

Regression in Detail

- Simple Linear Regression (Univariate Regression)
- multiple linear Regression (Multivariate Regression)
- Polynomial Linear Regression

Errors are calc. in Regression

- Mean Absolute Error
- Mean Squared Error

Get Best features

- ✓ - OLS (Backward elimination)
- ✓ - Gradient Descent (neural net - deep learning)

Label is always continuous numeric.

Deployment of model using pickle.

CLASSIFICATION

Supervised Learning features & label

Regression

features \Rightarrow n columns

label \Rightarrow 1 column

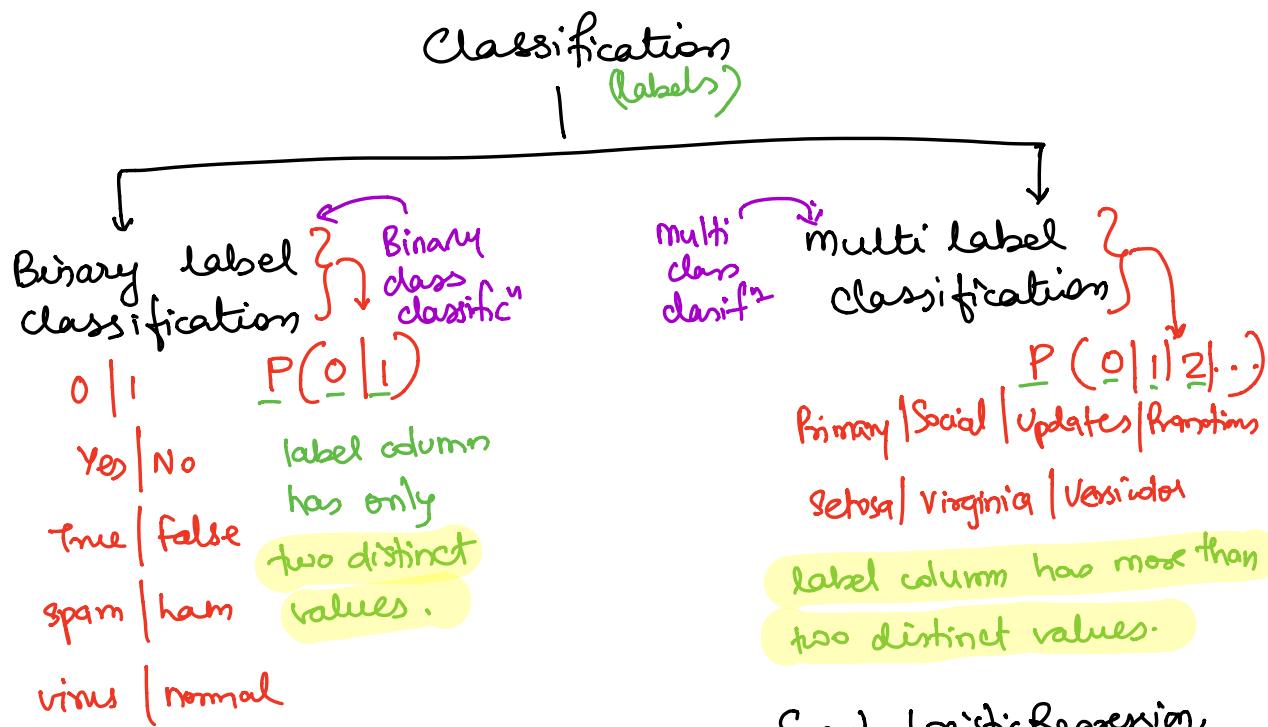
Cond": label must be
a continuous numeric
value

Classification

features \Rightarrow n columns

label \Rightarrow 1 columns

Cond": label must be
strictly categorical data
gender, M/F Y/N
T/F



All algs related to classification are supported.

