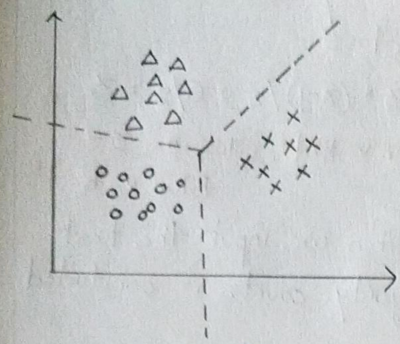
MULTI - CLASS CLASSIFICATION — One Vs ALL, (One vs Rest) One Vs One.

Multi-class Classification —



Multi class classification is the classification technique that allow us to categorize the test data into multiple class labels present in trained data as a model prediction.

2 types →

i) One vs All (One vs Rest) : N class instances the N binary classifier model.

ii) One vs One : N class instances then $\boxed{N*(N-1)/2}$ binary classifier models

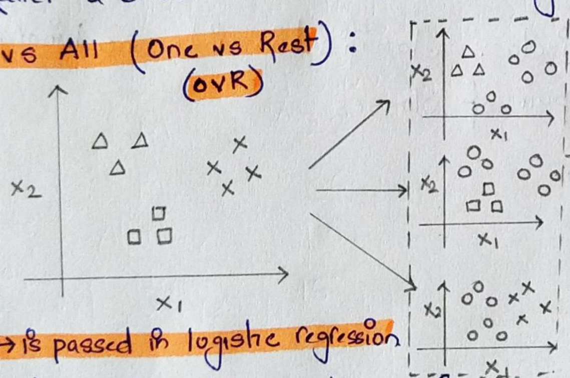
For example, consider a multi class classification problem with four classes: 'red', 'blue', 'green', and 'yellow'. This could be divided into six binary classification datasets as follows:

Classification 1: red vs blue.
Classification 2: red vs green
Classification 3: red vs yellow
Classification 4: blue vs green
Classification 5: blue vs yellow
Classification 6: green vs yellow.

$= (N*(N-1))/2$
$= (4*(4-1))/2$
$= (4*3)/2$
$= 12/2$
$= 6$.

This approach require each model predicts a probability of class. The argmax of these score (class index with largest score) is then used to predict a class.

One vs All (One vs Rest) : (OVR)



- We will generate same number of classifier as the class labels are present in the dataset.
- So here we have to create three classifier for three respective class
- we will have 3 dataset and have 3 sub models and based on output it give / assign class [gives probability]

OVR → is passed in logistic regression

Feature			Class
x_1	x_2	x_3	G
x_4	x_5	x_6	B
x_7	x_8	x_9	R
x_{10}	x_{11}	x_{12}	G
x_{13}	x_{14}	x_{15}	B
x_{16}	x_{17}	x_{18}	R

Overall Dataset

Feature			Green
x_1	x_2	x_3	+1
x_4	x_5	x_6	-1
x_7	x_8	x_9	-1
x_{10}	x_{11}	x_{12}	+1
x_{13}	x_{14}	x_{15}	-1
x_{16}	x_{17}	x_{18}	-1

Training set 1

Feature			Blue
x_1	x_2	x_3	-1
x_4	x_5	x_6	+1
x_7	x_8	x_9	-1
x_{10}	x_{11}	x_{12}	-1
x_{13}	x_{14}	x_{15}	+1
x_{16}	x_{17}	x_{18}	-1

Training set 2

Feature			Red
x_1	x_2	x_3	-1
x_4	x_5	x_6	-1
x_7	x_8	x_9	+1
x_{10}	x_{11}	x_{12}	-1
x_{13}	x_{14}	x_{15}	-1
x_{16}	x_{17}	x_{18}	+1

Training set 3

- Suppose a test datapoint (y_1, y_2, y_3) is passed suppose for Model 1, we get p(green)=0.9 the Model 2, P(Blue) = 0.04 and Model 3, P(Red) = -0.04. So based on probability we will assign highest probability = 0.9 to (green).

One Vs One (OvO) →

- For N class instances dataset, we generate $N*(N-1)/2$ binary classifier models.
- Using this classification approach, we split the primary dataset into one dataset for each class opposite to every other class.
- Last example have Green, Blue and Red $(N=3)$, $= 3*(3-1)/2 = 6/2 = 3$
  Classifier 1 = Green vs Blue, **Classifier 2 = Green vs Red**, **Classifier 3 = Blue vs Red**.
- Each binary classifier predicts one class label. When we input the test data to the classifier, then the model with majority counts is concluded as result.

One vs Rest / One vs All possible downside –

- It require one model for each class. This could be an issue for large dataset where we have many classes.

SUMMARY →

- One vs All / One vs Rest, splits a multi-class classification into one binary classification problem per class.
- One vs One strategy splits a multi-class classification into one binary classification problem per each pair of classes