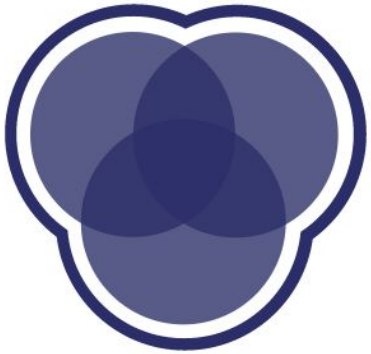


CDS

Cornell Data Science

# SVM

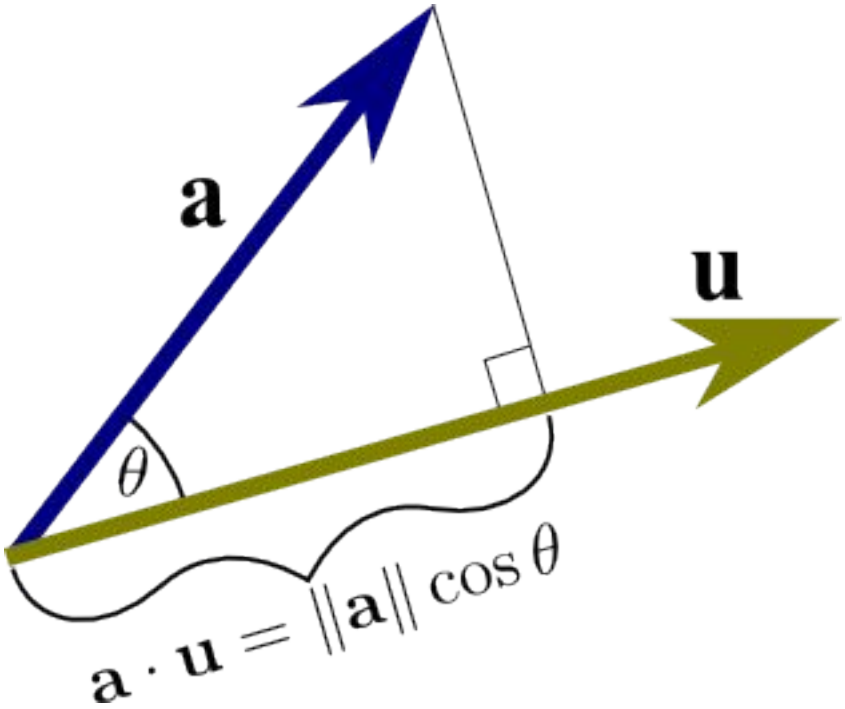
# Sanity Check

- Project A
  - any concern or questions?
  - Did everyone turn in their project?
- OH room changed to 122
  - Come come come
- Project B released today
  - Linear Regression
  - KNN Classification

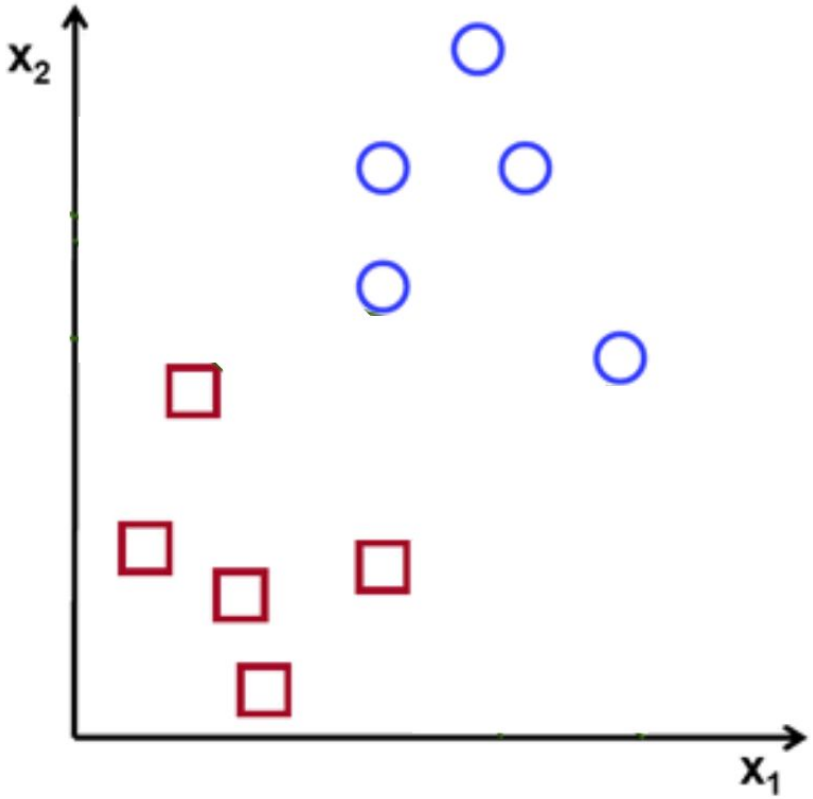
# Lecture Flow

- How it works (Derivations)
- What it is and how to use (Support Vector Machines)
- How to find the best one (Margins and Kernels)

