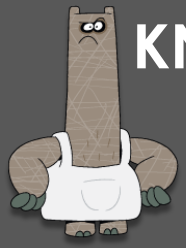
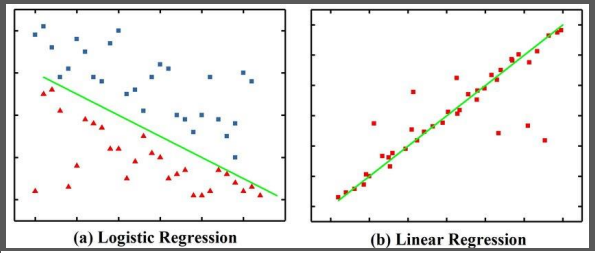
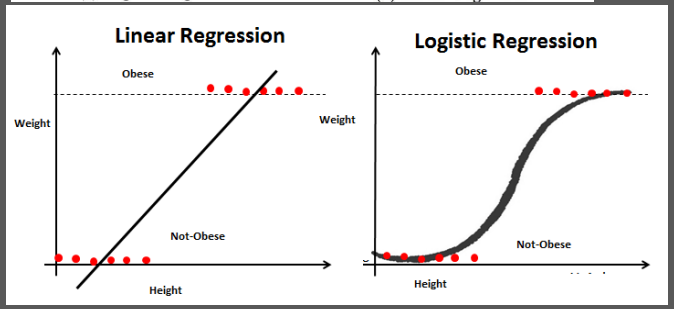


KNDN and SVM

Discussion Chowder

Reminder: we have learned that classifier is used to categorize data into groups

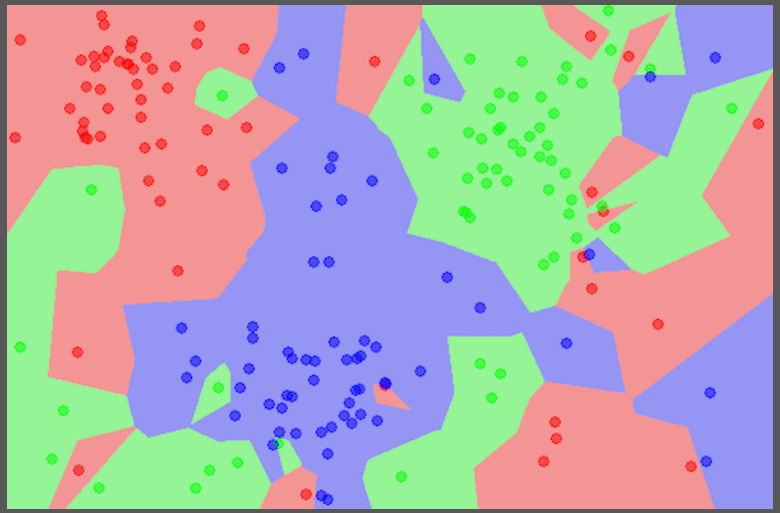




We learned logistic LA GRANGEsion

Which draws a line to separate data into each group

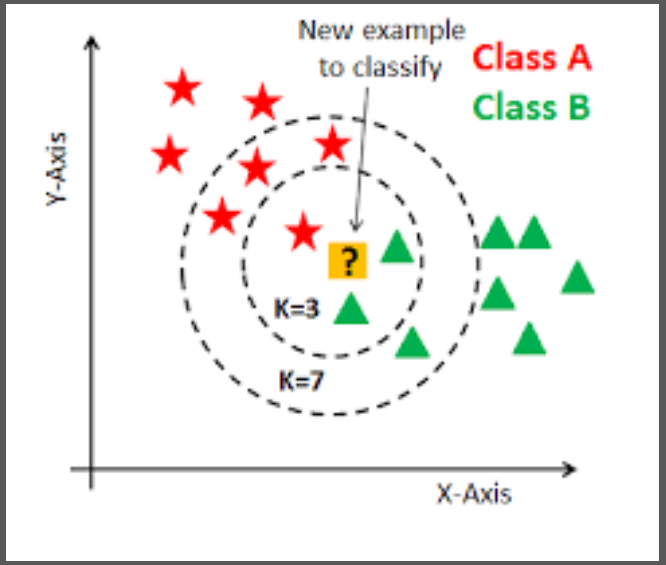
Problem: what if your data is not linearly seperable?



