## variables

random variable

categorical or numerical

### univariate

### multivariate

### categorical

qualitative

nominal variables

no order of precedence

country

ordinal variables

some order associated

(week days)

### numerical

quantitative

discrete

years of schooling

number of items

continuous

age

weight of a human being

altitude of a flying bird

## Machine Learning

### Supervised Learning

labelled data

100 matches

80 matches → training data

20 matches → testing data

build a model

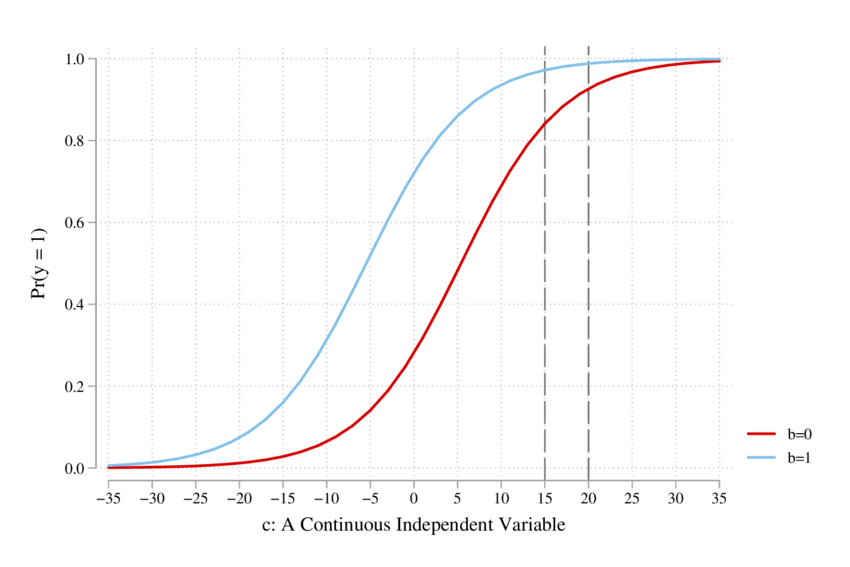
apply it on the

#### Linear Regression



#### Logistic regression





#### KNN



### SVM



### Unsupervised Learning

#### K Means Clustering



#### DBSCAN clustering

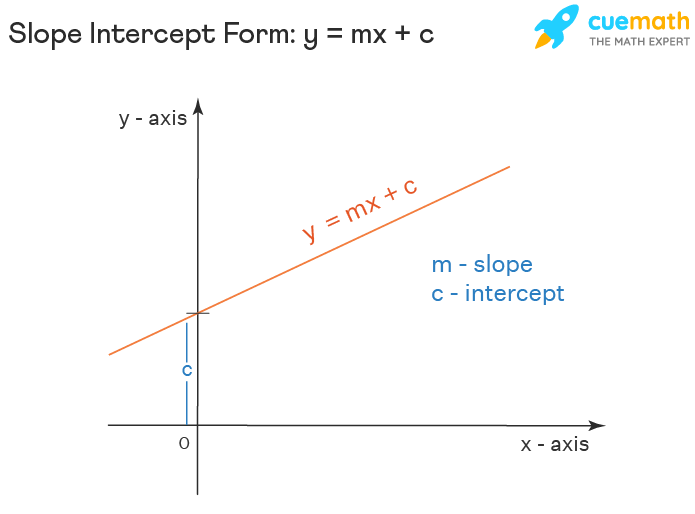


## Stages

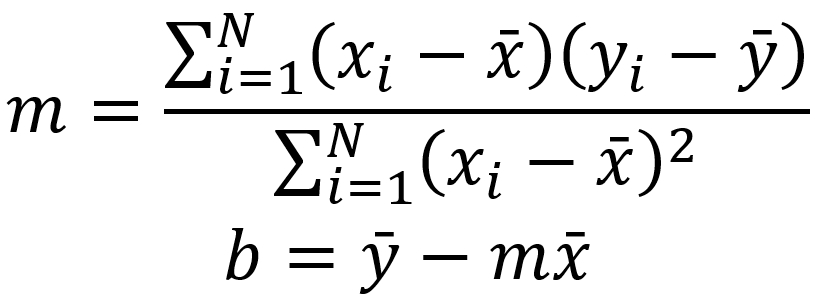
1. Data preparation
2. Model architecture specification
3. Design & execute of training process
4. evaluation of trained model

## Linear Regression

y = mx + b



Least Square method



## knn (k-nearest neighbour )

supervised

classification

k = num of similar properties

how similar

similarity metric

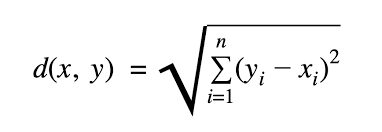


### Euclidean distance

differences = (x1 - y1) + (x2 + y2) + (x3 + y3) + (x4 + y4)

squared differences = (x1 - y1)2 + (x2 + y2)2 + (x3 + y3)2 + (x4 + y4)2

euclidean distance = square root of [ (x1 - y1)2 + (x2 + y2)2 + (x3 + y3)2 + (x4 + y4)2 ]



## Decision Tree

ID3

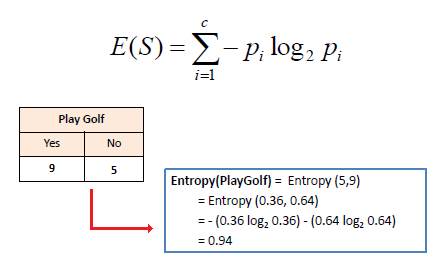
CART

random forests

### ID3

#### Entropy

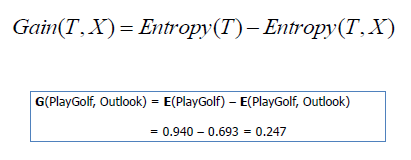
average rate at which information is provided by a random source of data



#### Information gain

amount of information gained about a random variable by observing another

random variable



Outlook = Overcast

play happens

Outlook = sunny

Humidity = high

no play

humidity = normal

play happens

Outlook = rain

windy = strong

no play

windy = normal

play happens

| **7** | sunny | mild | high | Weak | N |
| --- | --- | --- | --- | --- | --- |