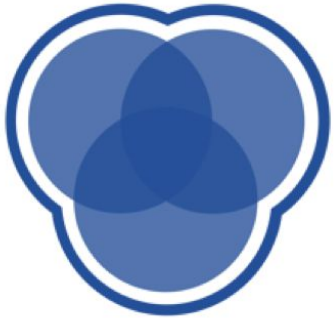
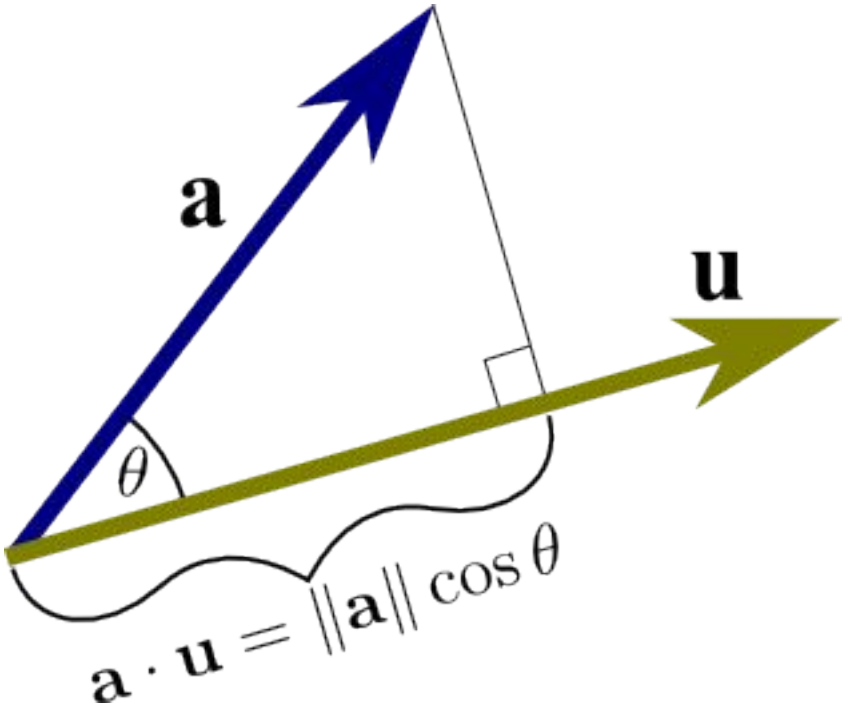


CDS Education

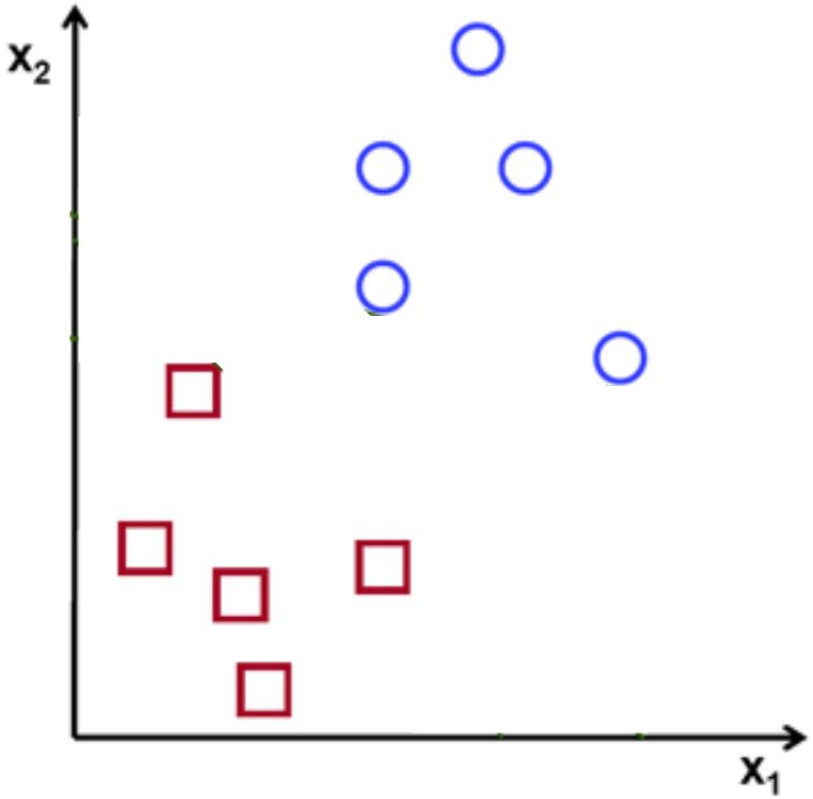
Introduction to Machine Learning for Python

