

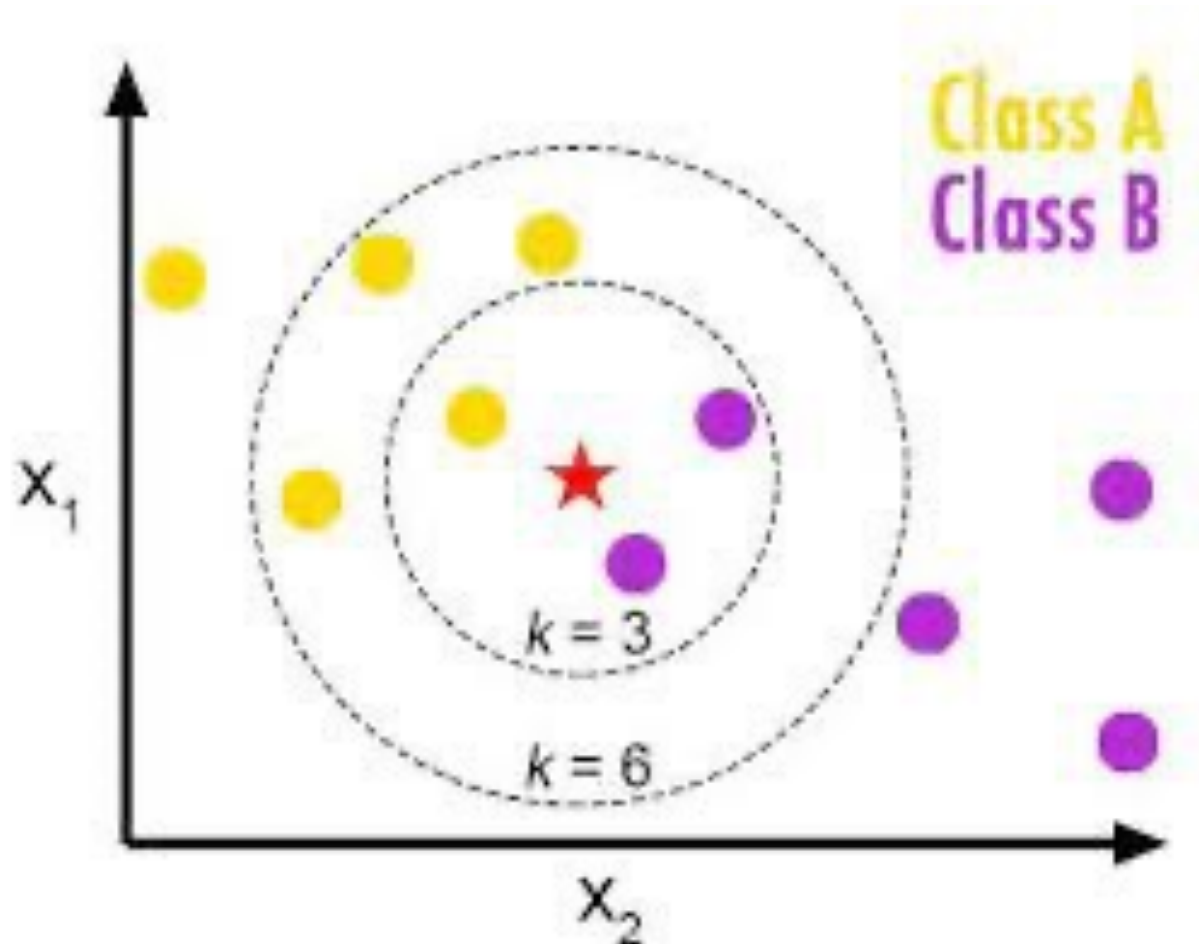
Data Science KNN and SVM

Steve Albury March 2022

KNN

- Supervised Learning
- Non Linear
- Simple model but powerful for
 - Categorisation
 - Regression
- Not so good in situations where the data is complex as can take a long time to model (e.g computer vision)

KNN – basic idea (from
towardsdatascience.com



KNN basic idea

- Looking for a target variable (usually called y)
- New data is compared to nearest k data points
- The new data point is predicted based on the classification of those nearest k data points.
- Based on the idea that things nearby will be similar
- A majority voting system is used
 - If there is a tie various techniques can apply including reducing / increasing k

KNN Tasks

- Towardsdatascience tutorial
- RealPython.com tutorial
- Ch 12 of book – write your own classifier

SVM Support Vector Machines

- A binary linear classifier (things are above or below a threshold level)
- Maximises distance between the two parts of a hyperplane
- Can use the 'kernel trick' to increase dimensions where no line can be drawn to separate the two parts – massively complicates measuring distance
- A useful guide to working with SVMs
 - <https://towardsdatascience.com/understand-support-vector-machines-6cc9e4a15e7e>

Visualising SVM (from towardsdatascience.com)