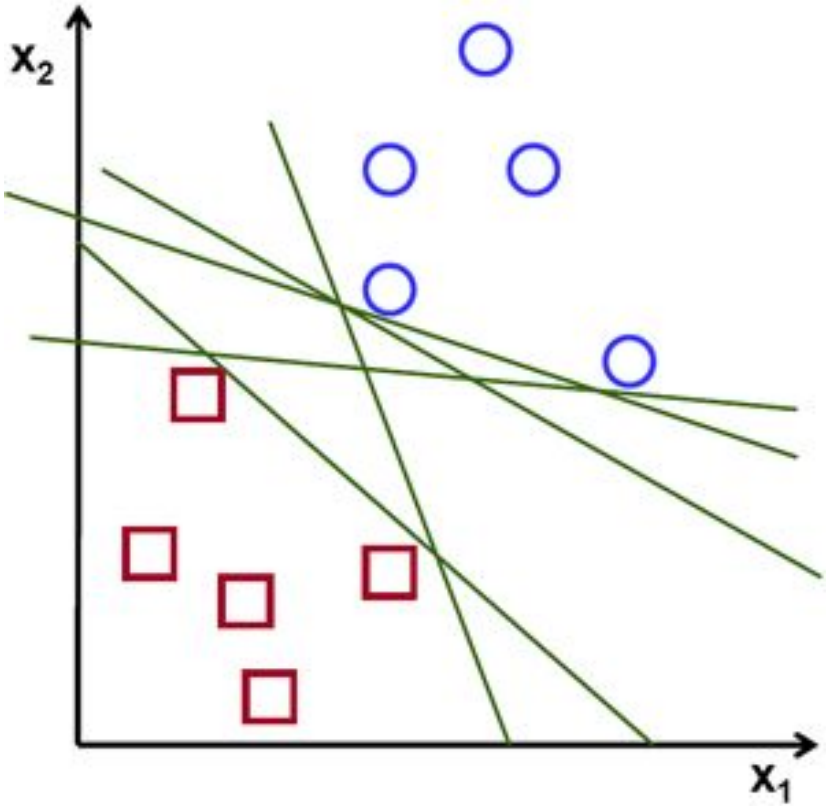
# Vector Dot Product



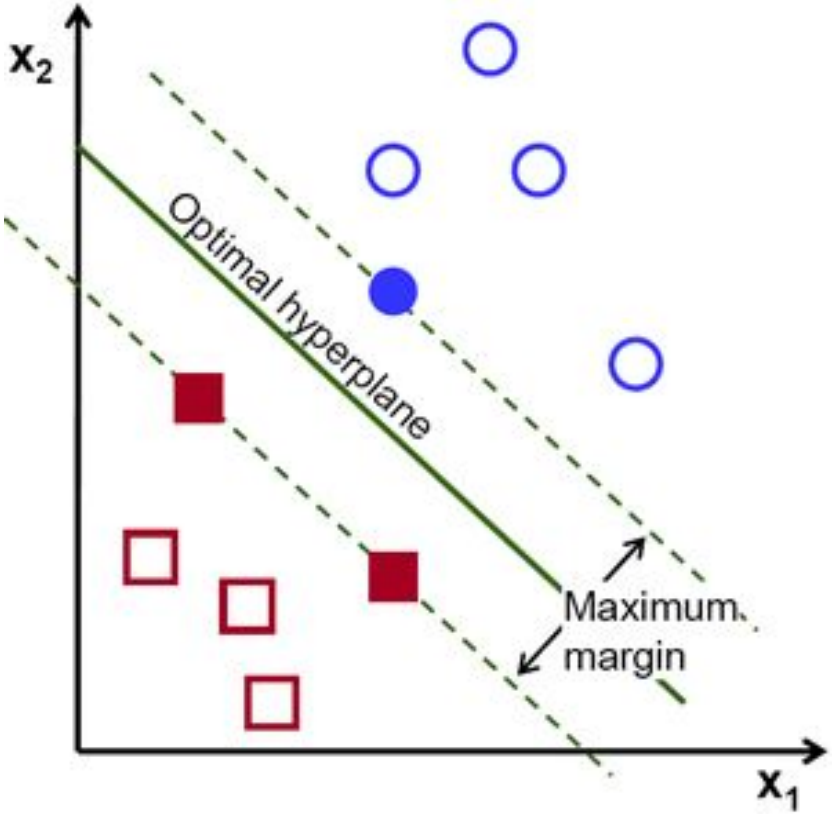
# Classify (+) and (-)