Support Vector Machine

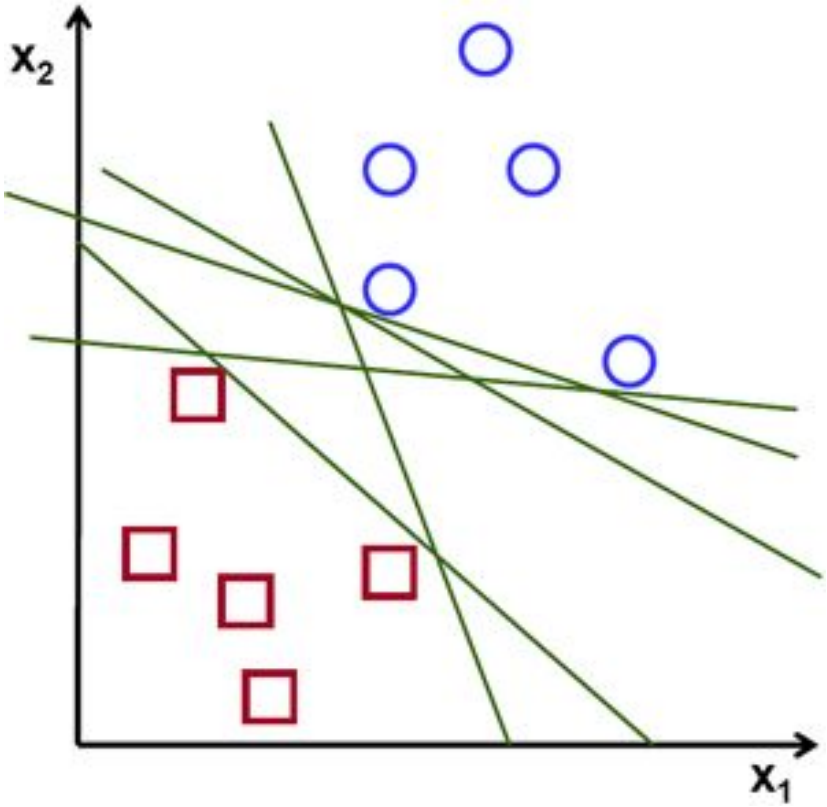
Vector Dot Product



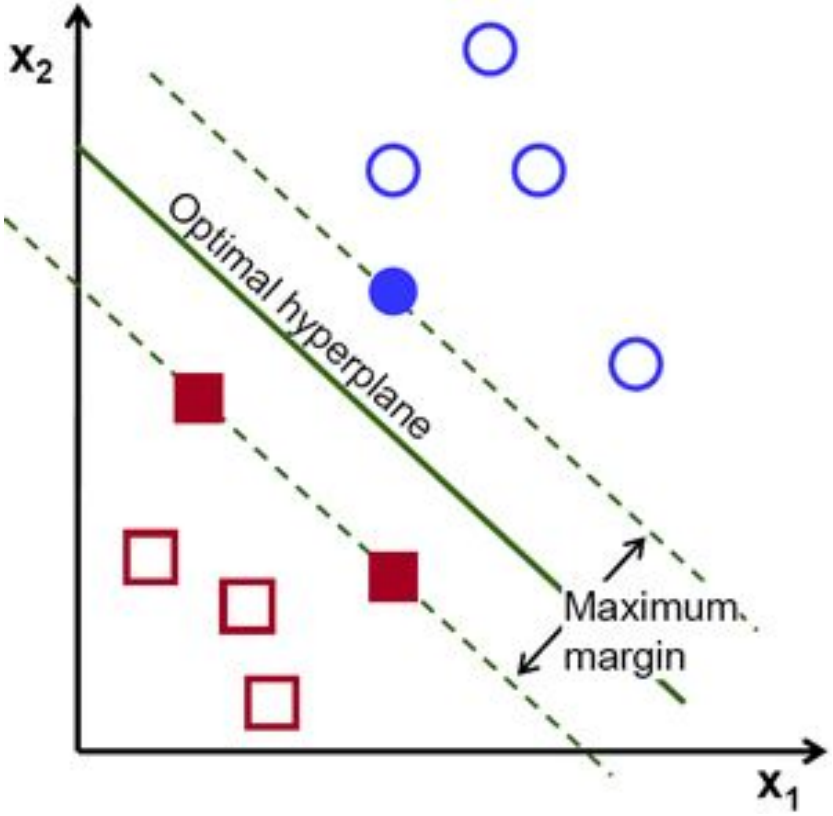
Classify (+) and (-)



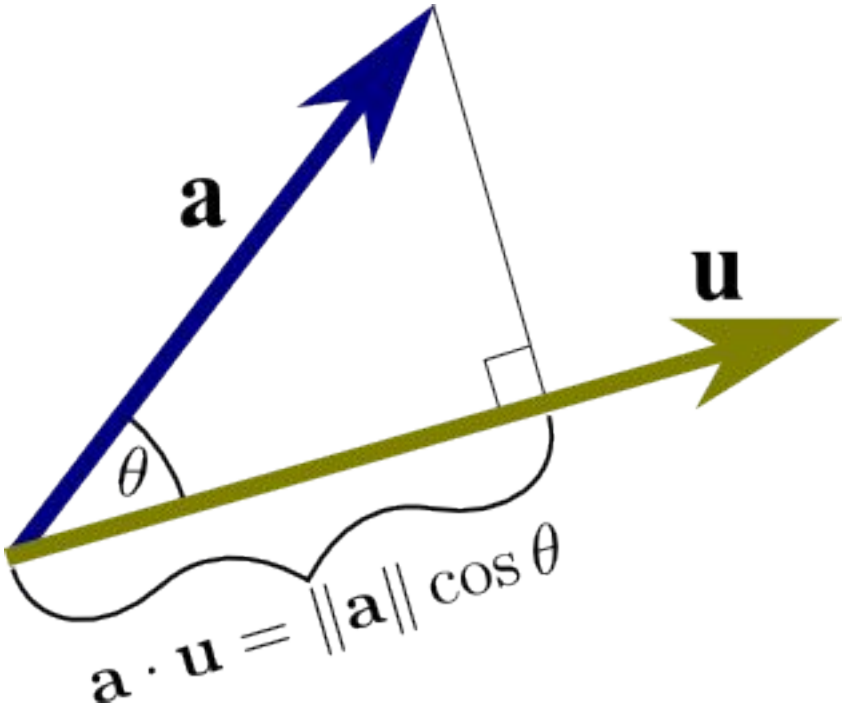
Which Hyperplane?



