

classification

it means dividing data to it similar properties making classes of them and it divide the data through decision boundary.

classification

Binary classification

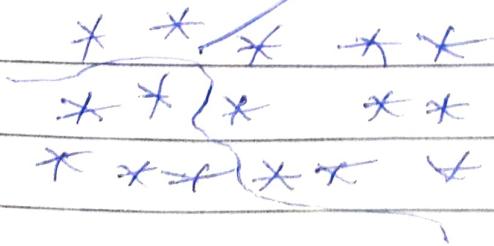
multi-classification



Binary classification.

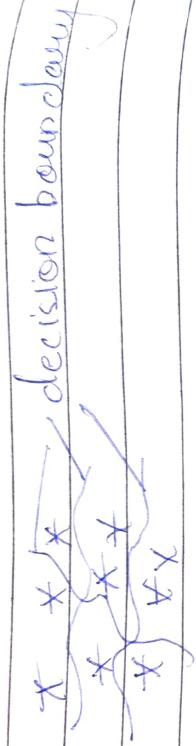
here classification can be done in two only eventually the model will draw decision boundary

decision boundary.



* Multiclassification

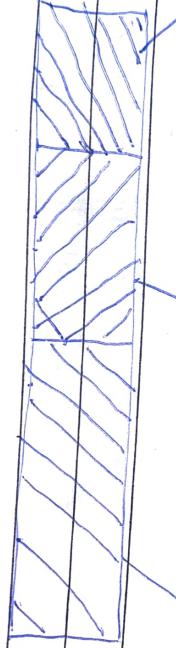
it means the model make multiple classes.



- * PPS → for categorical data finding correlation between them.
- * Correlation matrix → for variable data finding correlation between them.

Data split for model building

we don't always use 100% of data for building a model instead we will take 80 to 90% data for training 20% for test if accuracy is coming or not for validation will split more



- for training for validating for testing
- the data data data

Holdout method problem

overfitting issue in model building suppose we train the model and accuracy come around 80 to 90% but when we test on test data the accuracy came 40 to 60% it means it drastically drop it not learning proper. ceneses means it doesn't too much, but not understanding. if we use a called underfitting. so solution is to perform different evaluation methods.

* So when we should rely on model

= when accuracy is more than 70%

* in Supervised learning we have some algorithm:

for continuous variable
we will use logic behind linear regression

for discrete / categorical
we will use logic behind logistic regression

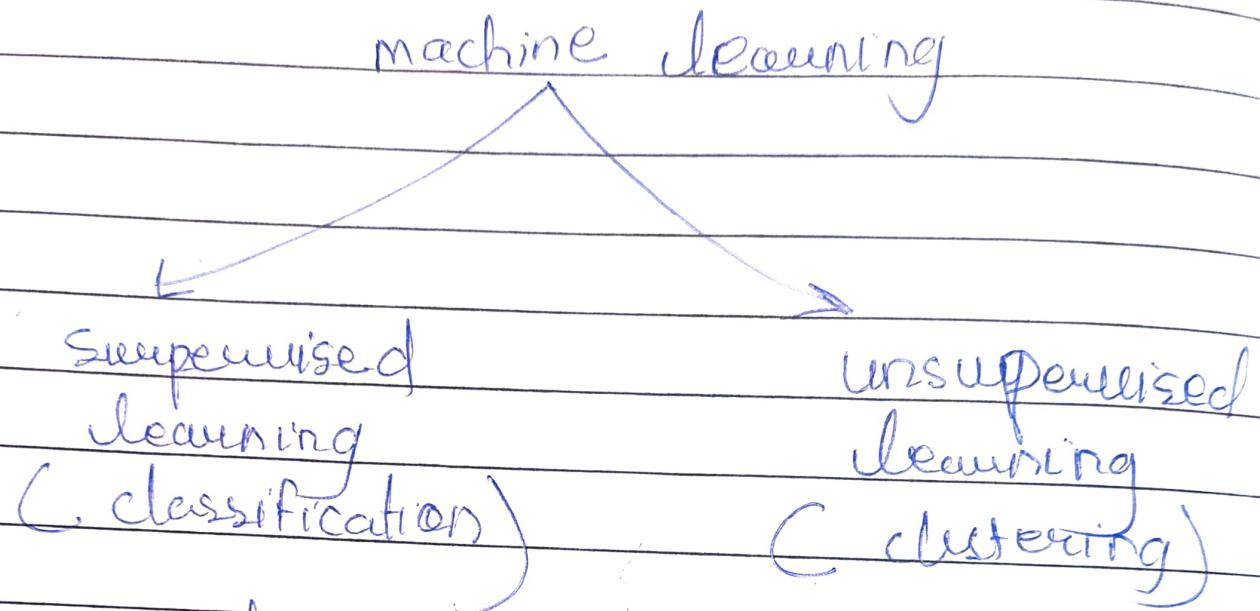
- a) decision tree
b) KNN
c) sum

and there many more but this is common used

* unsupervised here we prefer an

clustering Analysis

* then what is difference clustering and classification



example

NCC

LDA

VG

NN

KNN

Neural Networks

Support Vector Machine

Random Forest etc.