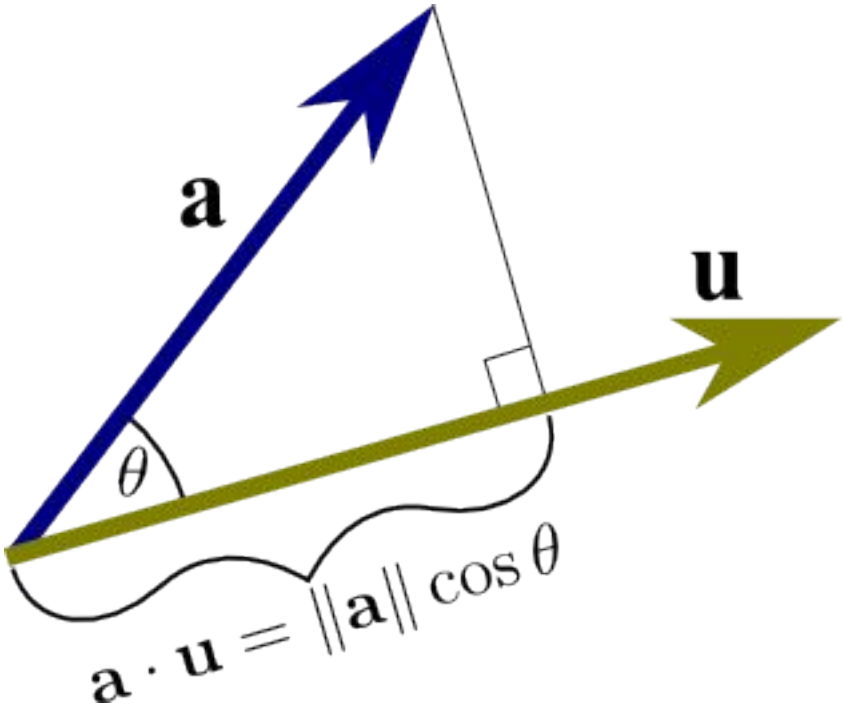

# Which Hyperplane?



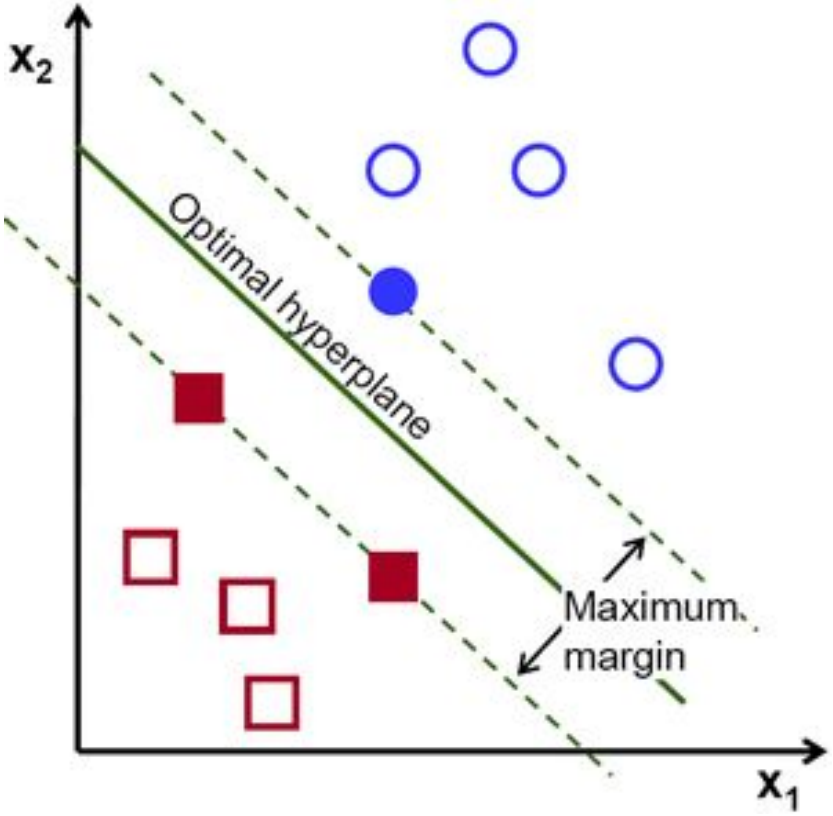

# Optimal Solution



# Vector Dot Product



# Optimal Solution



# Support Vector Machine

Memory efficient

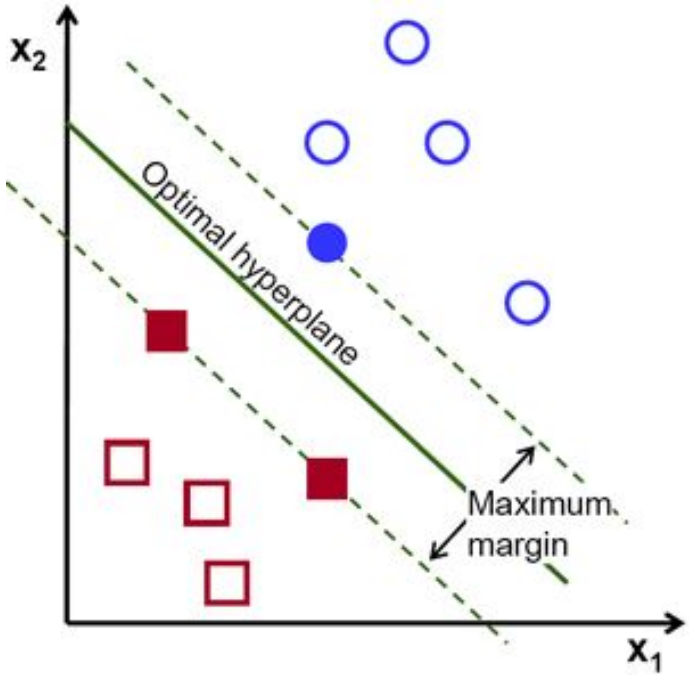
Used for classification in a higher dimension

Slow calculation time

Doesn't handle noise well

# Maximal Margin Classifier

- We want to find a **separating hyperplane**
- Once we find candidates for the hyperplane, we try to maximize the **margin**, the normal distance from borderline points
  - Only **Support Vectors** matter