Introducing: K-nearest neighbors (KNN)

KNN works by choosing a number (K), then find its K nearest data neighbors. After that, use majority vote to classify that data

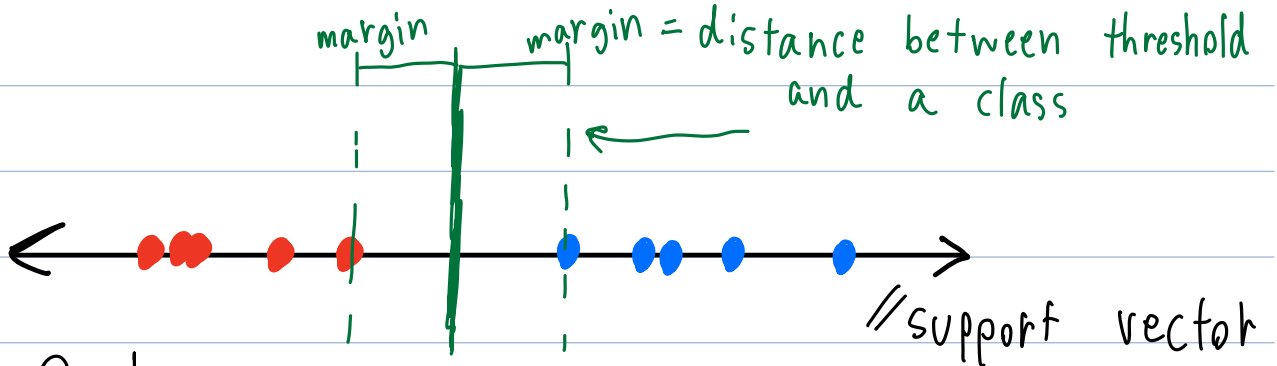
Example:



Problem: real data is not ideal. There are always outliers, so KNN may classify incorrectly

Solution: Support Vector Machines (SVM)

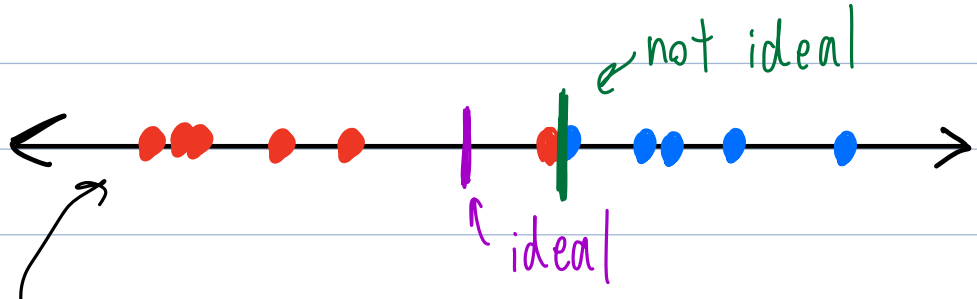
Ex



// support vector classifier //

Goal: Maximize the margin

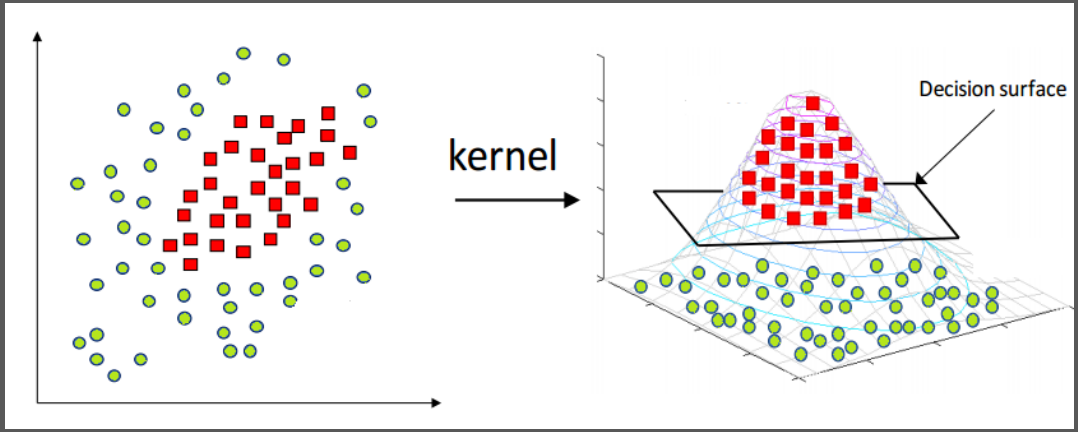
But what if the data looks like this?



