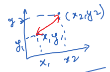
① $d = \frac{w^Tx + b}{\|w\|}$

Assumption

* Plane passing through origin

$w^Tx + \cancel{b}^{0} \Rightarrow w^Tx$





$\sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$

* 'w' is a unit vector.

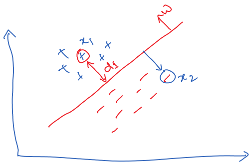
$\|w\| = 1$

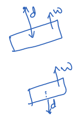
$\frac{w^Tx + \cancel{b}^{0}}{\cancel{\|w\|}^{1}}$

$d = w^Tx$

$y = \vec{w^T}x + b$

②





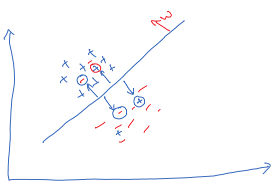
$d = w^Tx$

$d = w^Tx$

$d_1 = w^Tx_1$

$d_2 = -w^Tx_2$

*



$y = +ve \Rightarrow +1$

$= -ve \Rightarrow -1$

$y \in (+1, -1)$

$y * \text{distance} \Rightarrow y * w^Tx$

Case1: $y = +ve,\ d \to +ve$

$(+1) * (+ve) = +ve$

Case2: $y = -ve,\ d \to +ve$

$(-ve) * (+ve) \to -ve$

Case3: $y = -ve,\ d = -ve$

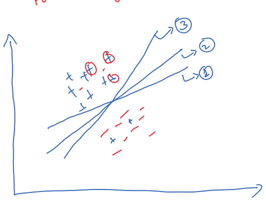
$(-ve) * (-ve) = +ve$

Case4: $y = +ve,\ d \to -ve$

$(+ve) * (-ve) \to -ve$

$y * w^Tx \to +ve$

* Whenever I am correctly classifying the point the product $(y * w^Tx)$ is '+ve' & whenever I am misclassifying the product $(y * w^Tx)$ is negative



$y * w^Tx \Rightarrow +ve \Rightarrow$ Correctly classified

$\Rightarrow -ve \Rightarrow$ misclassified

$\sum y * w^Tx$

① $\sum y * w^Tx = 30$ $\boxed{②}$ $\sum y * w^Tx = 88$ ③ $\sum y * w^Tx = 45$

Eg., 18, 19, 20, 8, 12, −8, −12, −3 = ☐ → } whichever answer is

= ☐ → } bigger, consider that as best segregation.

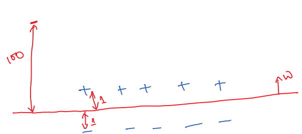
$\text{argmax} \sum y * w^Tx$

$\boxed{\sum y * w^Tx = 2}$

$\boxed{\sum y * w^Tx = -18}$

Case1

$\frac{1}{11}$ $\frac{10}{11}$

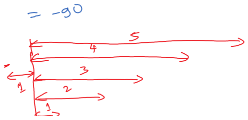


$\sum y * w^Tx = (+1)(+1) + (+1)(+1) + (+1)(+1) + (+1)(+1) + (+1)(+1) +$
$(-1)(-1) + (-1)(-1) + (-1)(-1) + (-1)(-1) + (-1)(-1) +$
$(-1)(+100)$

$= 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 - 100$

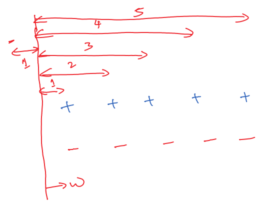
$= -90$

Case 2



$\frac{5}{11}, \frac{6}{11}$

Case 2



$\frac{5}{11}, \frac{6}{1}$

$$\sum y * w^T x = (+1)(+1) + (+1)(+2) + (+1)(+3) + (+1)(+4) + (+1)(+5) + (-1)(+1) + (-1)(+2) + (-1)(+3) + (-1)(+4) + (-1)(+5) + (-1)(-1)$$