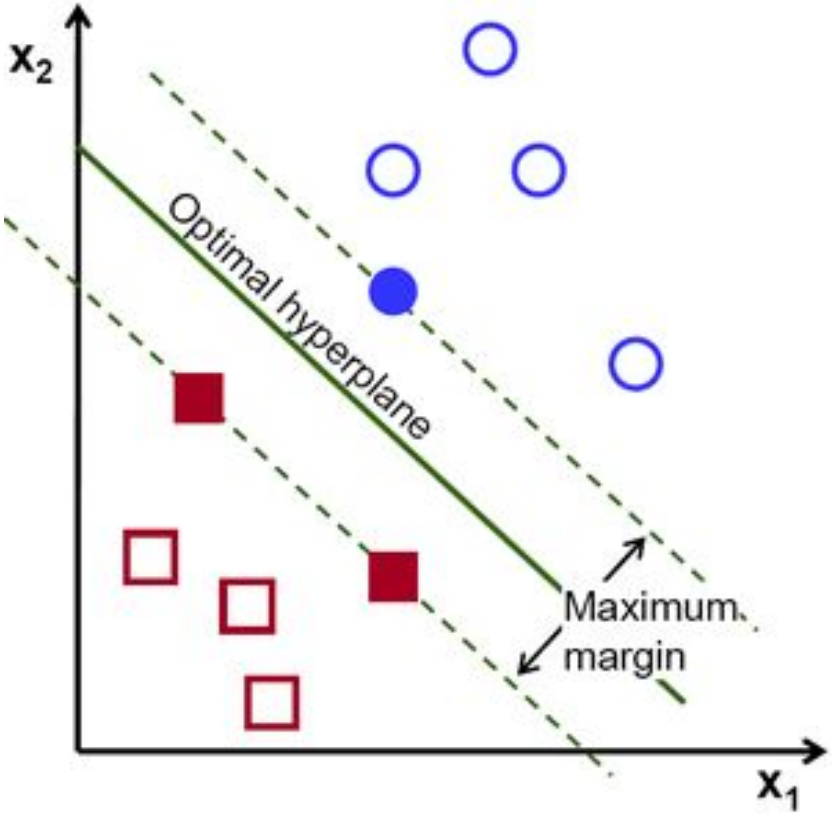
Optimal Solution



Vector Dot Product



Optimal Solution



Support Vector Machine

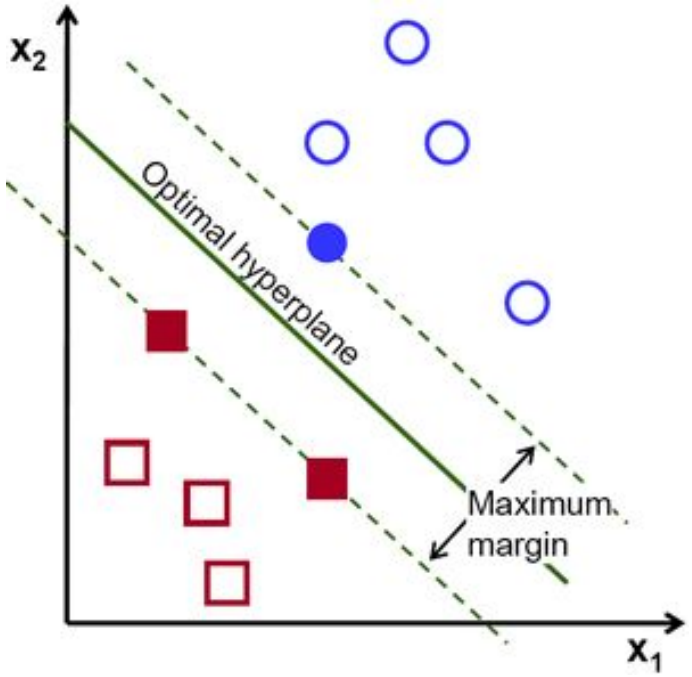
Memory efficient

Used for classification in a higher dimension

Slow calculation time

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

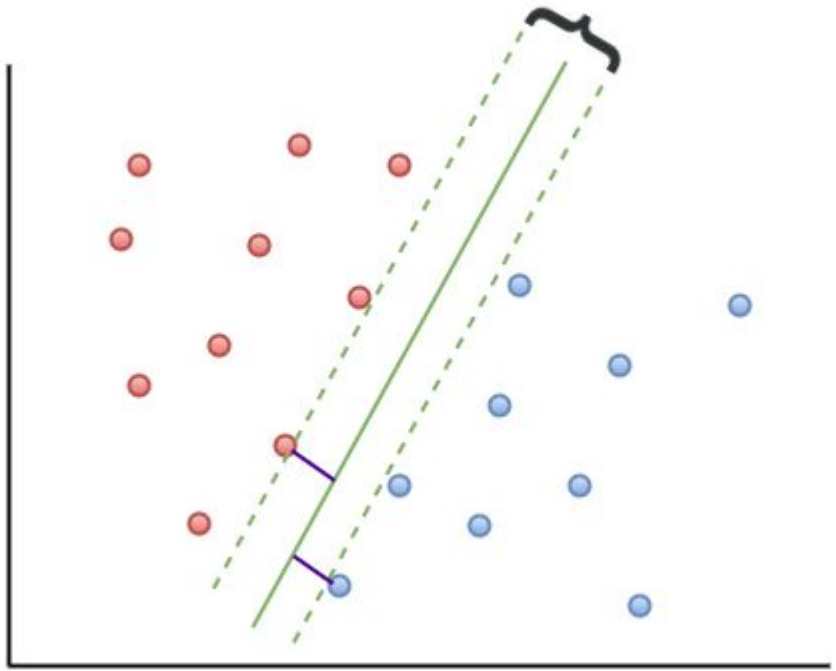


2 Dimensional Example

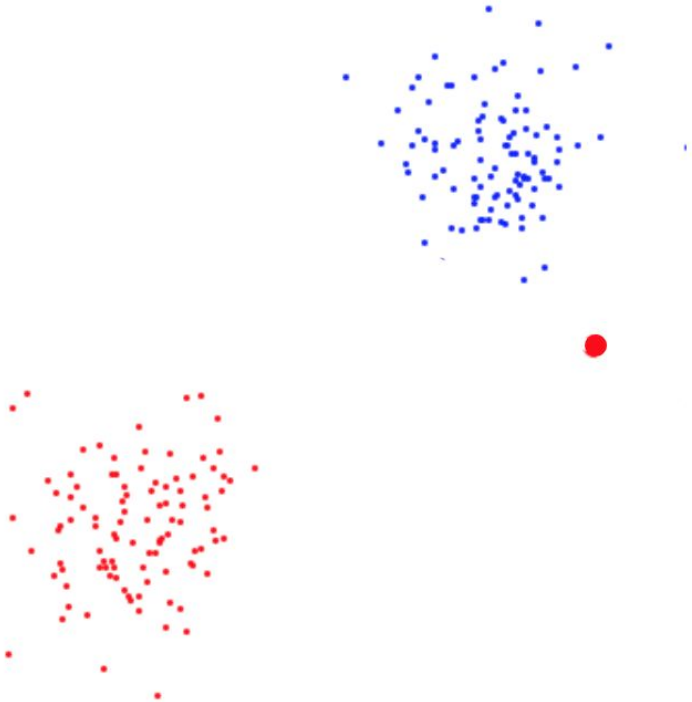
- The data points will be separated by a line

