



kuala lumpur

school of ai

Machine Learning 101

K-Nearest Neighbour Theory

Supervised Machine Learning

- we are given a labelled dataset consisting of training observations (x,y) and would like to capture the relationship between x and y .

What is K- Nearest Neighbour

- k-NN is based on feature similarity: How closely out-of-sample features resemble our training set will determine how we classify a given point
- An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors

K-NN Algorithm Steps

1. Label train data according to the classes
2. Calculate distance of new sample to “train” data
 - a. Euclidean distance (popular choice)

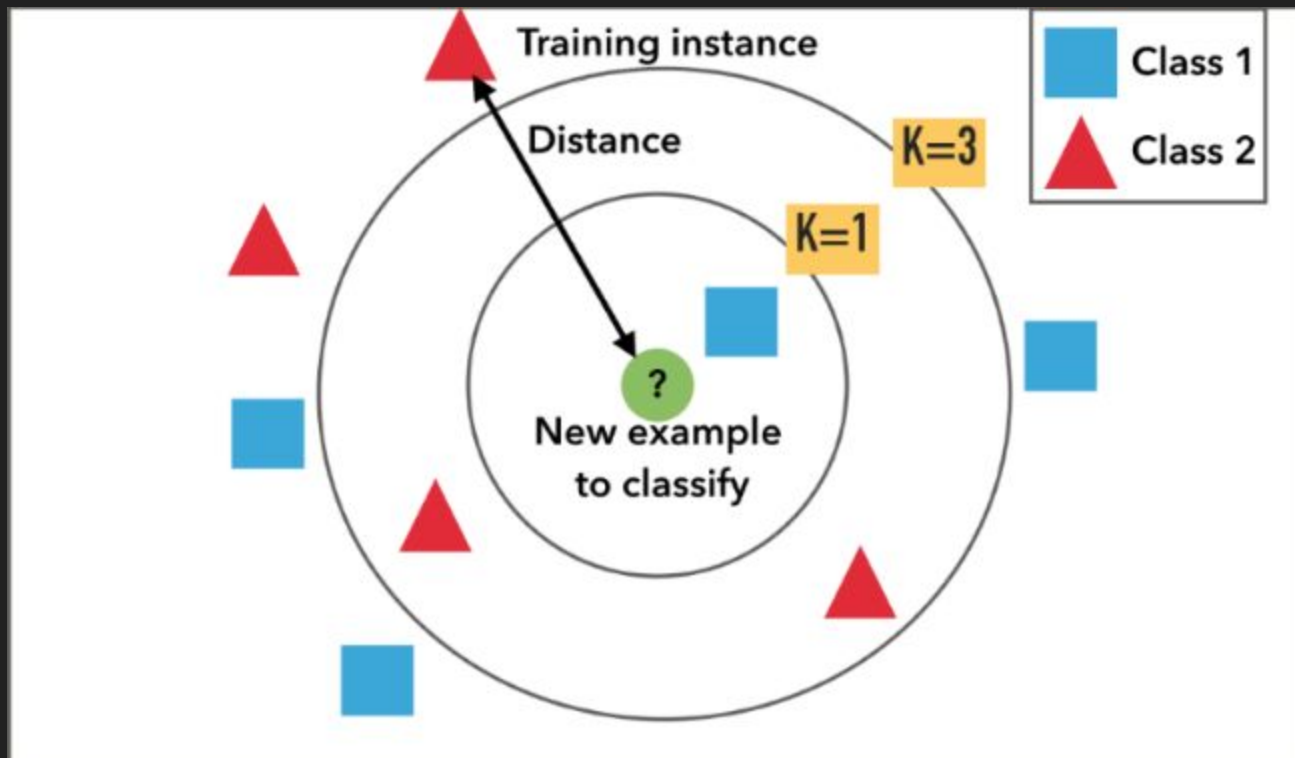
$$= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}.$$

3. Classify the new samples to the given classes

What is K in K-NN ??

- K here means the number of closest neighbour of the sample
- K is a hyper parameter that you have to set
- if $k = 1$ then the sample will be included in that class of the closest first neighbour.
- if $k = 3$ the algorithm will look at the 3 points closest to the sample. it will then see the 2 out of 3 points which are closest to it.



Extra Information

- Instance-based learning means that our algorithm doesn't explicitly learn a model. Instead, it chooses to memorize the training instances which are subsequently used as "knowledge" for the prediction phase. Concretely, this means that only when a query to our database is made (i.e. when we ask it to predict a label given an input), will the algorithm use the training instances to spit out an answer.
- ## non instance based is when the model would learn from the data and give answer based on learned results
- Non-parametric (it means that it does not make any assumptions on the underlying data distribution) , model structure is determined from the data

References

<https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7>

<https://sebastianraschka.com/faq/docs/lazy-knn.html>

<https://github.com/huseinzol05/Machine-Learning-Numpy/blob/master/clustering/K-nearest-neighbors/knn-iris.ipynb>

<https://kevinzakka.github.io/2016/07/13/k-nearest-neighbor/>

[https://en.wikipedia.org/wiki/Hyperparameter_\(machine_learning\)](https://en.wikipedia.org/wiki/Hyperparameter_(machine_learning))

<https://scikit-learn.org/stable/modules/neighbors.html>