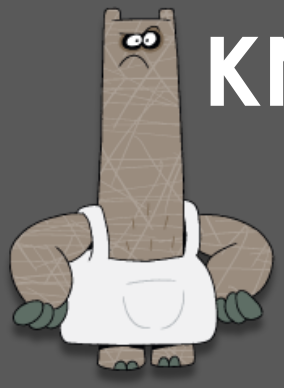
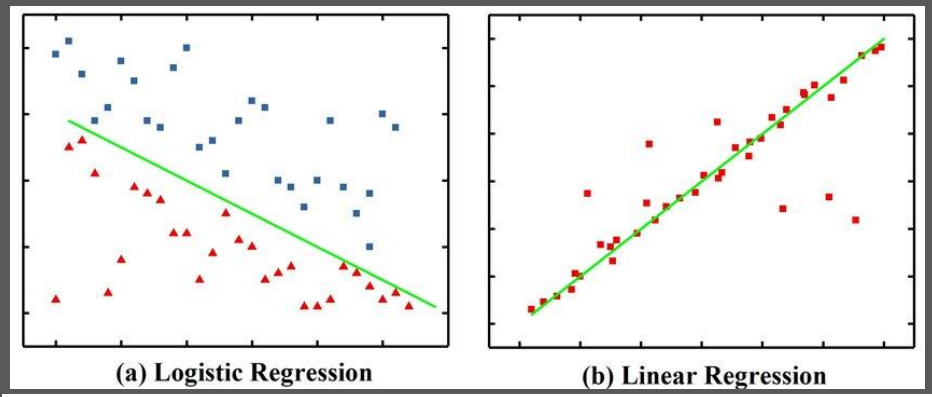
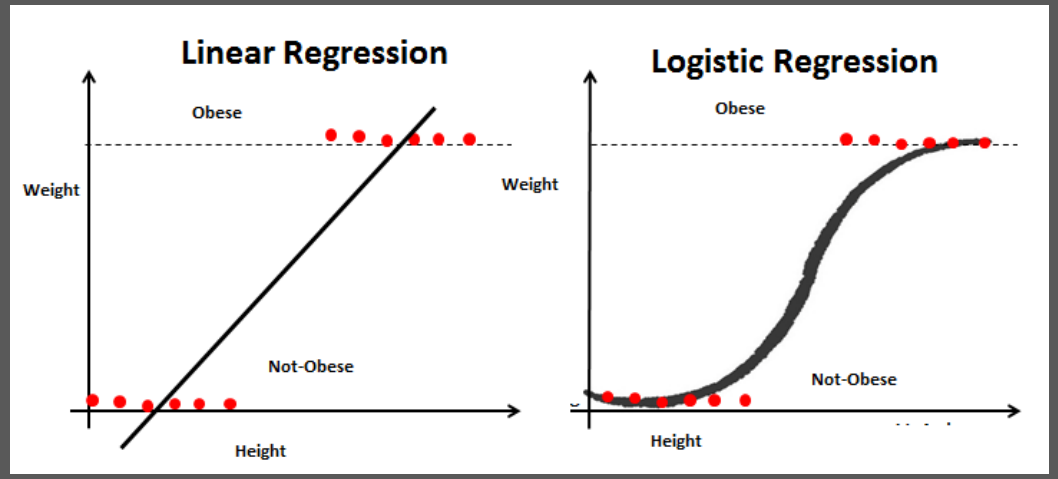