$= 1$

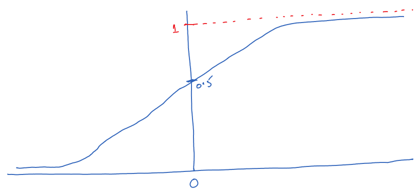
Case 1	Case 2
-90	< 1

$\sum y * w^T x$

* So whenever outlier is coming $\sum y * w^T x$ is failing.
So we need to modify the eq^n

Sigmoid function

$\sigma(y) = \frac{1}{1 + e^{-y}}$



$\sigma(x) = \frac{1}{1 + e^{-x}}$

$\text{argmin} \sum y * w^T x$

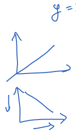
$\sigma(y * w^T x) = \frac{1}{1 + e^{-y * w^T x}}$

$\text{argmax} \sum \sigma(y * w^T x)$

$= \text{argmax} \sum \frac{1}{1 + e^{-y * w^T x}}$

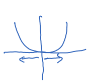
Monotonical functions.

$y = x$



$\Rightarrow$ Monotonically $\uparrow$ing

$\Rightarrow$ Monotonically $\downarrow$ing



$\Rightarrow y = x^2$

@ $x = 0$, $y = 0$

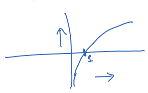
$\log x \Rightarrow$ Used to make calculation easier.

$\log(ab) = \log(a) + \log(b)$

$\log(a^x) = x \log a$

$\log\left(\frac{a}{b}\right) = \log a - \log b$

$\log x \rightarrow$ is a monotonical function



* Wherever you apply monotonically function to any function u will get same minimum point.

$\text{argmax} \sum \frac{1}{1 + e^{-y * w^T x}}$

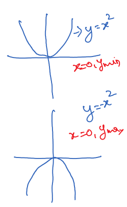
$\text{argmax} \sum \log\left(\frac{1}{1 + e^{-y * w^T x}}\right)$

$= \text{argmax} \sum \left(\cancel{\log 1}^{\,0} - \log(1 + e^{-y * w^T x})\right)$

$= \text{argmax} \sum \left(-\log(1 + e^{-y * w^T x})\right)$

$= -\text{argmax} \sum \log(1 + e^{-y * w^T x})$

$= \text{argmin} \sum \log(1 + e^{-y * w^T x})$



$\boxed{\sum y * w^T x}$

$\Downarrow$

$\sigma(\sum y * w^T x)$

$\text{argmin} \sum \log(1 + e^{-y * w^T x})$

$\hookrightarrow$ Best suggestion

$\sum \log(1 + e^{-y * w^T x}) \Rightarrow$ solve problem of outlier

↳ Best segu..

$\arg\min \sum \log(1 + e^{-y*w^Tx}) \Rightarrow$ solve problem of outliers

$\arg\min \sum \log(e^{-y*w^Tx})$

$\arg\min \sum -y w^Tx \log e$

$-\arg\min \sum y*w^Tx$

$\arg\max \sum y*w^Tx \Rightarrow$

EDA

Cont. → Cat. → Discrete

Univariate Analysis

X = Continuous

Hist plot, distplot

X = Categorical / Discrete

Count plot

Bivariate Analysis

X & Y → Continuous

Scatter plot

X → Categorical   Y → Cont.

Bar plot

X & Y → Categorical

cat plot.

Pregnancies → 0, 1, 2, 3.. 17 → Discrete → Categorical

Age → → Discrete → Continuous

1 km = 1000m

2 km, 500m

2, 500

Scaling

* scaling brings all values in the same range.