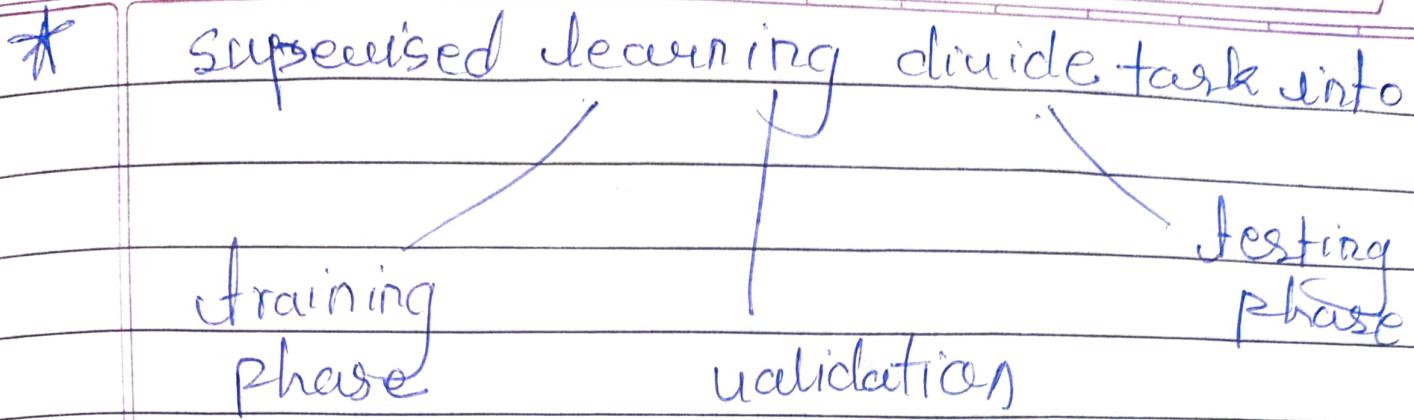
example

ICA

SOM

GRAs

Hierarchical
K-means etc.



Classification

* Discriminatory power of humans/machines to recognize objects is classification

clustering

* may not have prior information
* we want to arrange similar object together
* we want to arrange dissimilar object in different bags or group or clusters.

Hold out method.

To get the more good accuracy with less error rate. That's why we use evaluation methods for example going to split training set and test set for 20/10.

give data more for training in order to get more Accuracy.

and

If you give more data in test then you will know error rate.

and one of the evaluations method is hold out method.

✓. which Holdout method does us.

it splits the dataset into training and test datasets

but here should we split the baby

~~lets take~~ one example

~~Class~~

here what happens

you haven't given any classes of

~~training~~ = I to train the error rate become high that's why we use classification

+ = training

+ = test

+ = error rate

+ = high that's why we use classification

Sacrification which does is it simple
make equal parts of same size
after sacrifice the split should
be like this.

after stratification

PAGE No.

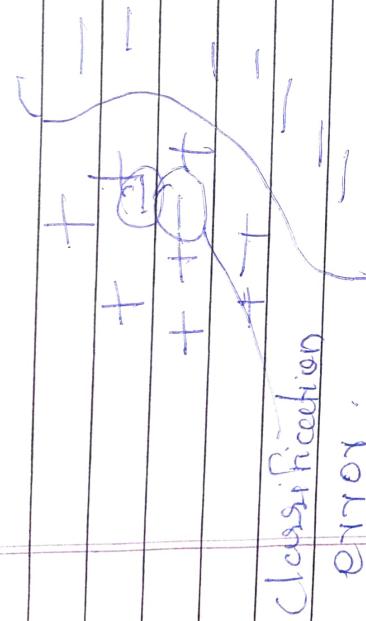
DATE / /

class

+ training
+ test
+ 2g

there are different methods also
Cross validation Bootstrap

Classification



To evaluate the accuracy of classification models we can use one technique known as confusion matrix. There are many techniques also.

The objective of classifications algorithms is to make purely homogeneous regions.

So we should optimize the decision boundary.

* classification : Step -

①

Give a collection of records -

②

Find a model a accurate model .

③ Goal :- Previously unseen records
should be assigned to classes accurately
as possible. (Test set)

*

Note :- when we use dimension
reduction or classification the
model accuracy is increase and PCA
is the best for classification.

* example of classification task

i) predicting tumour cells as benign
or malignant .

②

categorizing news stories as finance
weather, entertainment, sport
etc.

Classification techniques.

- 1) Decision tree
 - 2) Neural Networks
 - 3) Naive Bayes
 - 4) Support vector machine
 - 5) k- Nearest Neighbour
- many more