KN~~D~~N and SVM

Discussion **C**howder

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

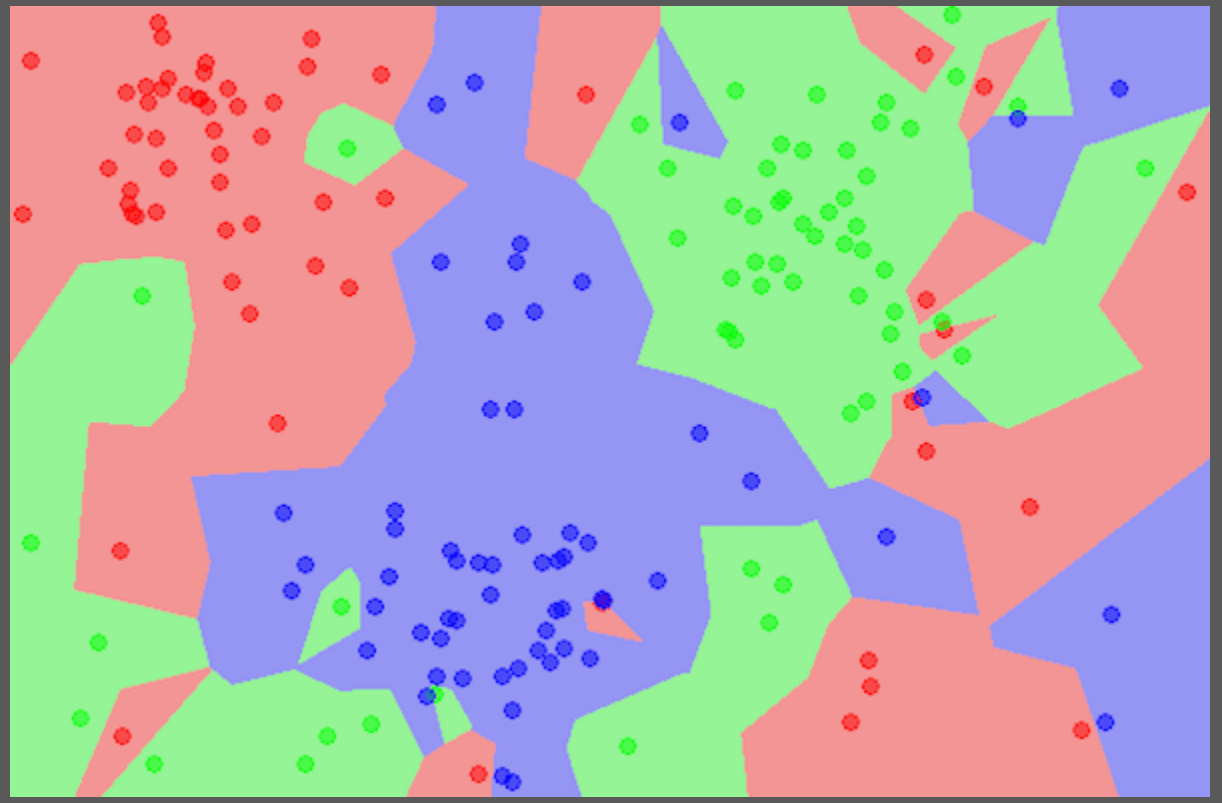




We learned Lo(oney)gistic Regression

Which draws a line to separate data into each group

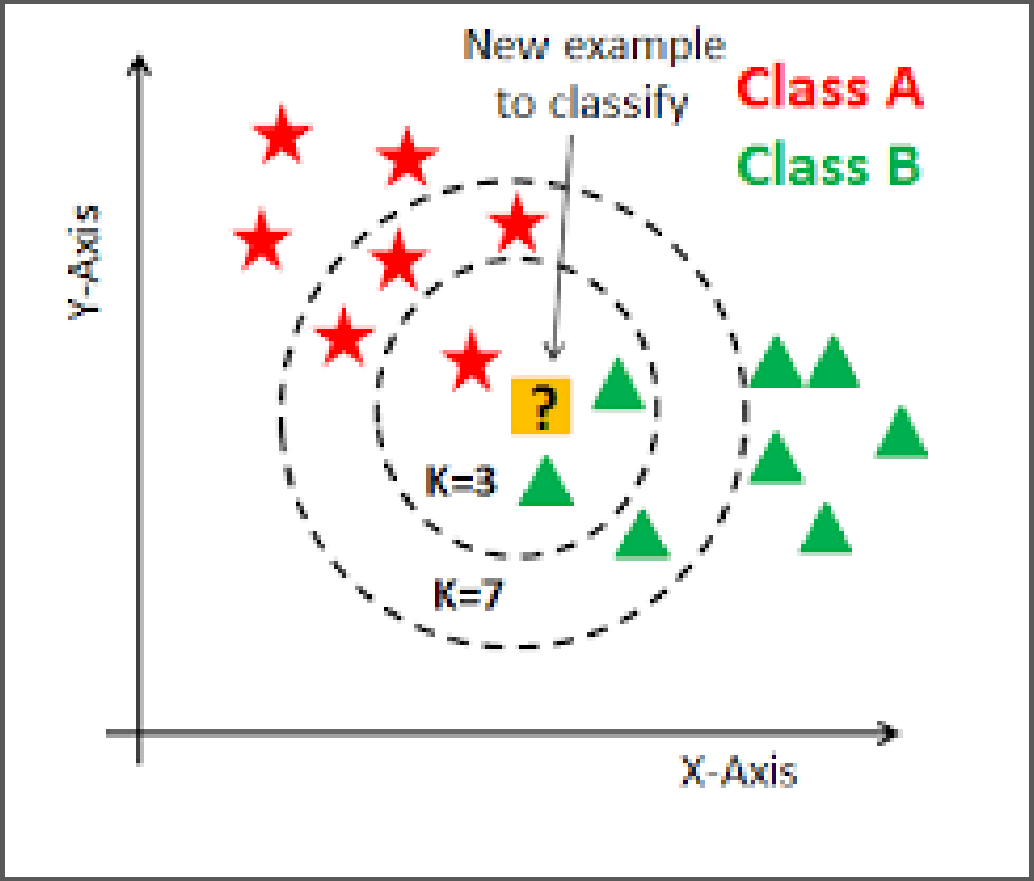
Problem: what if your data is not (F)lin(tston)early seperable?