$$y = mx + b$$

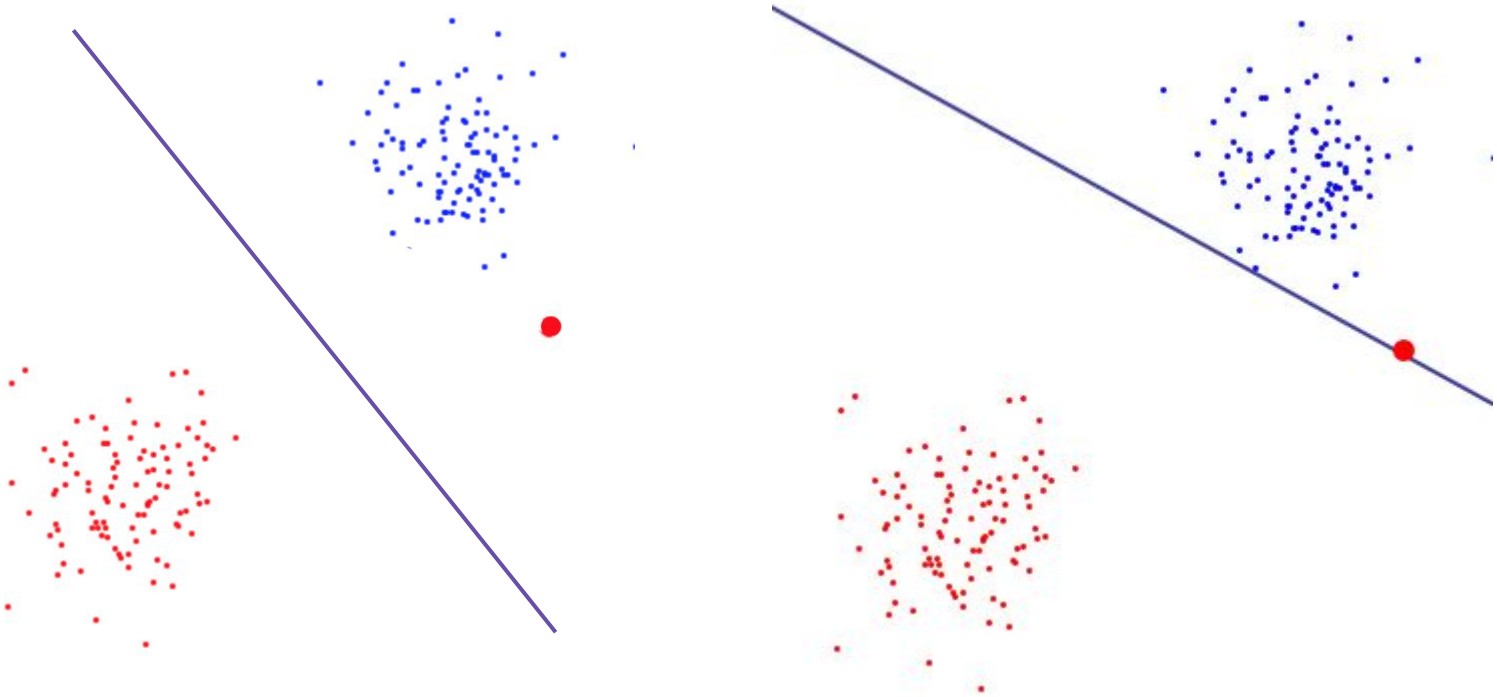
- Tweak parameters to find best line of separation



What if..