Soft SVC: allow some misclassified data

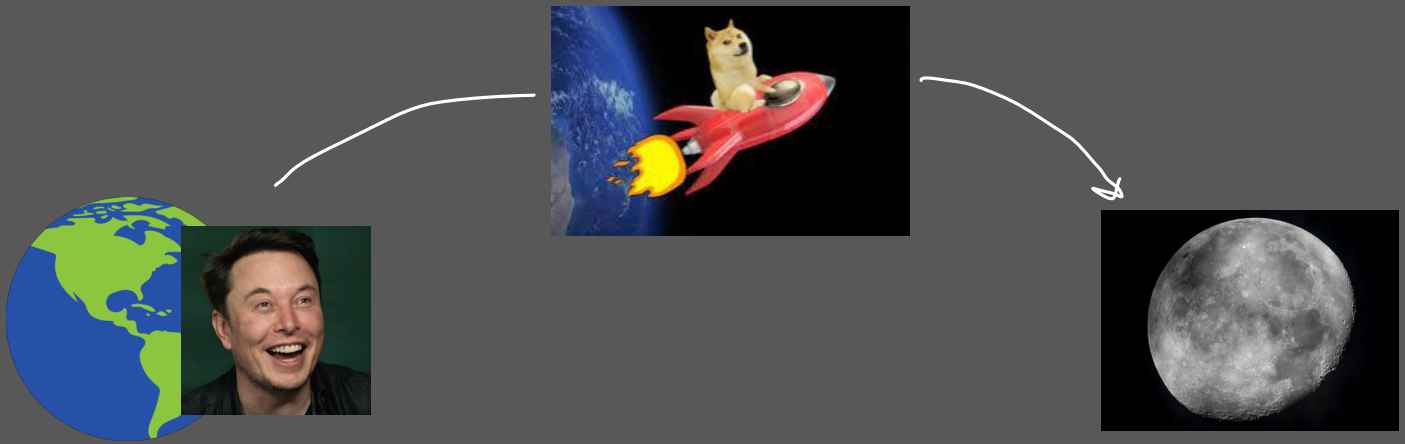
Problem: But what if the data is still linearly inseparable?

Solution: Kernel // transform data points



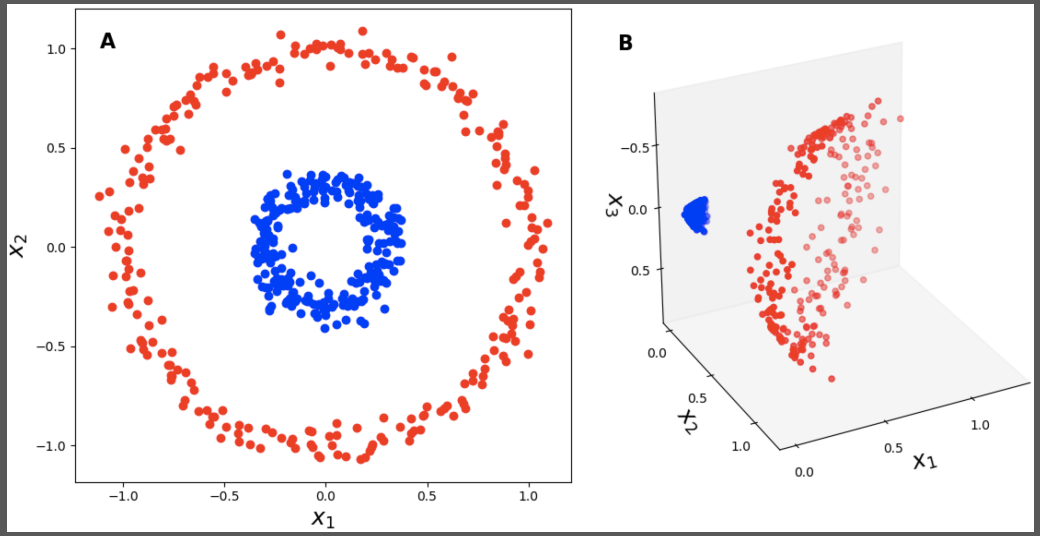
We call this Support Vector Machine

Analogy



Want to calculate data by the moon scale, so send everything to the moon and calculate there.

world x $\Longrightarrow$ world x^2



But you know, I learned something today

- For linearly inseparable data, we can use KNN or SVM
- KNN finds test data's k-nearest neighbors and its majority class
- SVM draws decision boundary. It can calculate decision boundary in higher degree