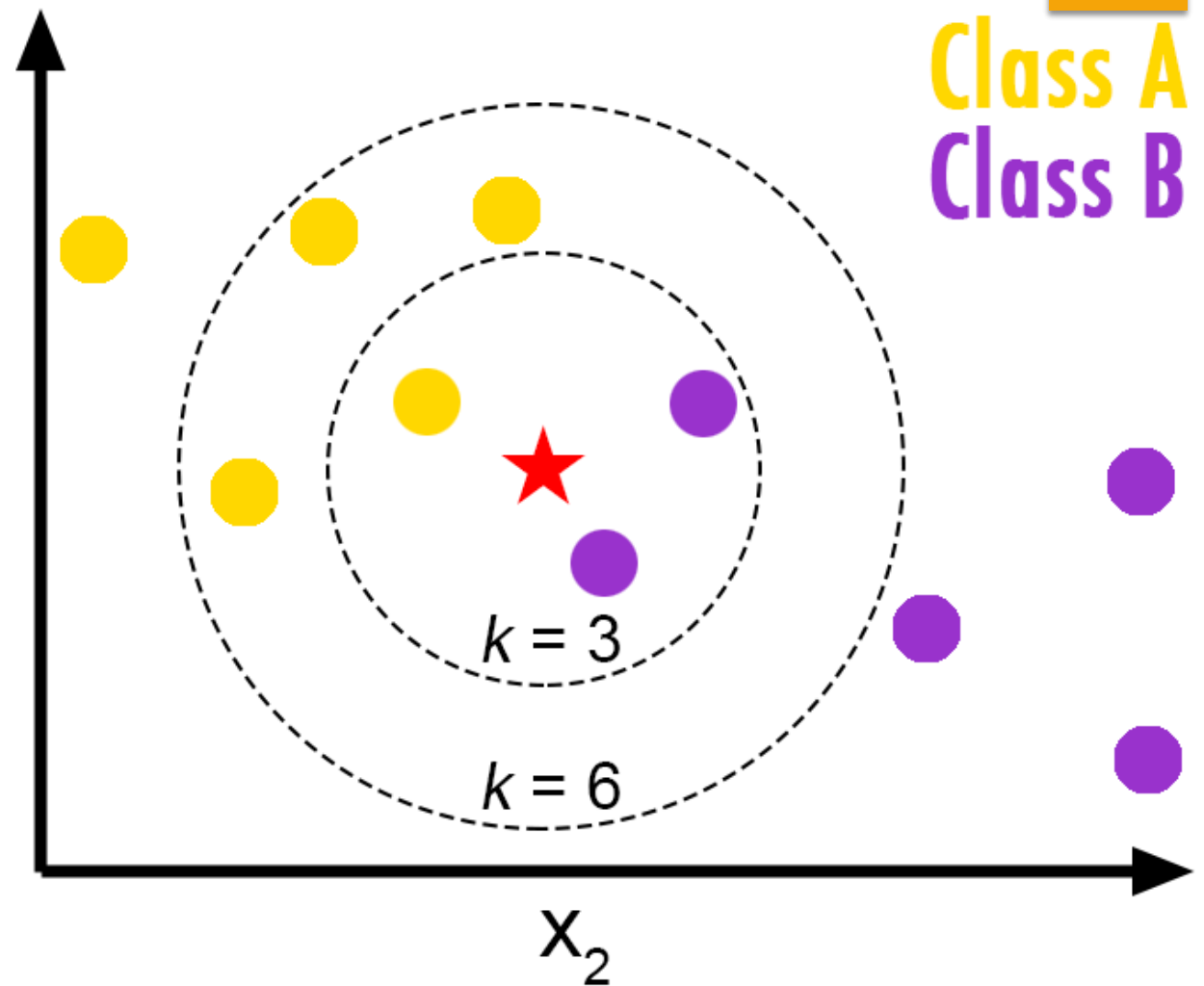


K- Nearest Neighbours

- ▶ K is number of neighbors
- ▶ Small K = overfitting due to learning localized patterns
- ▶ large K = highly generalize model and may miss complex patterns.



# KNN

- ▶ Can be used for both classification as well as regression predictive problems.
- ▶ Mainly used for classification predictive problems.
- ▶ **Lazy learning algorithm**
  - ▶ No specialized training phase
  - ▶ Uses all the data for training while classification.
- ▶ **Non-parametric learning algorithm**
  - ▶ doesn't assume anything about the underlying data.
- ▶ Finds result based on closely matching data points.
- ▶ Distance Based algo hence "Scaling" is required.

## Pros:

- ▶ KNN algorithm uncomplicated and easy to apply.
- ▶ There are only two metrics to provide in the algorithm. value of  $k$  and distance metric.
- ▶ Work with any number of classes not just binary classifiers.
- ▶ It is easy to add new data to algorithm.

## Cons:

- ▶ computationally expensive.
- ▶ High memory storage
- ▶ Hard to work with categorical features.
- ▶ Prediction is slow for big Number of features.
- ▶ sensitive to the scale of data and irrelevant features.