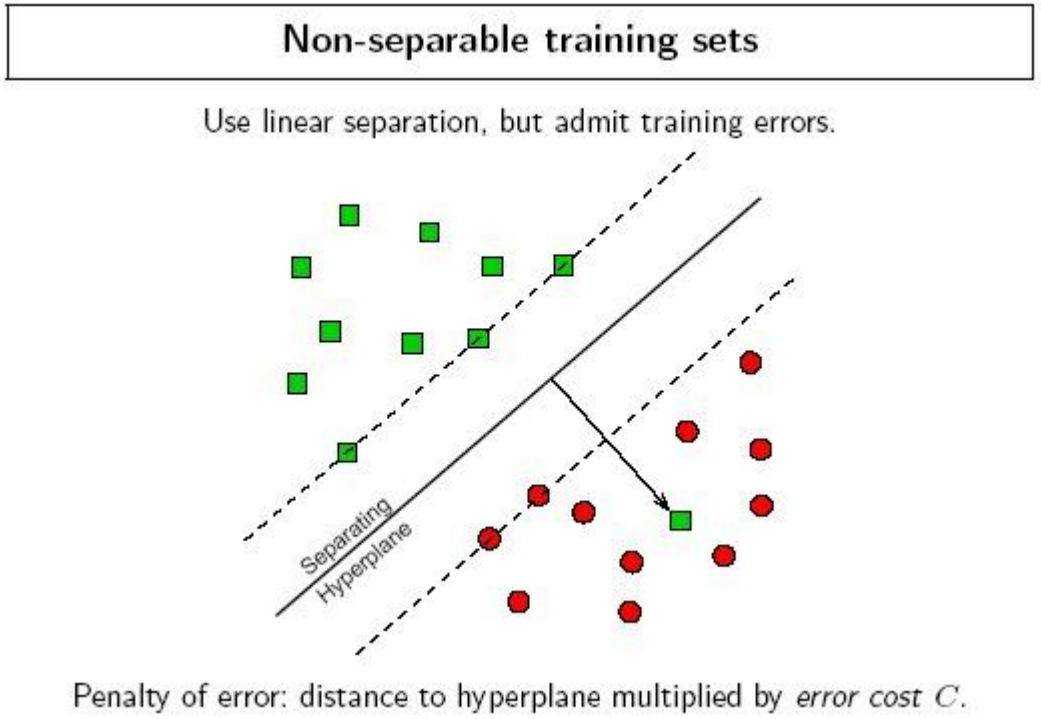


Which one is better



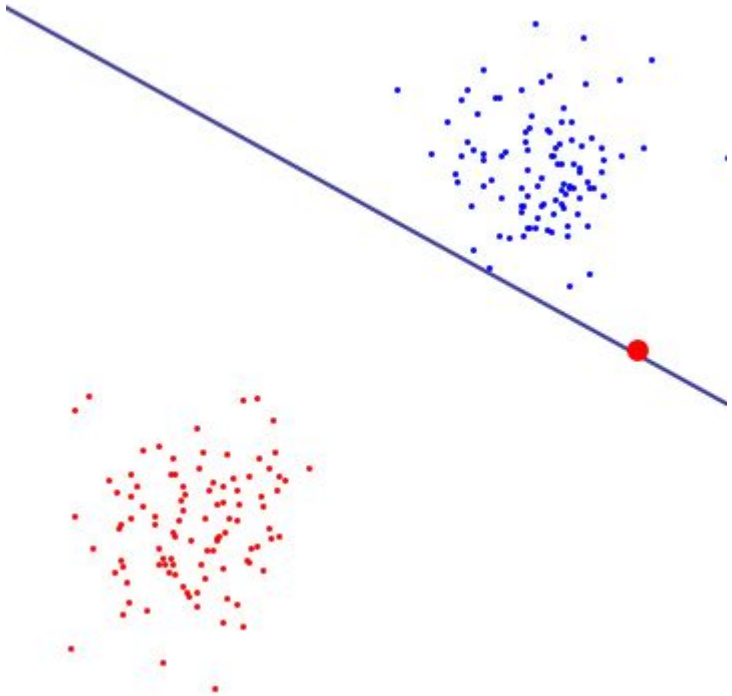
Margins

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



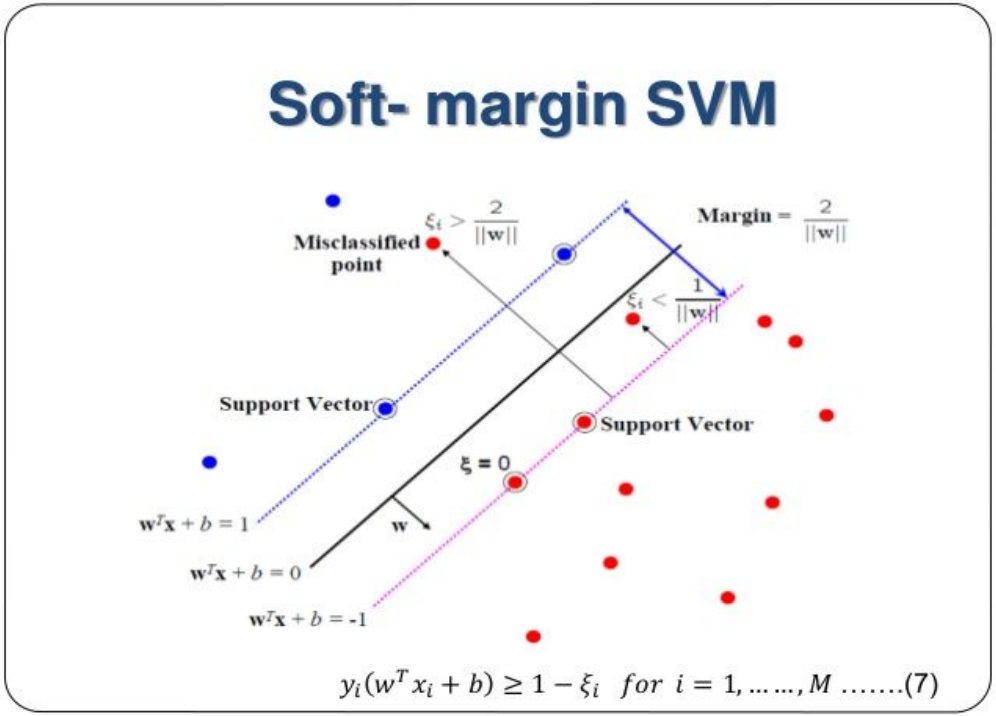
Hard Margins

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

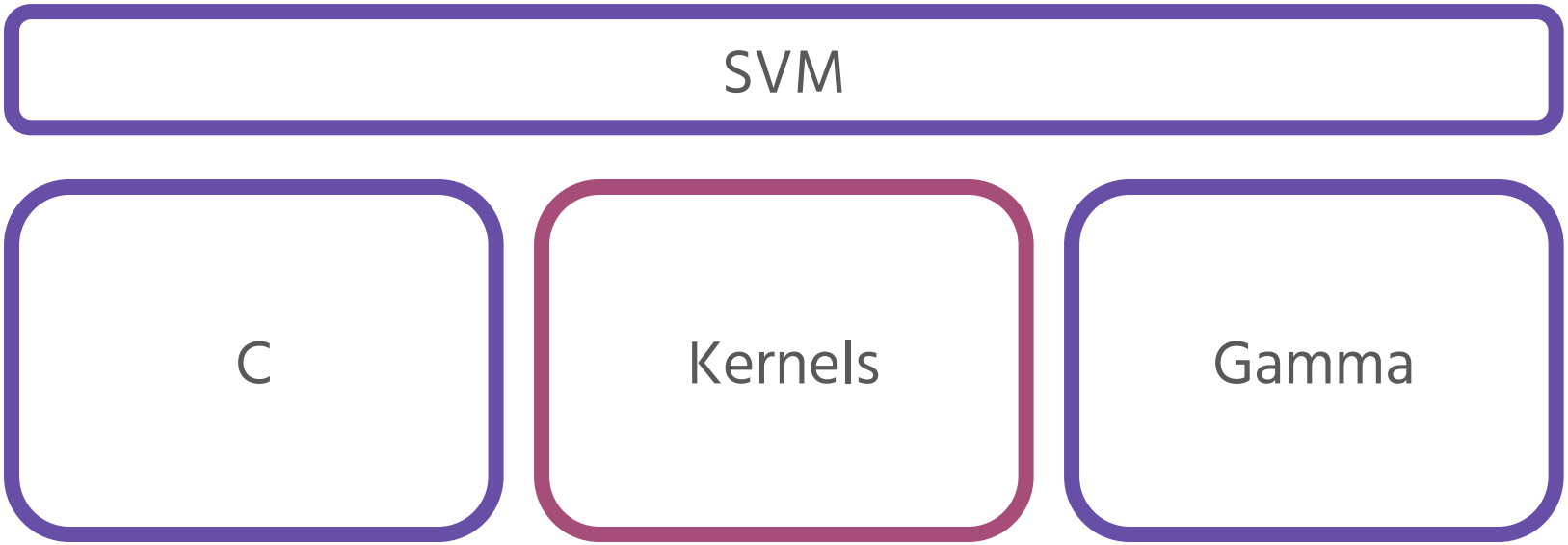


Soft Margins

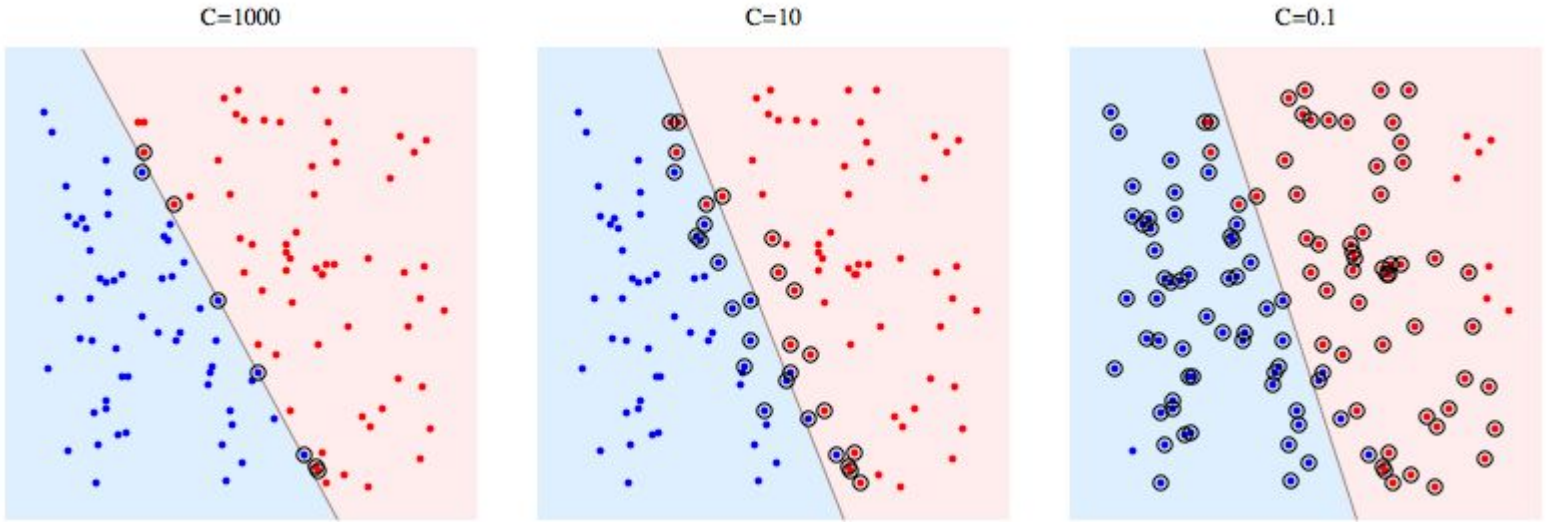