Pregnancies	Insulin
0	100
2	120
8	116
3	102
9	103

MinMax Scaler

[0, 1]

$$\text{Scale} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$x_{min} = 22$

$x_{max} = 32$

Age	Age
25	0.3
22	0
28	0.6
23	0.1
29	0.7
30	0.8
32	1

Scale

$\frac{25-22}{32-22} = \frac{3}{10} = 0.3$

$\frac{22-22}{32-22} = \frac{0}{10} = 0$

$\frac{28-22}{10} = \frac{6}{10} = 0.6$

$\frac{23-22}{10} = \frac{1}{10} = 0.1$

$\frac{29-22}{10} = \frac{7}{10} = 0.7$

$\frac{30-22}{10} = \frac{8}{10} = 0.8$

$\frac{32-22}{10} = \frac{10}{10} = 1$

min → 0 , max → 32

[0, 1]

Standard Scaler

$$\text{Scale} = \frac{x - \mu}{\sigma}$$

[-3 to +3]

Standard ND

97.3%

$\mu = 27, \sigma = 4$ (say)

Age	Age
25	-0.5
22	-1.25
28	.
23	.
29	.
30	.
32	

$\frac{25-27}{4} = \frac{-2}{4} = -0.5$

$\frac{22-27}{4} = \frac{-5}{4} = -1.25$

Evaluation Metrics

1, 0

200 → 160, 40

① Accuracy Score

$$Acc = \frac{\text{No. of correct prediction}}{\text{Total number of prediction}}$$

Eg., $acc = \frac{\cancel{160}^{4}}{\cancel{200}_{5}} = \frac{4}{5} = 0.8 = 80\%$

* Accuracy score is not good evaluation metric for imbalance data.

Balanced data v/s Imbalanced data

Eg., 1000 → 1, 0

	1 : 0	
50 : 50	(500 : 500)	→ Perfectly Balanced data
60 : 40	(600 : 400)	Balanced data
70 : 30	(700 : 300)	Balanced data
80 : 20	(800 : 200)	Imbalanced data
90 : 10	(900 : 100)	Imbalanced data

8 → 5, 3

Target	Pred
1	1 ✓
0	0 ✓
0	1 ✗
1	1 ✓
1	0 ✗
1	1 ✓
0	1 ✗
1	1 ✓

$Acc = \frac{5}{8} = 0.625$

8 → 7, 1

Target	Pred
1	1
1	1
1	1
1	1
$\boxed{0}$	→ 1
1	1
1	1
1	1

$acc = \frac{7}{8} = 0.875 = 87.5\%$

Confusion Matrix

		Actual	
		+ve	-ve
Pred	+ve	TP	FP
	-ve	FN	TN

2×2

TP → True positive
FP → False positive
TN → True Negative
FN → False Negative

TP
→ What is ur prediction
↳ Are u correct?

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

② Precision score

$$Pr = \frac{TP}{TP + FP}$$

FP↓

Out of total '+ve' prediction by model how many are actual '+ve'

		Actual	
		+ve	-ve
Pred	+ve	TP	FP
	-ve	FN	TN

→ Total +ve prediction by model

2×2

② Recall Score

$$Re = \frac{TP}{TP + FN}$$

FN↓

		Actual	
		+ve	-ve
	+ve	TP	FP

$$Re = \frac{TP}{TP+FN} \qquad FN\downarrow$$

Out of total actual '+ve' how many are actually +ve

		Actual +ve	Actual -ve
pred	+ve	TP	FP
	-ve	FN	TN

2×2

Total actual +ve

E.g., Spam / Ham
(1) (0)

- S → S
- H → H
- S → H
- H → S ✓ (0 → 1) FP↓

Cancer Yes / No
(1) (0)

- Yes → Yes
- No → No
- Yes → No ✓ (1 → 0) FN↓
- No → Yes

$$Pr = \frac{TP}{TP+FP\downarrow}\downarrow \uparrow$$

$$Re = \frac{TP}{TP+FN}\downarrow \uparrow$$

③ F1-score.

$$F1 = \frac{2\, Pr * Re}{Pr + Re}$$

FP↓ FN↓

Confusion Matrix

- Acc
- Pr
- Re
- F1

E.g., Olympics

- G → G
- B → B
- G → B
- B → G

E.g.,

200 → 150 (+ve), 50 (-ve)