Introducing: K-(Phi)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:

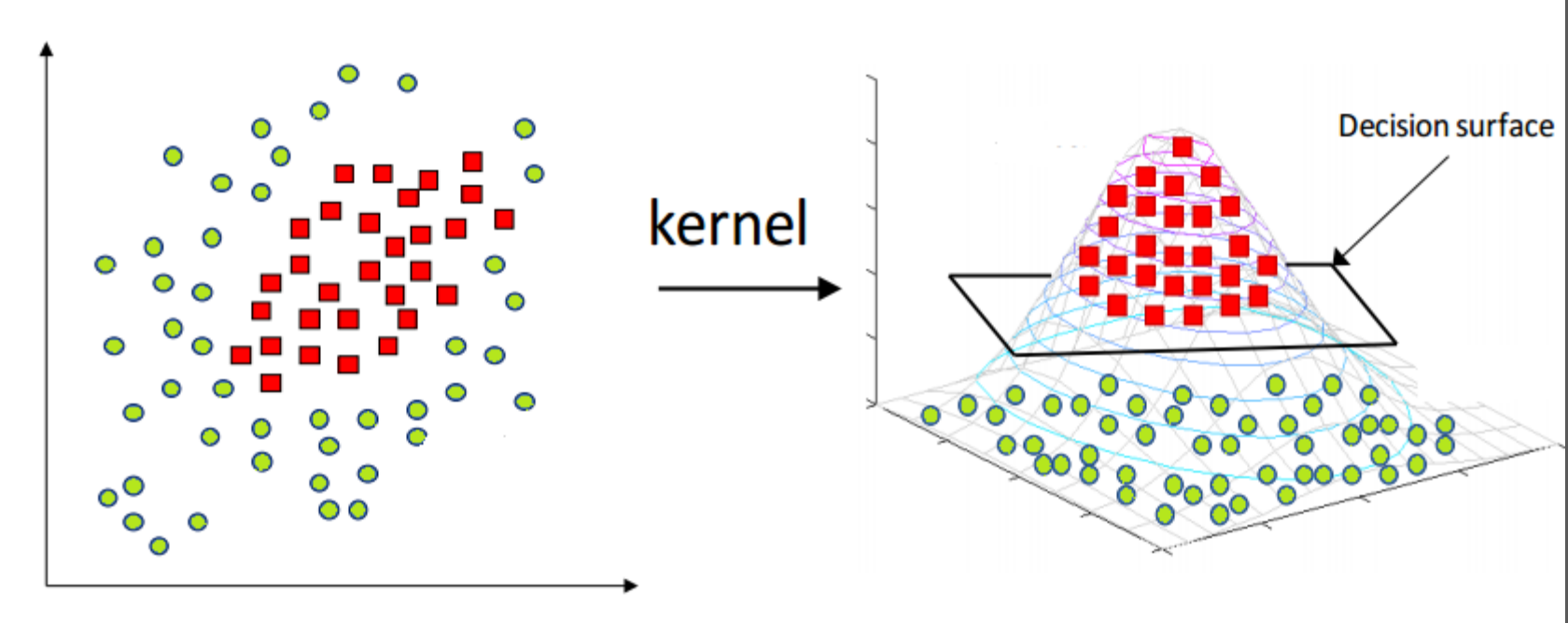


Problem: real data is not ideal. There are always outliers, so KNN may classify incorrectly