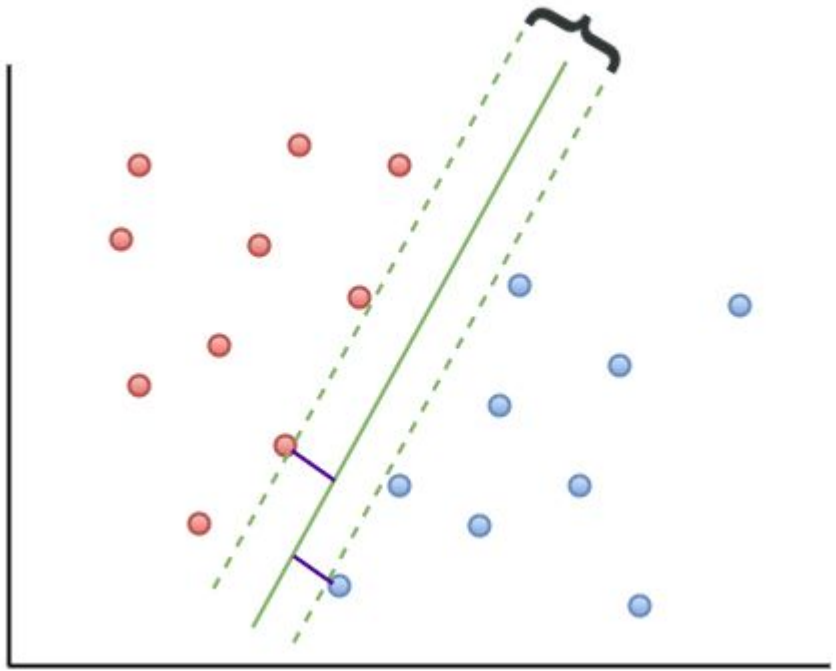



# 2 Dimensional Example

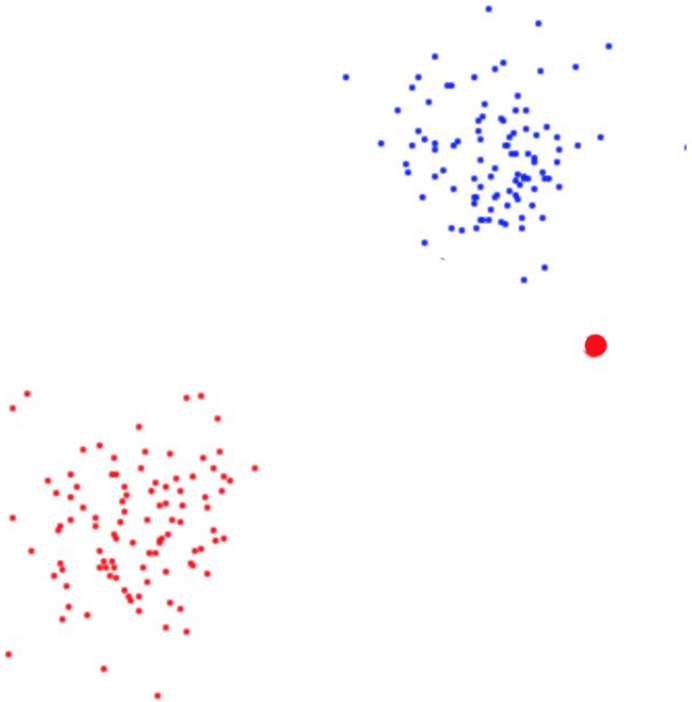
- The data points will be separated by a line

$$y = mx + b$$

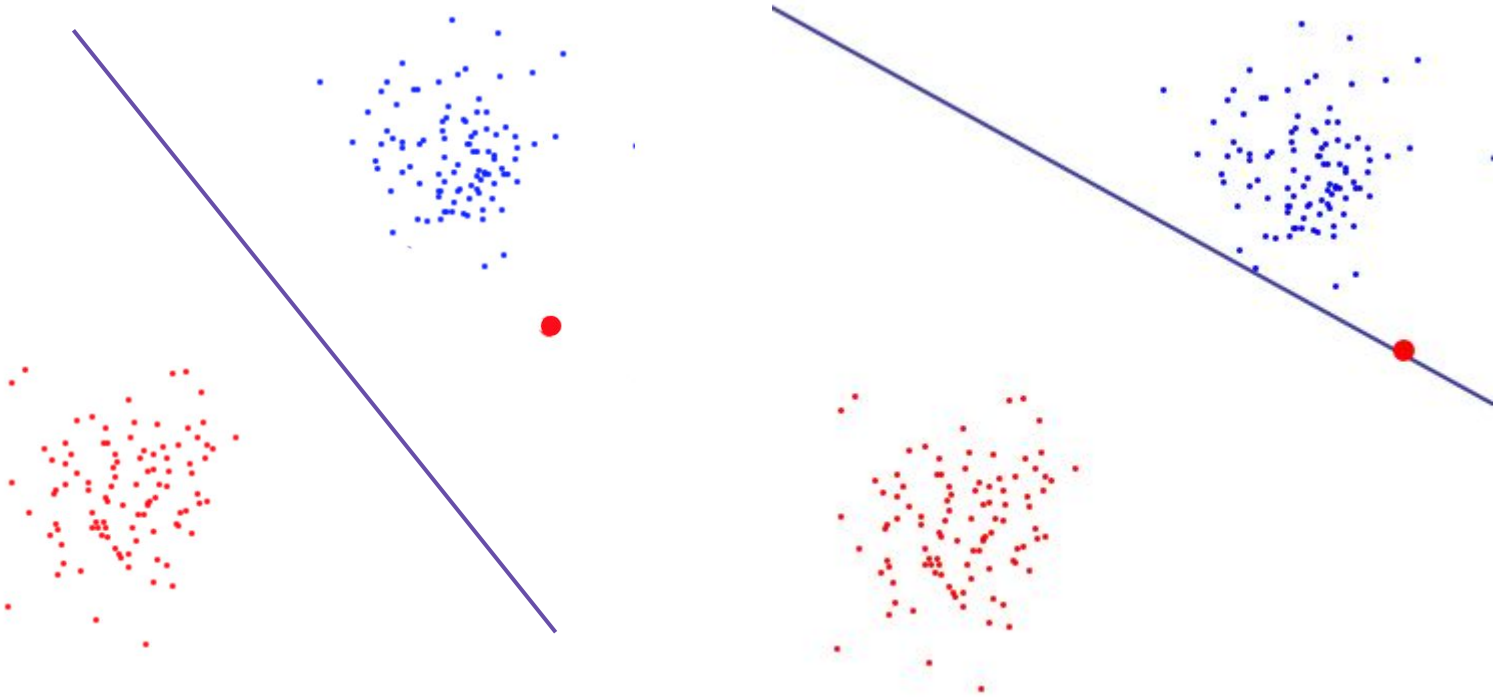
- Tweak parameters to find best line of separation