150 (+ve) → 120 (+ve) (TP), 30 (-ve) (FN)

50 (-ve) → 20 (-ve) (TN), 30 (+ve) (FP)

		actual +ve	actual -ve	
Pred	+ve	120 TP	30 FP	150
	-ve	30 FN	20 TN	50
		150	50	200

$$acc = \frac{TP+TN}{TP+FP+TN+FP} = \frac{120+20}{200} = \frac{\cancel{140}^{7}}{\cancel{200}_{10}} = 0.7\% = 70\%$$

$$Pr = \frac{TP}{TP+FP} = \frac{120}{120+30} = \frac{120}{150} = 0.8 = 80\%$$

$$Re = \frac{TP}{TP+FN} = \frac{120}{120+30} = \frac{120}{150} = 0.8 = 80\%$$

$$F_1 = \frac{2 \times Pr \times Re}{Pr + Re} = \frac{2 \times 0.8 \times 0.8}{0.8 + 0.8} = 0.8 = 80\%$$

*

$$TPR = \frac{TP}{TP+FN}$$

$$TNR = \frac{TN}{TN+FP}$$

$$FPR = \frac{FP}{\quad}$$

TP	FP
FN	TN

$TNR = \frac{TN}{TN+FP}$

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

$FNR = \frac{FN}{FN+TP}$

TP	FP
FN	TN

Acc
Pr.
Re.
F1
ROC-AUC

MSE
RMSE
MAE
R^2
adj R^2

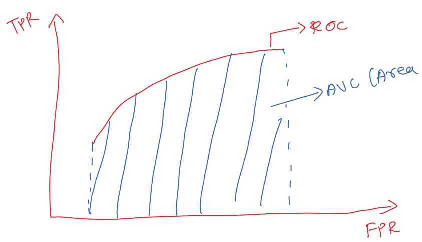
⑤ ROC - AUC curve.

ROC → Receiver operating characteristic.

→ This is the graph b/w TPR & FNR

>0.5 → 1
<0.5 → 0

Target	prob	Threshold = 0.5	Thr = 0.3	Thr = 0.7
1	0.6	1	1	0
0	0.3	0	1	0
0	0.4	0	1	0
1	0.8	1	1	1
1	0.7	1	1	1
0	0.2	0	0	0
1	0.9	1	1	1
		⇓ TPR & FPR	⇓ TPR & FPR	⇓ TPR & FPR



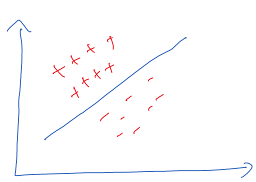
* ROC-AUC curve is used to compare different algorithm.
* Whichever algorith gives highest ans, consider that one as best model for ur dataset

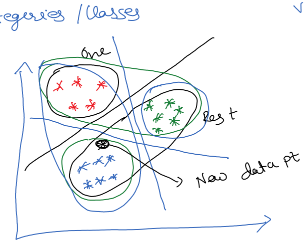
Eg., $ROC\text{-}AUC_{LR} = 0.7$
$ROC\text{-}AUC_{KNN} = 0.78$
$ROC\text{-}AUC_{DT} = 0.9$ ← Best algorithm

Multiclass Classification

* More than 2 categories / classes

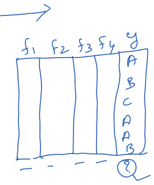
setosa
versicolor
verginica





Multiclass → One v/s Rest (OVR)

No. of classes = No. of model



M1	M2	M3
A v/s Rest	B v/s Rest	C v/s Rest
Prob. of A	Prob B	Prob C.
⇓ 0.3	⇓ 0.5	⇓ 0.2.

$\underbrace{0.3 \quad 0.5 \quad 0.2}_{\text{Class} - B}$