

Lecture 5

Supervised Algorithms: KNN and SVM

By: Nazerke Sultanova

k-Nearest Neighbours

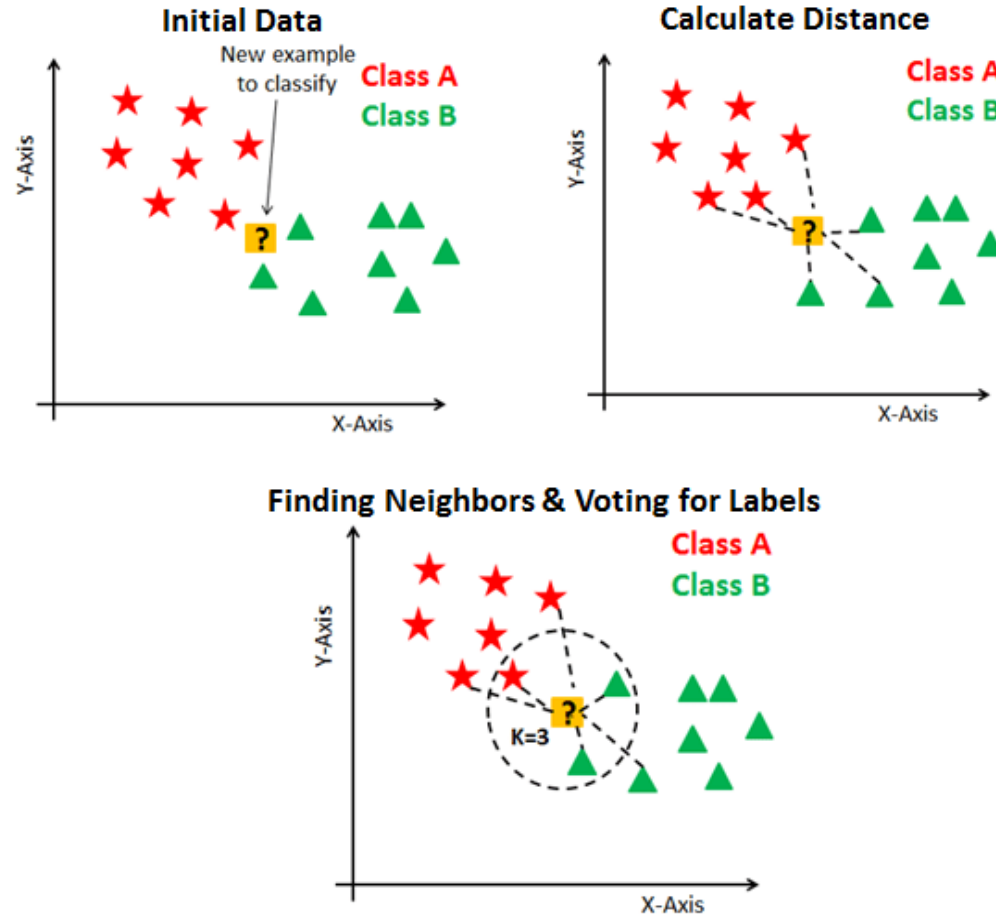
- KNN is a non-parametric and lazy learning algorithm.
- Non-parametric means that the model structure determined from the dataset.
- Lazy algorithm means that all training data used in the testing phase.
- This makes training faster and testing phase slower and costlier.

kNN Algorithm:

KNN has the following basic steps:

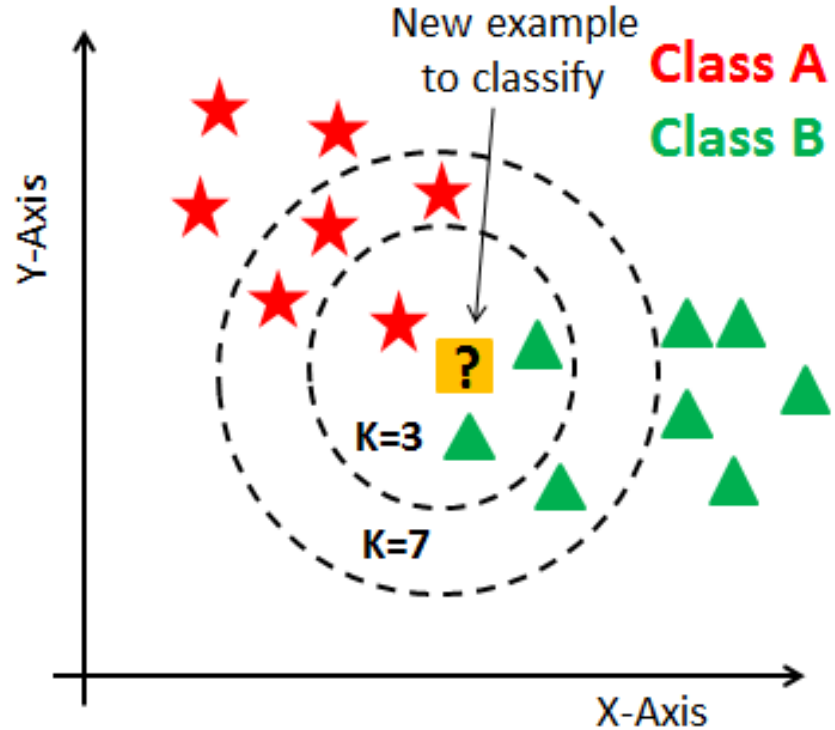
- Calculate distance
- Find closest neighbors
- Vote for labels

kNN Algorithm:



How to choose the optimal number of neighbors?

- Generally, Data scientists choose as an odd number if the number of classes is even.
- Also check by generating the model on different values of k and check their performance.



kNN advantages

- Classifier immediately adapts as we receive new training examples
- But computational complexity grows linearly with the number of samples

Distance metrics:

$$d(x^i, x^j) = \sqrt[p]{\sum_k [x_k^i - x_k^j]^2}$$

- Euclidean distance if we set the parameter $p = 2$
- Manhattan distance if we set the parameter $p = 1$

sklearn.neighbors.KNeighborsClassifier

```
class sklearn.neighbors. KNeighborsClassifier (n_neighbors=5, weights='uniform', algorithm='auto', leaf_size=30,  
p=2, metric='minkowski', metric_params=None, n_jobs=None, **kwargs) \[source\]
```

Classifier implementing the k-nearest neighbors vote.

Read more in the [User Guide](#).

Parameters: **n_neighbors** : *int, optional (default = 5)*

Number of neighbors to use by default for `kneighbors` queries.