- Used in non-linearly separable datasets
- Allow for misclassification
- Can account for “dirty” boundaries



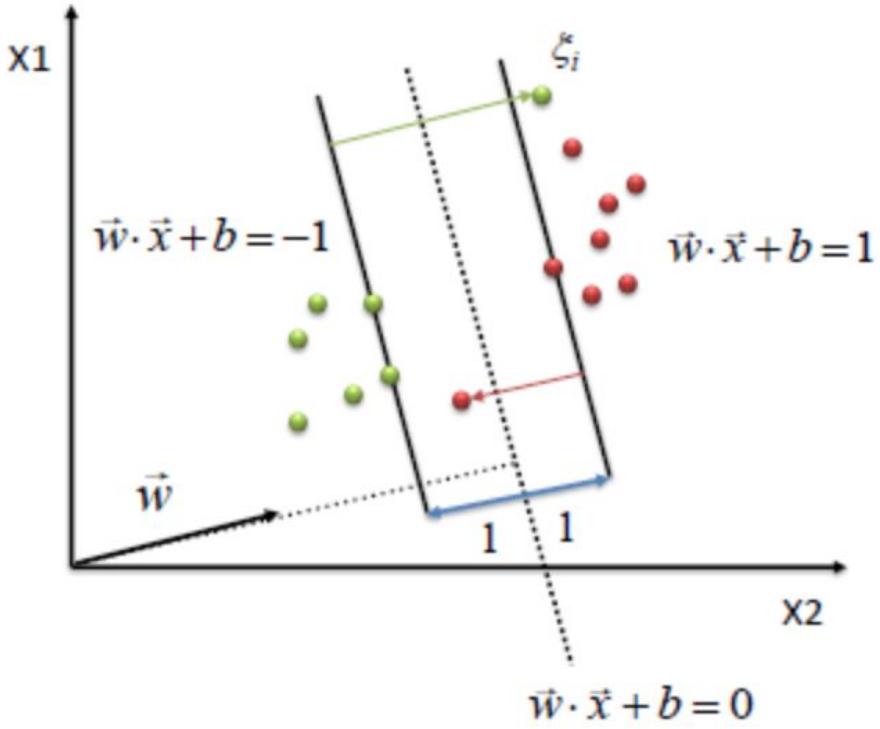
Hyper-Parameters



C Penalty

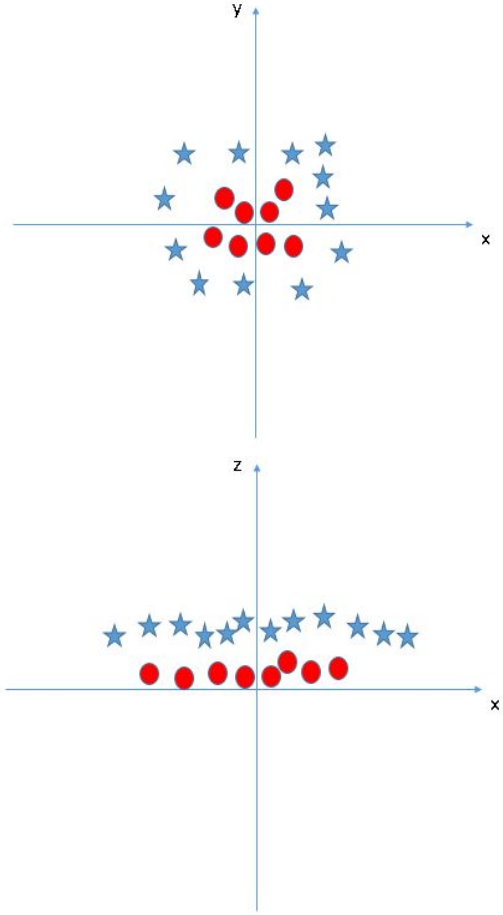


C Penalty


X1
ζ_i
$\vec{w}\cdot\vec{x}+b=-1$
$\vec{w}\cdot\vec{x}+b=1$
$\vec{w}$
1
1
X2
$\vec{w}\cdot\vec{x}+b=0$