# What if..

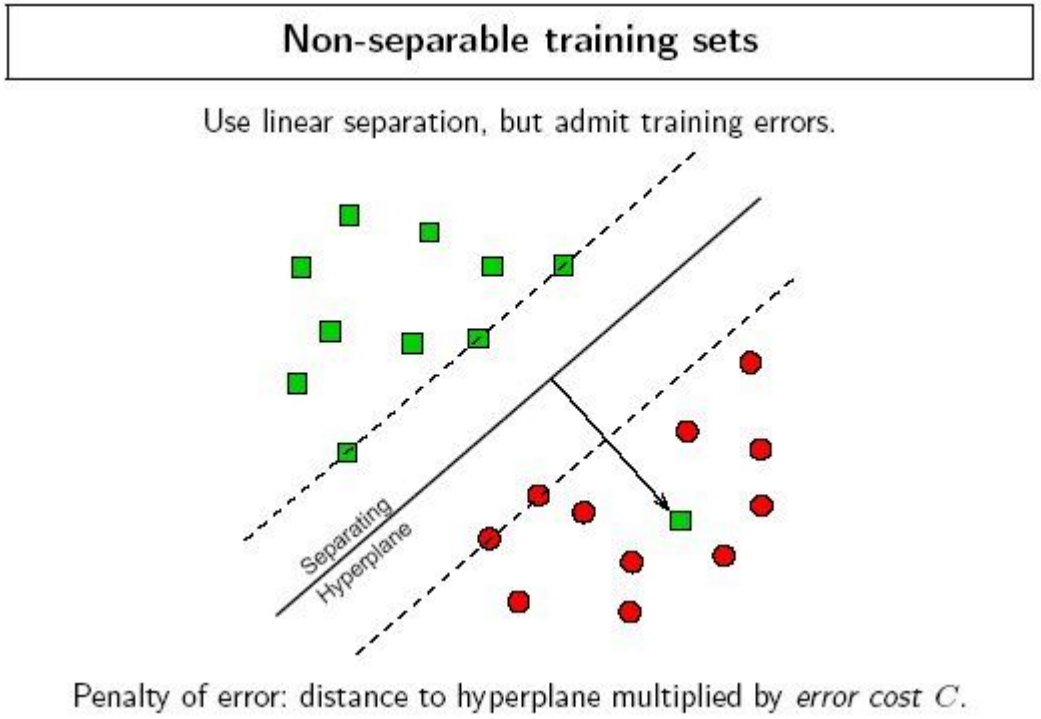


# Which one is better



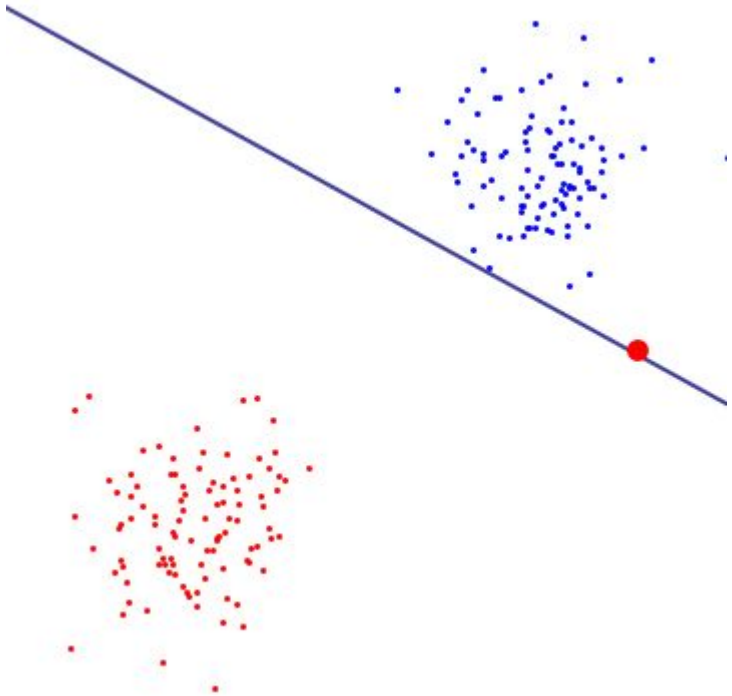
# Margins

- Cost function to penalize for errors
- Hard margins vs. Soft margins



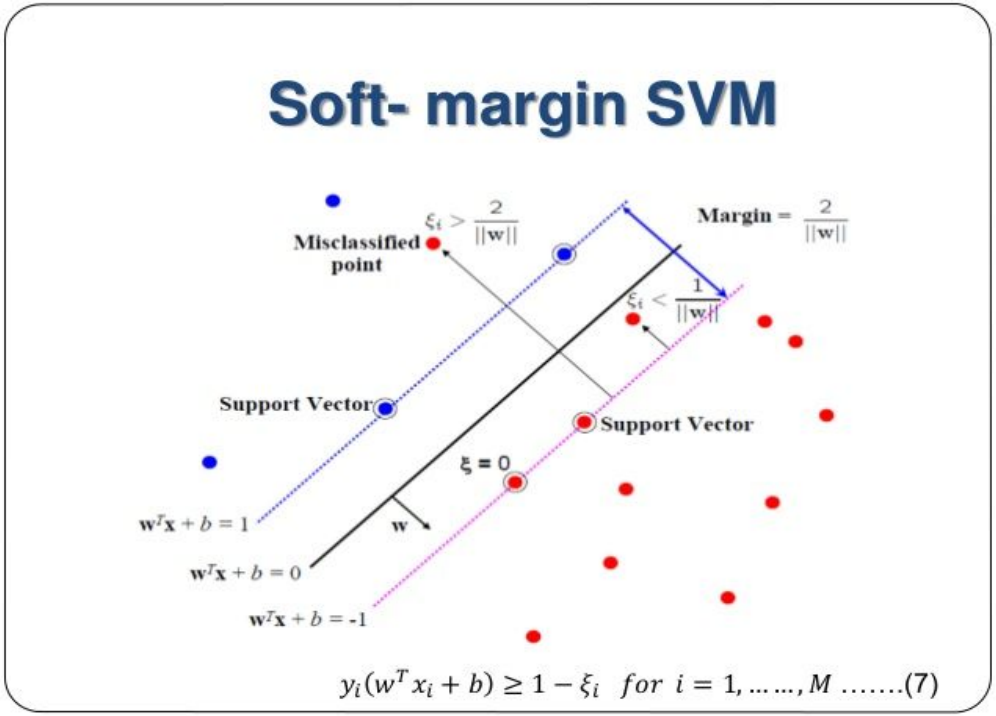
# Hard Margins

- High penalty value
