KNDN and SVM
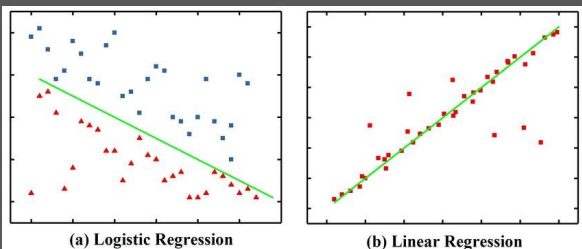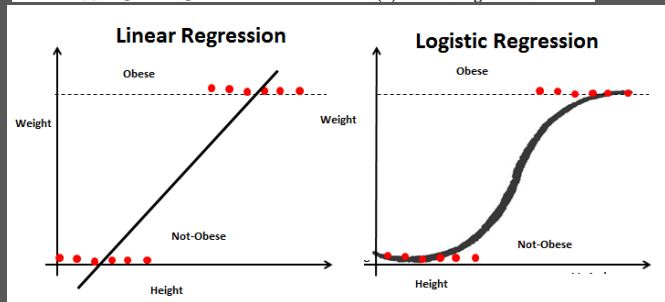
Discussion **C**howder

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

(a) Logistic Regression
(b) Linear Regression

Linear Regression
Obese
Weight
Not-Obese
Height

Logistic Regression
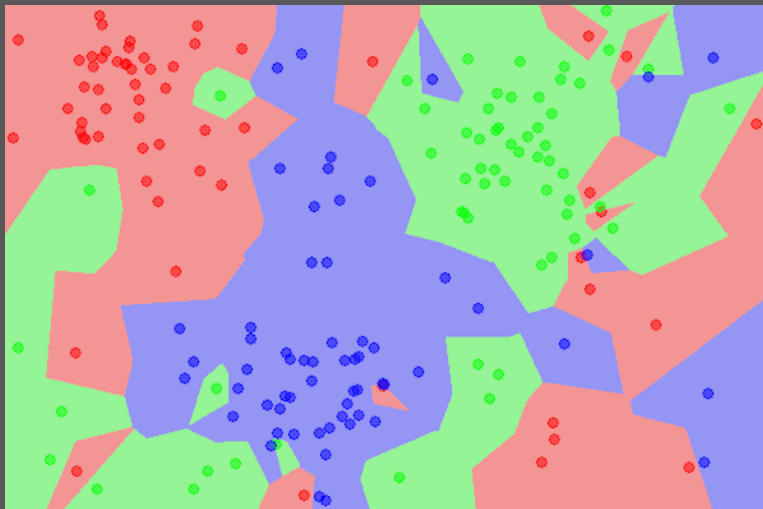Obese
Weight
Not-Obese
Height

We learned Lo(oney)gistic Regression

Which draws a line to separate data into each group

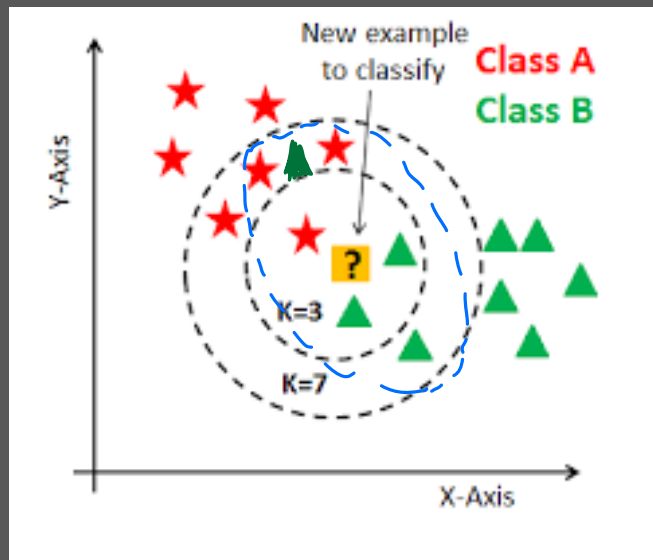Problem: what if your data is not (F)lin(tston)early seperable?