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

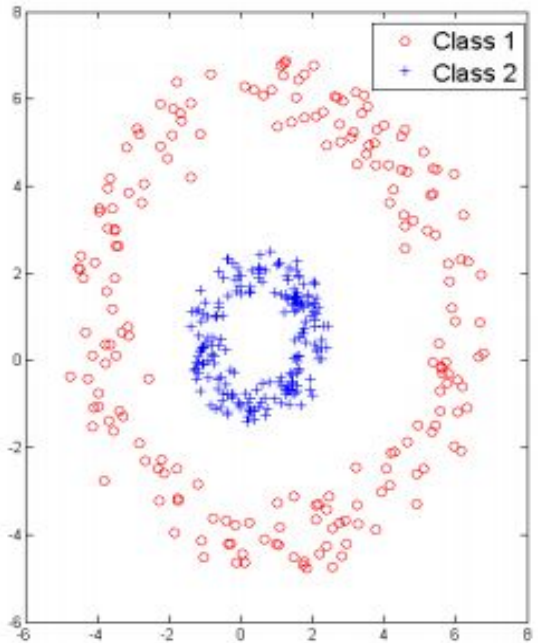
Poly

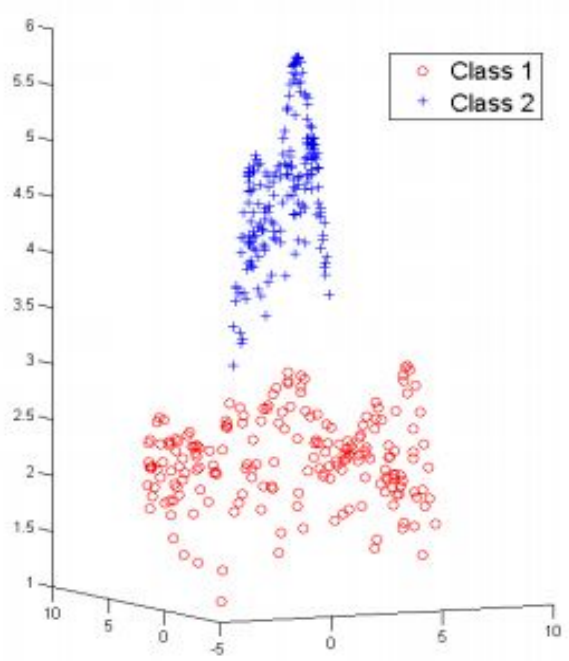
sigmoid

Pre-computed

etc

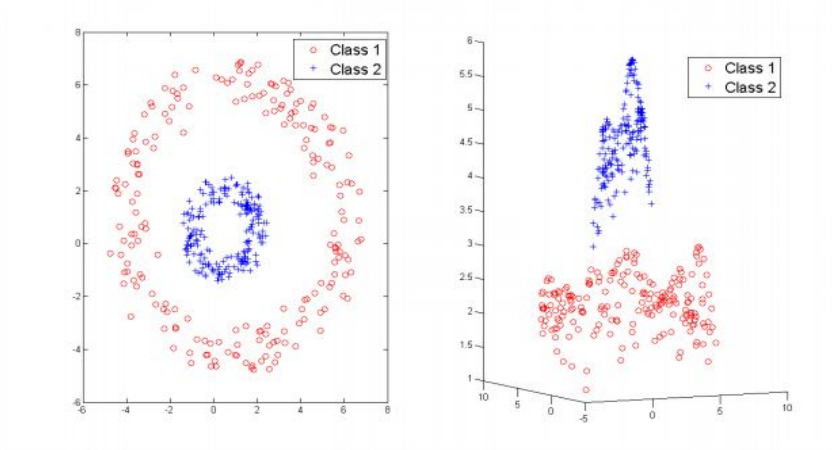
Gamma γ





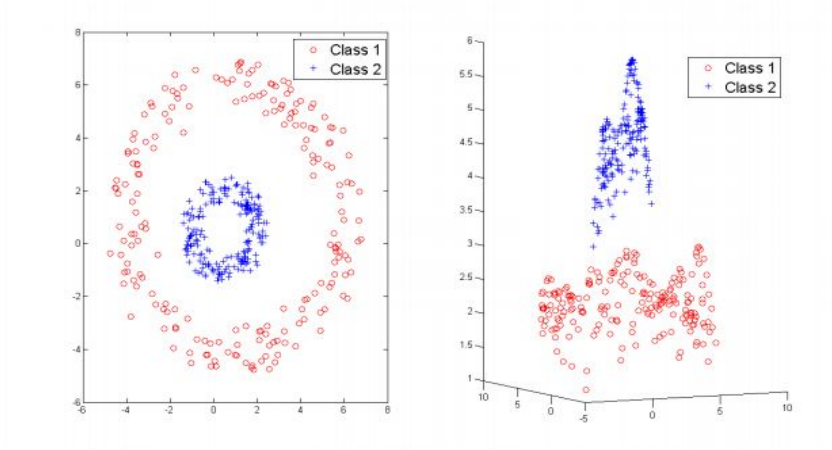
(Gaussian) Radial Basis Function (RBF)

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i K(x, x_i) = \sum_{i=1}^{N} \alpha_i y_i \exp(-\gamma \|x - x_i\|^2)$$