- The hyperplane can be dictated by a single outlier

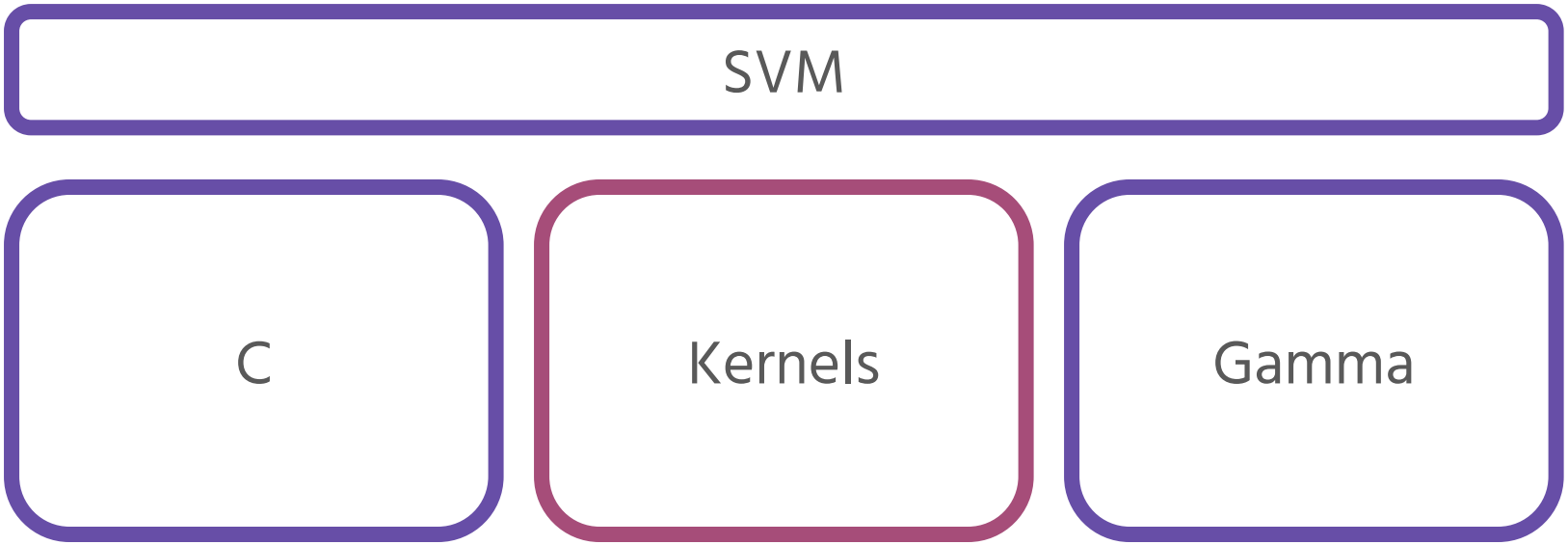


# Soft Margins

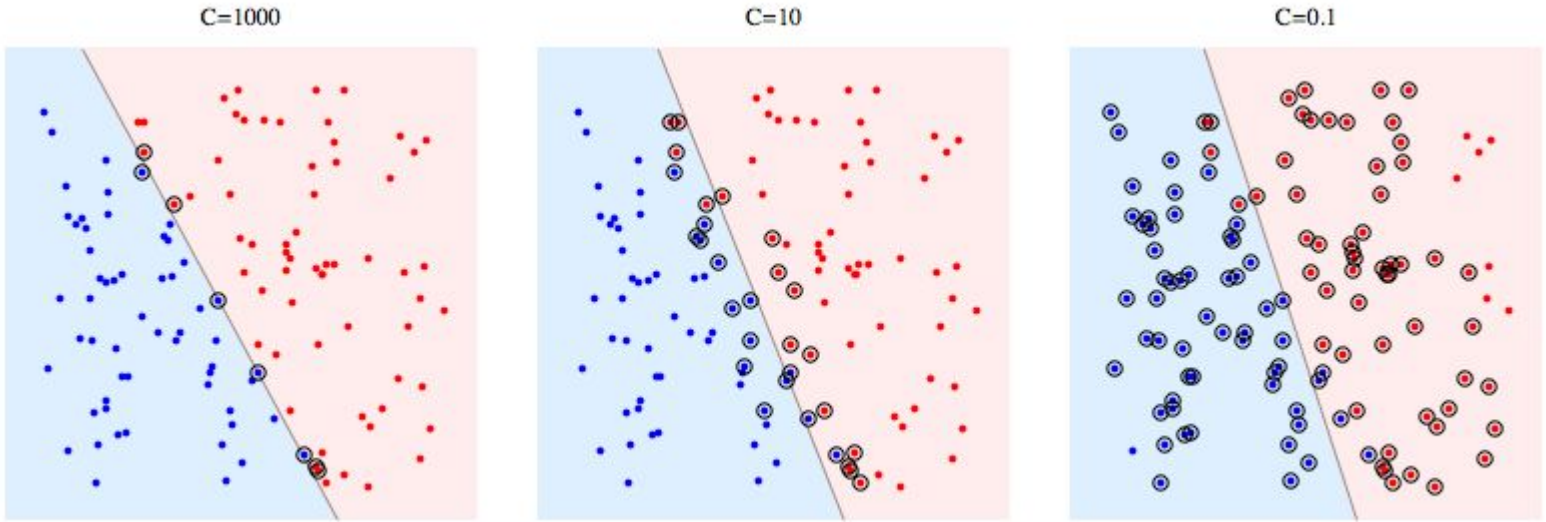
- Used in non-linearly separable datasets
- Allow for misclassification
- Can account for “dirty” boundaries



