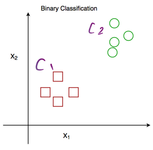
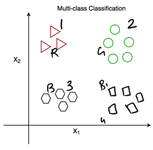
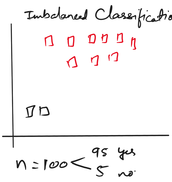
Classification: to classify based on features
y = continuous → Regression
y = categorical → Classification.





Imbalanced Classification.



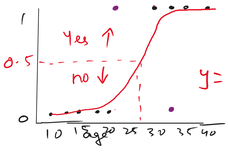
n = 100 < 95 yes, 5 no

Types of Classification Algorithm:

1. Logistic Regression — Sigmoid
2. Naive Bayes — Probability
3. K Nearest Neighbour — Distance metrics
4. Decision Tree — Condition
5. Support Vector Machine — Vectors
6. Random Forest — DT

Logistic Regression: binary classification.
↳ two 0/1, n/y, +1/-, etc.

Sigmoid function.



age = 26 ?

$$y = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$ → range - 0 to 1

assumption: y = 0.7, 0.3, 0.5

x	y
10	0.3
15	0.4
20	0.45
25	0.49
30	0.51
35	0.7

β_0 β_1

Threshold	when to use
0.1 - 0.4	chance of risk
0.5	balance
0.6 - 0.9	high confidence

Types of Logistic Regression

1. Binary Logistic Regression: 2 class - 0,1 ; y/n
2. Multinomial Logistic Regression: 3 or more - R, B, B
3. Ordinal logistic Regression: 3 or more - +ve, nav, -ve

1. Confusion Matrix

		Predicted: -ve	Predicted: +ve
Actual Real	-ve	TN True Negative	FP False Positive
	+ve	FN False Negative	TP True Positive

2. Accuracy Score: $\frac{TP+TN}{TP+FP+TN+FN}$ — how well the model is performing

3. Precision: $\frac{TP}{TP+FP}$ — +ve prediction.

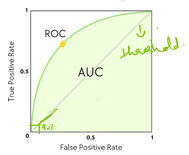
4. Recall: $\frac{TP}{TP+FN}$ — Correct +ve prediction

5. F1 Score: $2 \times \frac{Precision \times Recall}{Precision + Recall}$ — Note: F1 ≈ accuracy

6. AUC-ROC: Area Under Curve - Receiver Operating characteristics.
ROC-AUC

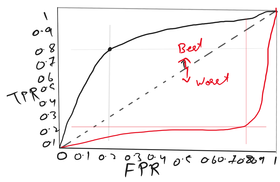
more the area under curve better the model is

points used to make area.



FPR (False Positive Rate) = FP/FP+TN

TPR = True Positive Rate = $\frac{TP}{TP+FN}$



TPR	FPR
0.8	0.2
0.2	0.8