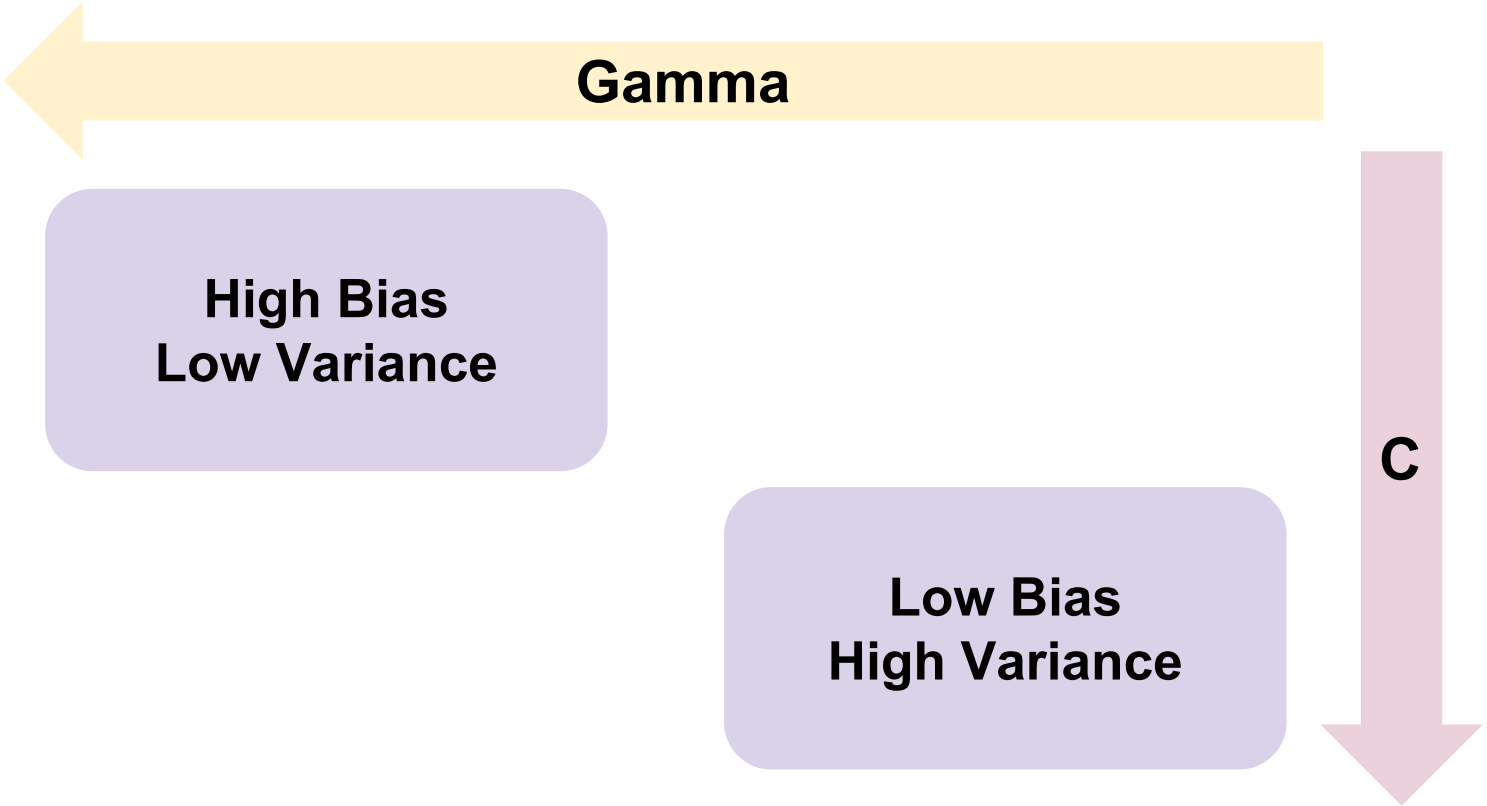


(Gaussian) Radial Basis Function (RBF)

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i K(x, x_i) = \sum_{i=1}^{N} \alpha_i y_i \exp(-\gamma \|x - x_i\|^2)$$

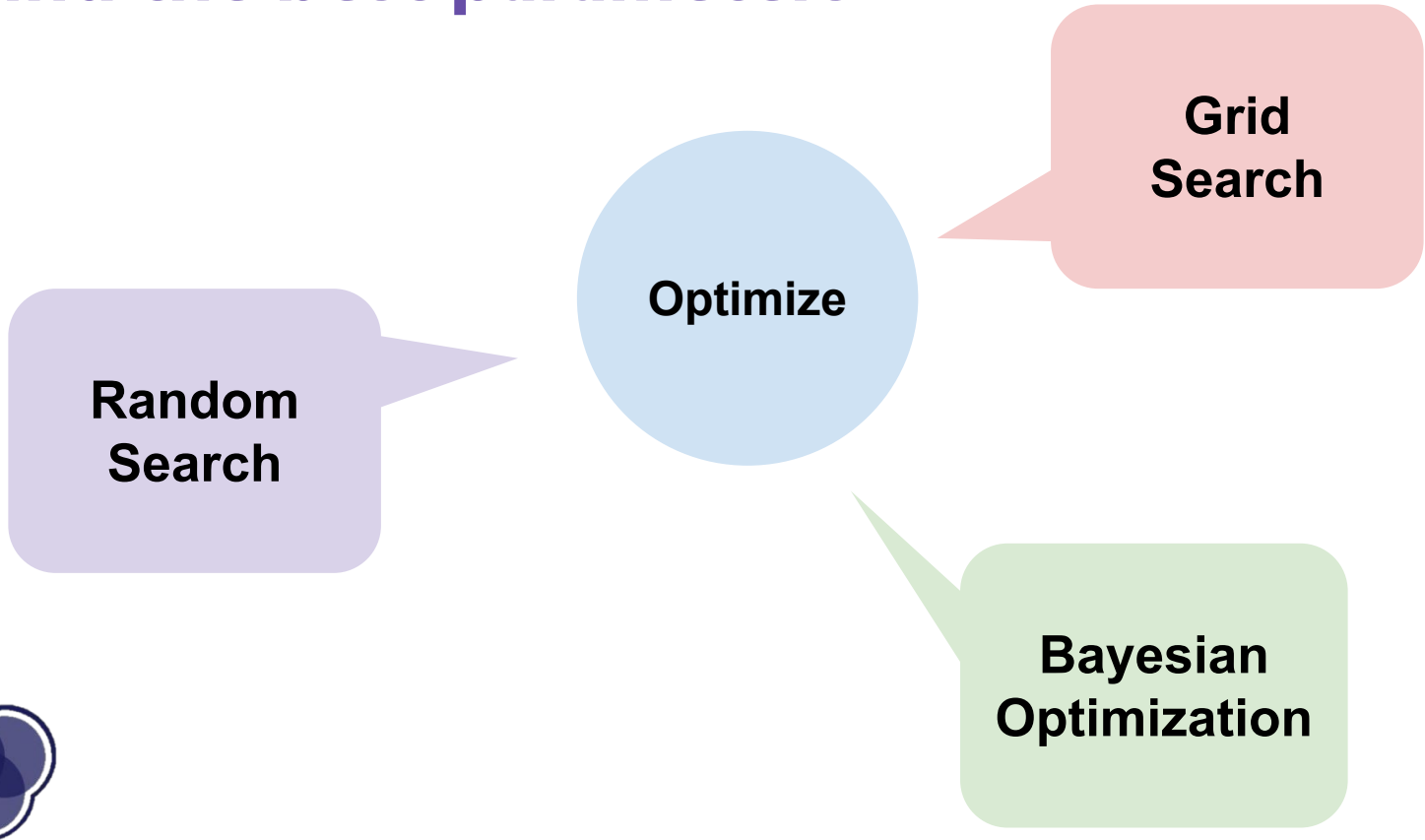


C and Gamma



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