# Hyper-Parameters

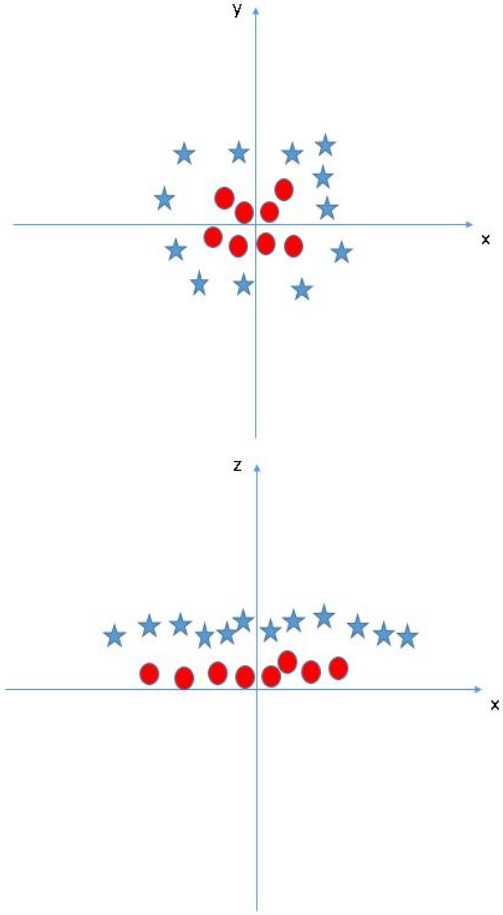



# C Penalty




# Kernels

- You cannot linearly divide the 2 classes on the *xy* plane at right
- Introduce new feature, $z = x^2 + y^2$ (**radial kernel**)
- Map 2 dimensional data onto 3 dimensional data. Now a hyperplane is easy to find. (Imagine slicing a cone!)