Except Logistic Regression, all algos related to classification are supported

Classification Algs

- ① Logistic Regression → binary class
- ② K-Nearest Neighbourhood → binary class | multi class
- ③ Decision Tree Classification → binary class | multi class
- ④ Random Forest Classification → binary class | multi class
- ⑤ Naive Bayes Classification → binary class | multi class

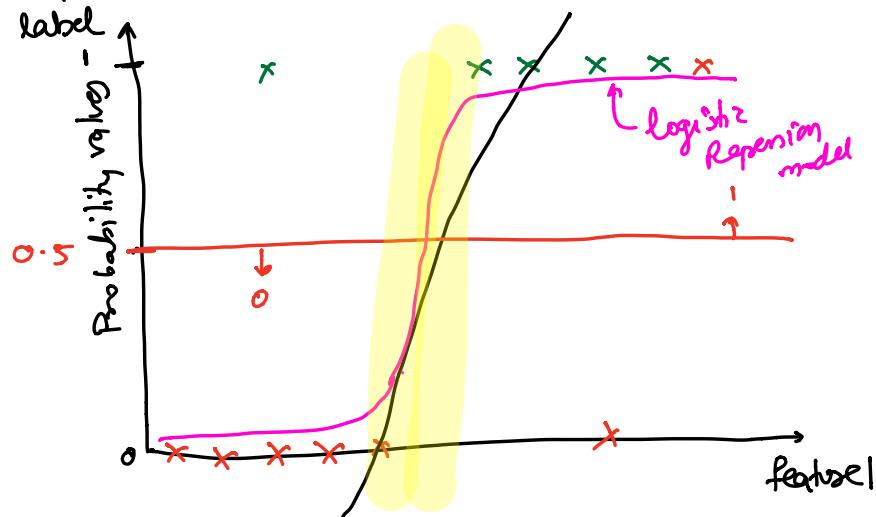
Logistic Regression

This algorithm is used for binary-class classification.

Note:- Even though we say it supports only BINARY CLASSIFICATION, technically you can implement the same in MULTI-CLASS CLASSIFICATION. However, either of the following 3 parameters will get affected when implementing LR.

- ① Accuracy
- ② Precision
- ③ Recall
- ④ F1 score

Why name is Logistic REGRESSION?



Add Sigmoid fn to linear regression.

Your data is
linearly
separable.

History:
Linear
Regression
 y = Probability
values between
0 and 1

Threshold factor
 $0 - 0.5 \rightarrow 0$
 $0.5 - 1 \rightarrow 1$
freq distribution

Linear Regression $\Rightarrow y = mx + b$ ————— ①

Sigmoid fⁿ $\Rightarrow P = \frac{1}{1 + e^{-y}}$ (Standard formula)

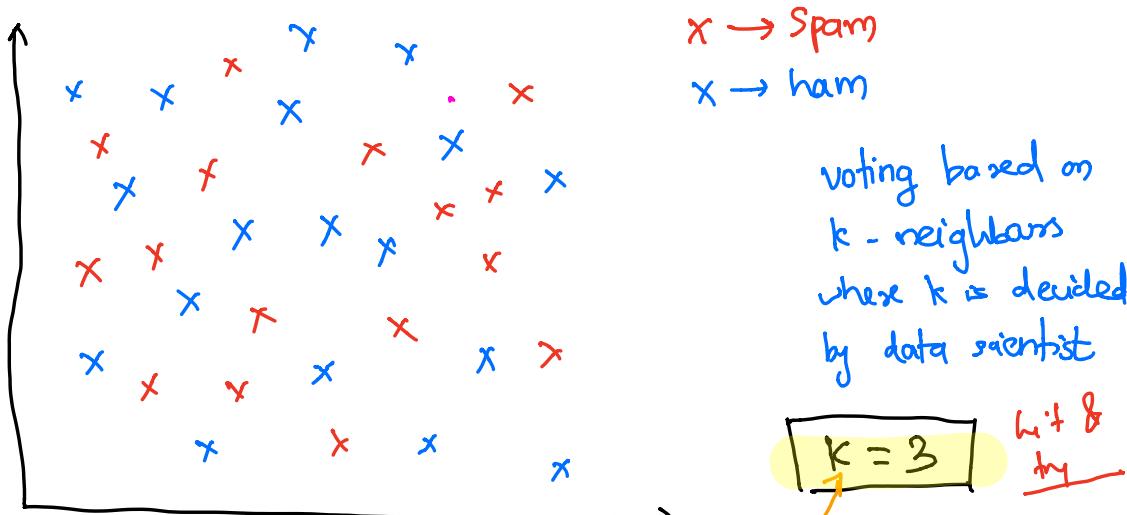
Substitute 'y' from ①

$$P = \frac{1}{1 + e^{-(mx+b)}}$$

$$P = \frac{1}{1 + e^{-mx - b}}$$

... Logistic Regression
equation
(intercept
slope)

