Lec 8 classification

Why classification ?

main idea : do label

for machine learning : let computer learn and distinguish the
result from clustering.

for classifing people : we can find whether a target person is
in a specific group or not.

for biology : the classification helps us do diagnosis, judge
normal and tumor.

what is classification ?

there are a number of attributes, one group is one of attributes.

How to do classification

training face : input the the data with lable

prediction face : predict the data without lable due to its attributes

K - nearest neighbors

training : label the data due to their contributes.

predicting : calculate the unlabeled record and return the closest class

running KNN : $\begin{cases} \text{Euclidean distance (contributes considered)} \\ \text{k (similar data considered)} \end{cases}$

① Normalization

② Compute distance between other data and target data

③ Identify k most similar data: k=2, so choose 2
  similar data to compare with target data
④ Find the label from these 2 data.

Clurstering vs Classification

	Clustering	Classification
Goal	Find similarity (clusters) in the data	Assign class to the new data
Data	Data without class	Training data with class and testing data without class
Classes	Unknown number of classes	Known number of classes
Output	The cluster index for each point	The class assignment of the testing data
Algorithm	One phase	Two phases (training and application)

{ unsupervised learning : analyse & cluster unlabelled data
  supervised learning : classify & predict labelled data