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
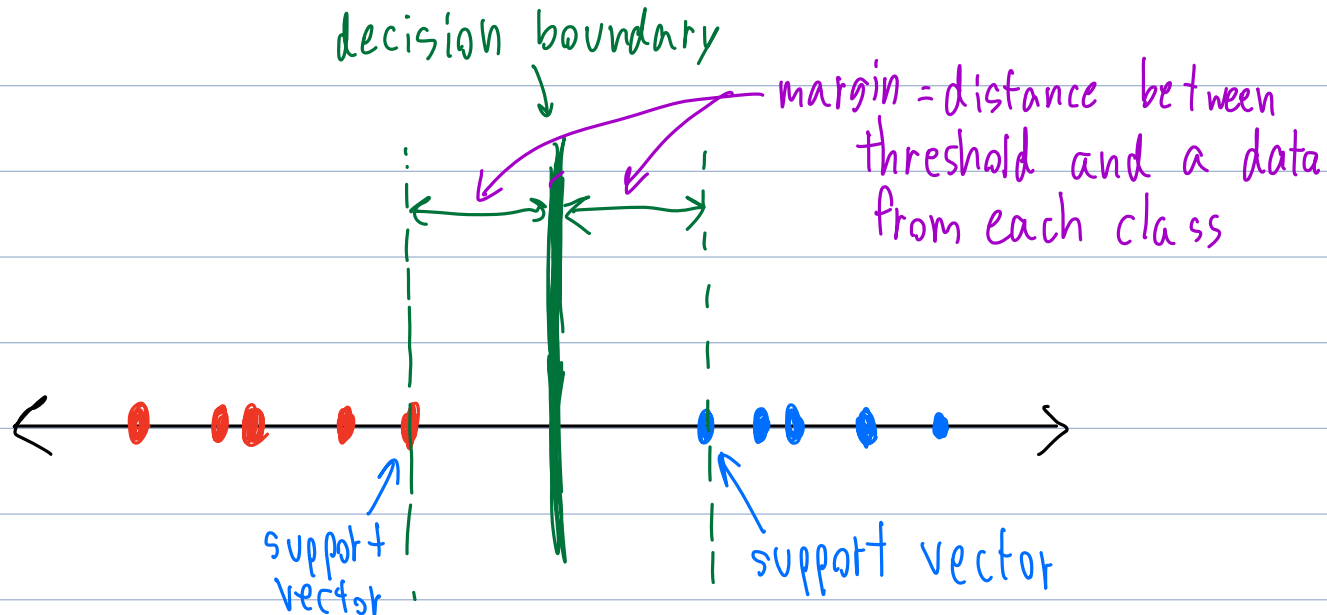

I love democracy.

Example:

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

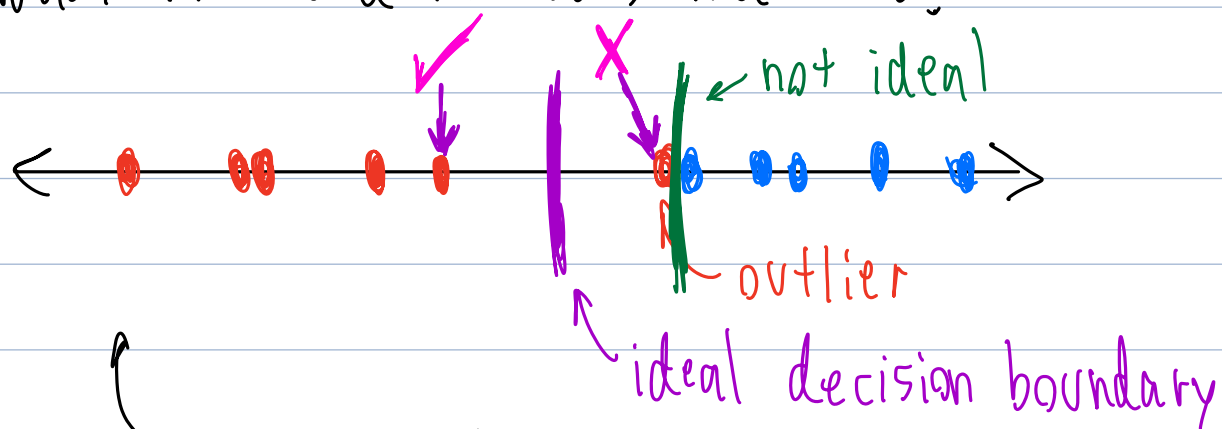**Solution**: Sup(er)port Vector Ma(n)chines (SVM)