K-NN (K-nearest neighbours)



To calc the distance betn two points:

① Euclidean distance $\rightarrow \oplus$

② Manhattan distance

\rightarrow blue = 2
red = 1

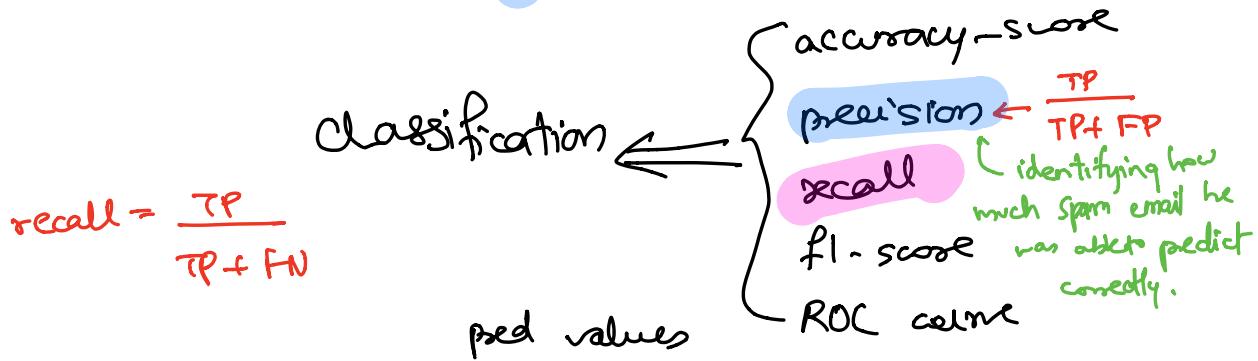
number of nearest pt

It is suggested to use odd nos for 'k' value.

Confusion Matrix
(Binary label)

		predicted values	
		spam	not spam
true values	spam	TP 52	FN 0
	not spam	FP 28	TN 0

accuracy = $\frac{52+0}{52+0+28+0} = 65\%$
 $\approx 65\%$
 $= \text{precision}$.



Setosa

Versicolor

True values

False values

True values

False values

True values

False values

		Setosa	Versicolor	Virginica
true values	Setosa	True values	False values	False
	Versicolor	False	True values	False
False values	Virginica	False	False	True values

Can we define
 $TP \Leftarrow$
 $TN ?$

what is accuracy?

Its the ratio of positive predicted values to the total values.

	Setosa	Versicolor	Virginica
Setosa	9	0	0
Versicolor	0	7	1
Virginica	0	0	13

$$\text{accuracy} = \frac{9+7+13}{9+0+0+0+7+1+0+0+13} = \frac{29}{30} = ?$$

I want to check precision for

$$\text{Virginica} \Rightarrow \frac{13}{13+0+1} = \frac{13}{14} = 92.85\%$$

$$\text{Setosa} \Rightarrow \frac{9}{9} = 1 = 100\%$$

$$\text{Versicolor} \Rightarrow \frac{7}{7} = 1 = 100\%$$

precision

= w.r.t to what you are looking for.

recall

= Setosa $\Rightarrow 100\%$.

Versicolor $\Rightarrow 87.5\%$.

Virginica $\Rightarrow 100\%$.

How to calculate?

Confusion Matrix

F1 score
weighted average of Precision & Recall.

$$\frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})}$$

Harmonic Mean

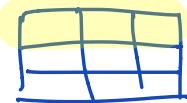
Accuracy
(Model wide)



Precision
(Label wide)



Recall
(Label wide)



When to use which metric for CLASSIFICATION?

- ① Check whether the supplied dataset is a balanced dataset or not.

Balanced dataset means, your label weightage must be same throughout the dataset

- ② If the dataset is BALANCED, go for ACCURACY
- ③ If dataset is unbalanced, check for domain requirement and based on that check for precision or recall or f1 score.