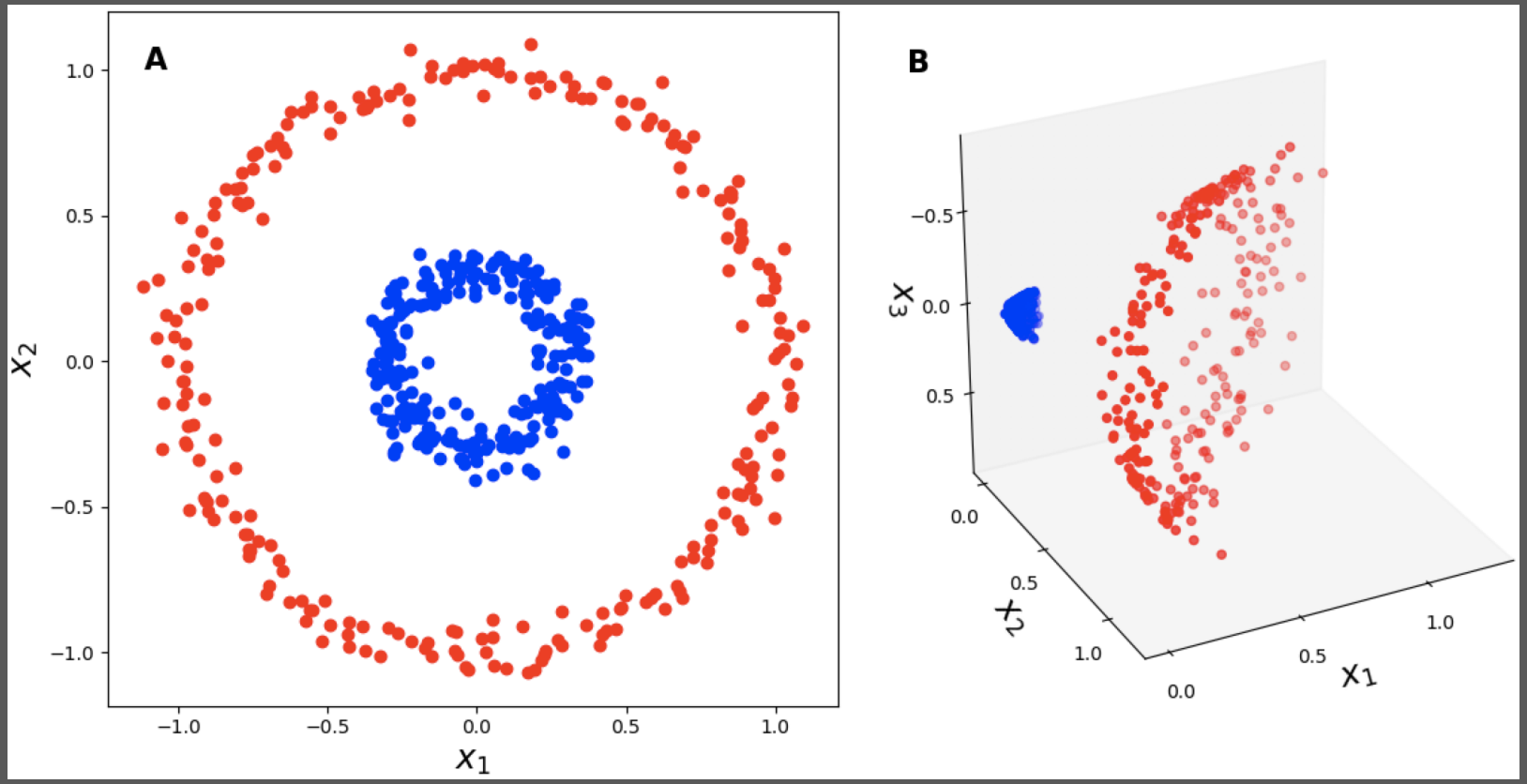
Solution: Sup(er)port Vector Ma(n)chines (SVM)

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

Solution: Kernel



Analogy


A
B
1.0
0.5
0.0
−0.5
−1.0
x_2
x_1
x_3
x_4

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