# Kernels

RBF

linear

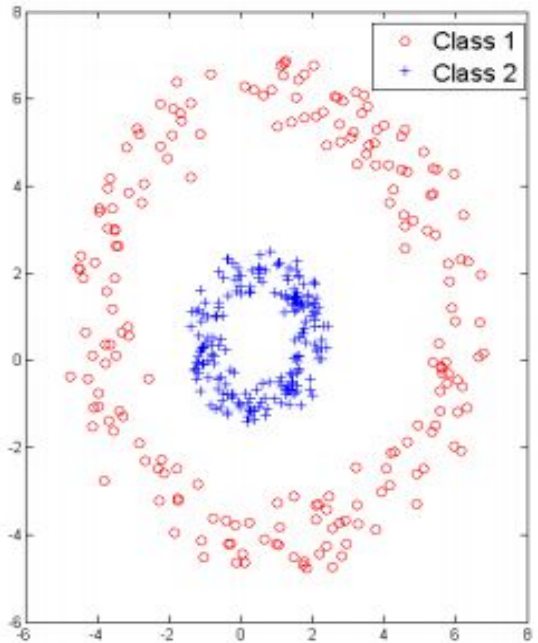
Poly

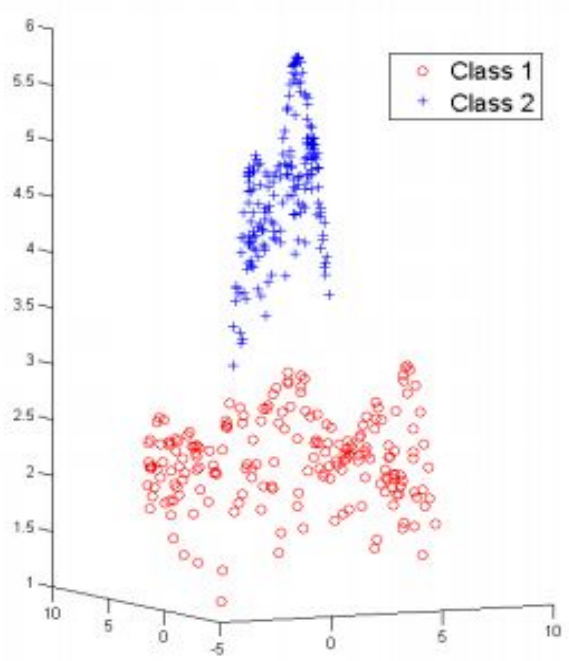
sigmoid

Pre-computed

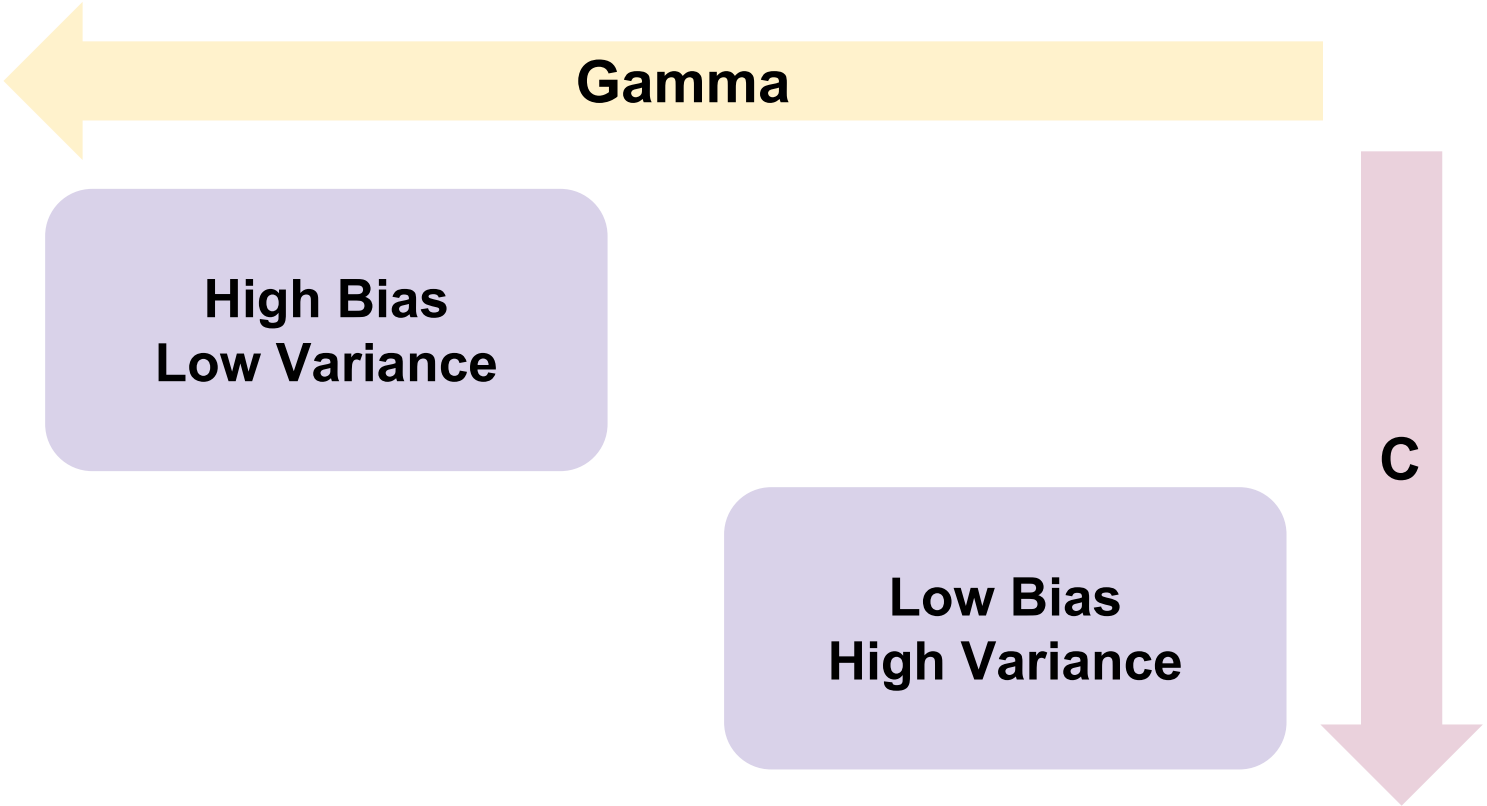
etc

# Gamma γ



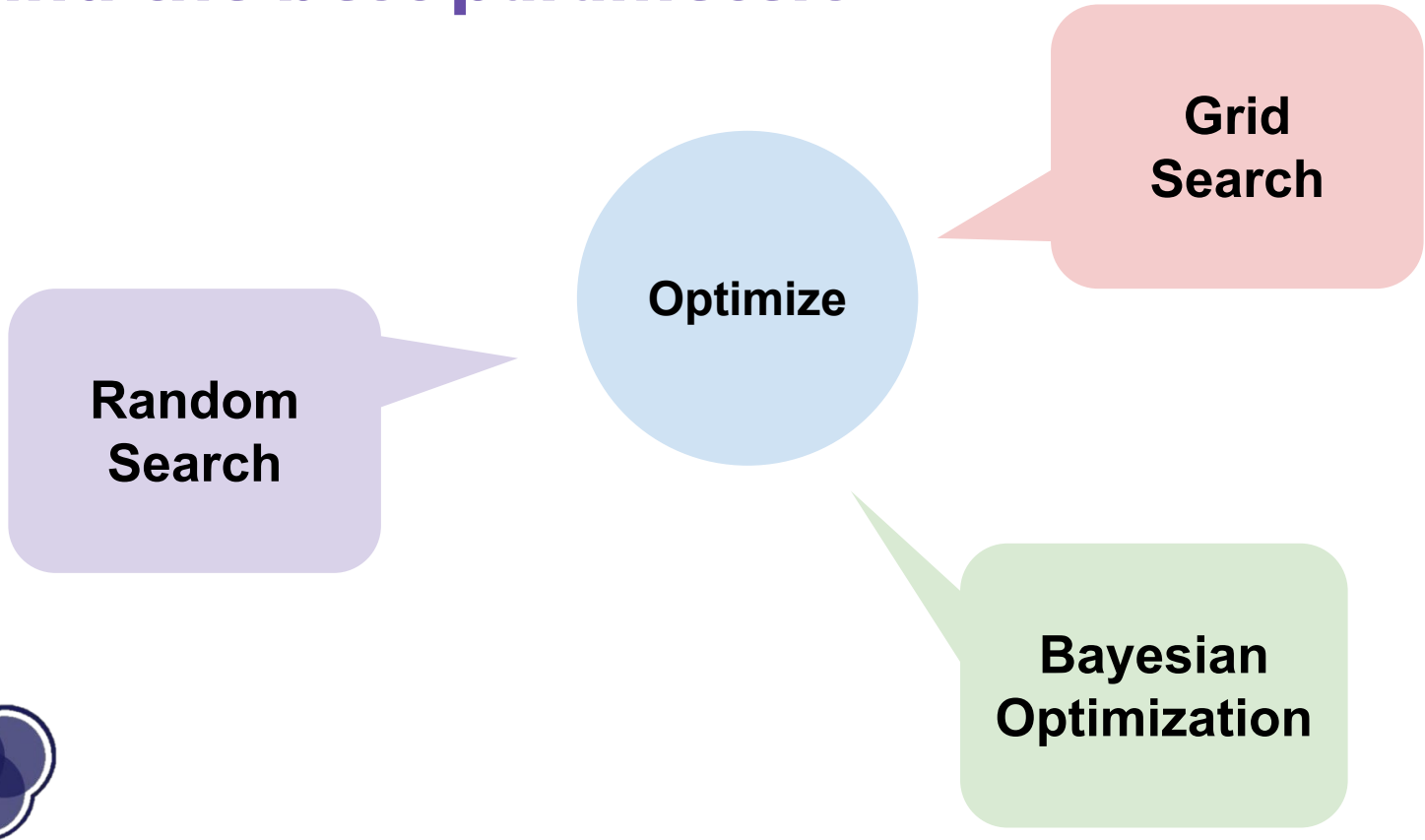




# C and Gamma



# **Demo:** Classification of Iris Species

# Find the best parameters



# Find the best parameters: Grid-Search



# Coming Up

**Your problem set:** Project part B

**Next week:** Logistic Regression and Decision Trees