<u>Ex</u>

decision boundary

margin = distance between threshold and a data from each class

support vector

support vector

Goal: maximize the margin

"Support Vector Classifier"

But what if the data looks like this?

not ideal

outlier

ideal decision boundary
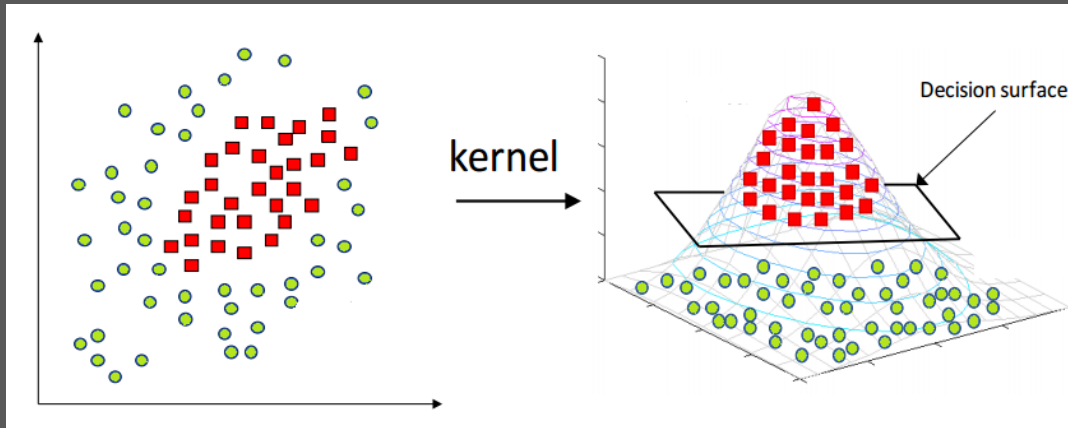
"Soft SVC": allows some misclassified or outliers data

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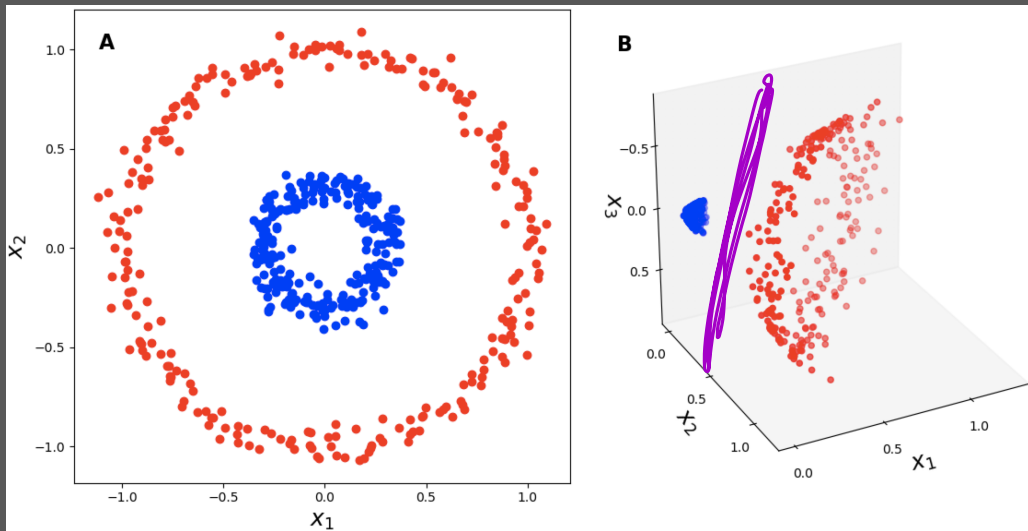
Solution: Kernel

Analogy   Support vector Machine
        ↳ SVC + kernel

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

2D

(2,2)

$\Downarrow$

3D

(4, 4, 4)

World x $\xrightarrow{\text{Kernel}}$ World x^2

A

B

$[x, y] \implies [x^2, x \cdot y, y^2]$

But you know, I learned something today

Break until

1:42

– Colab Day 6

– HW2 p2.1

– MAL implem. idea

– group work time

- 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