Data Science 7

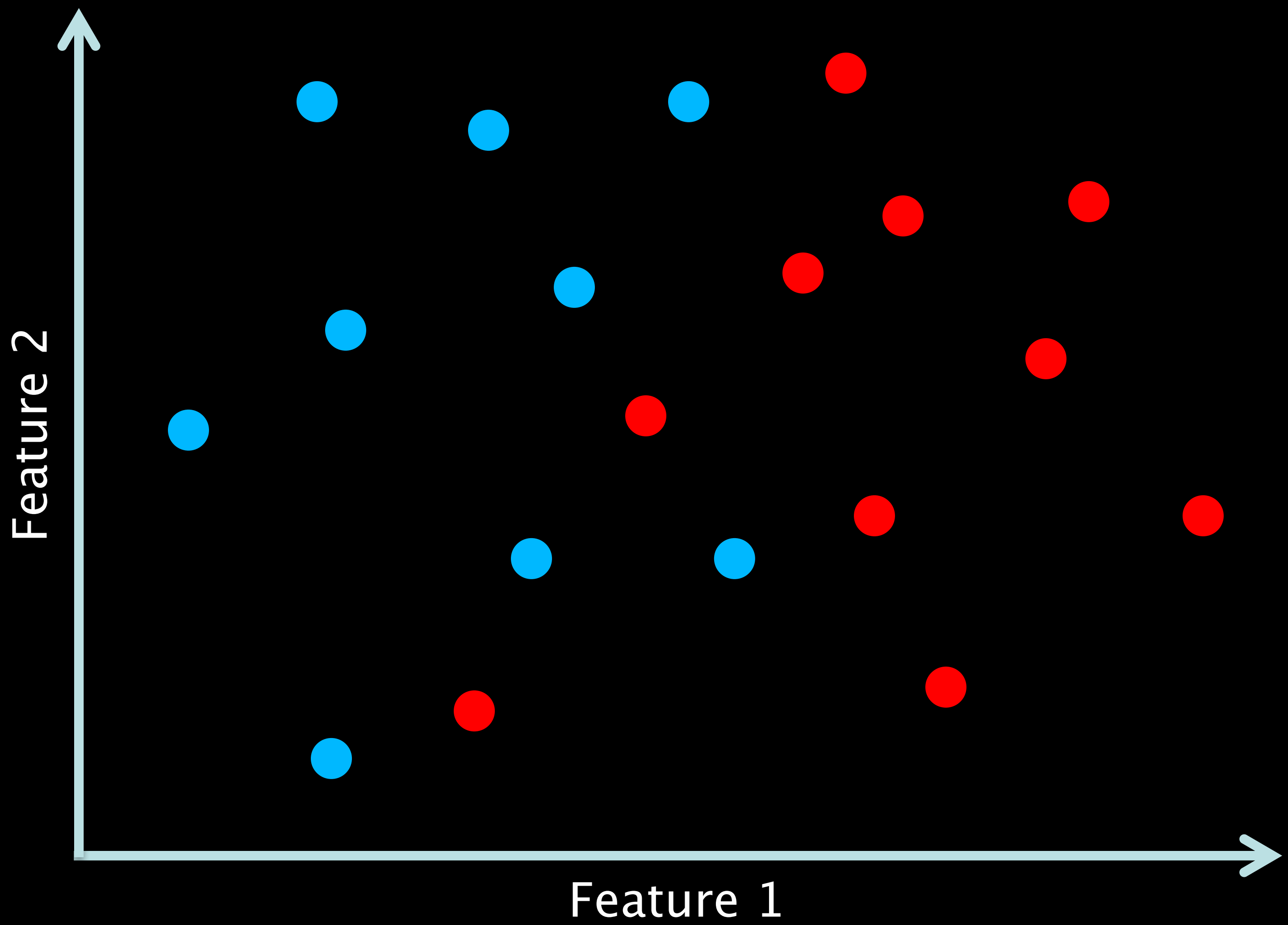
TILBURG UNIVERSITY

Understanding Society

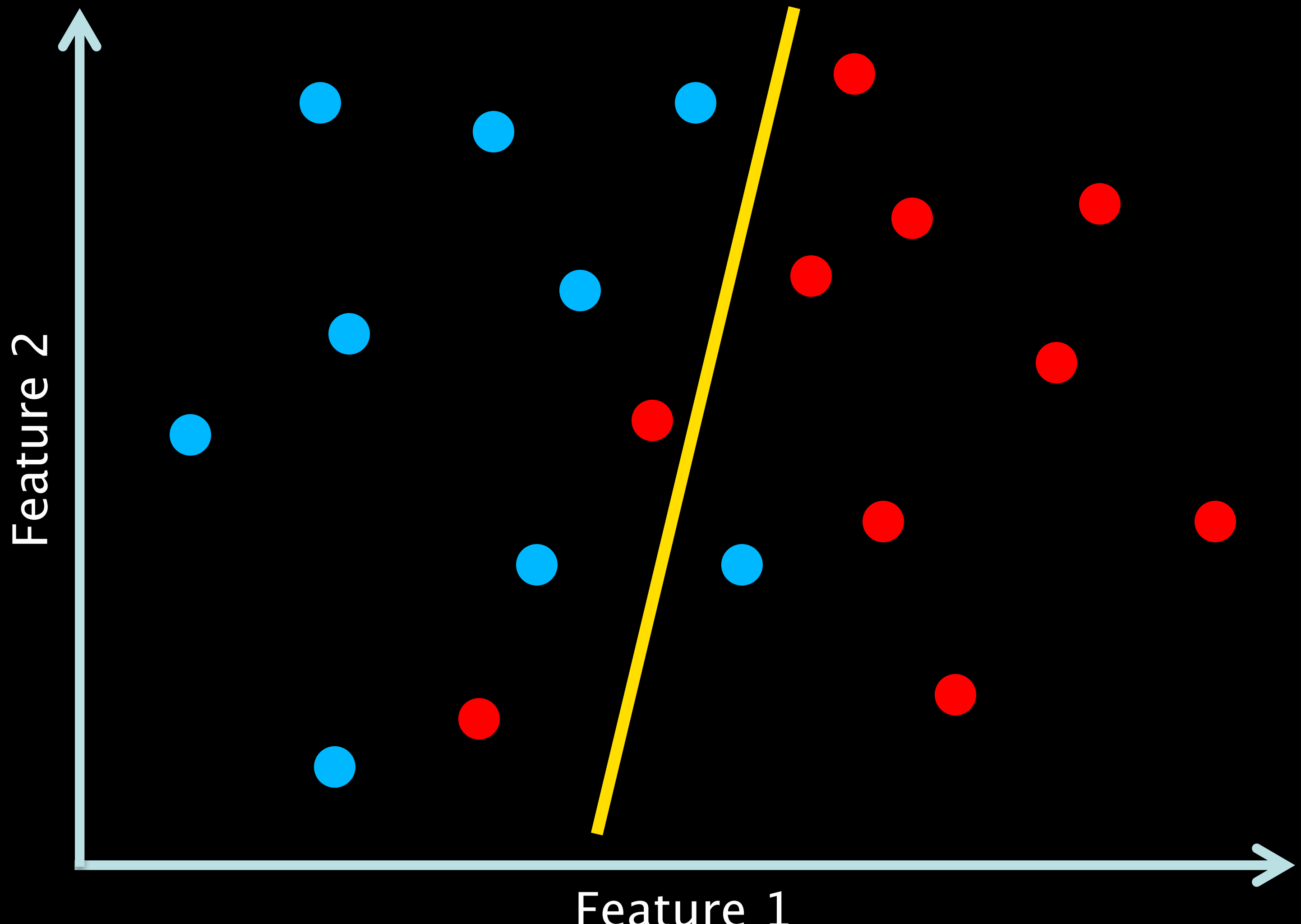
Overview

- Classification
- Model complexity = complexity of the decision boundary
- Support Vector Machines / Kernel Machines
- Kernels
- Orange

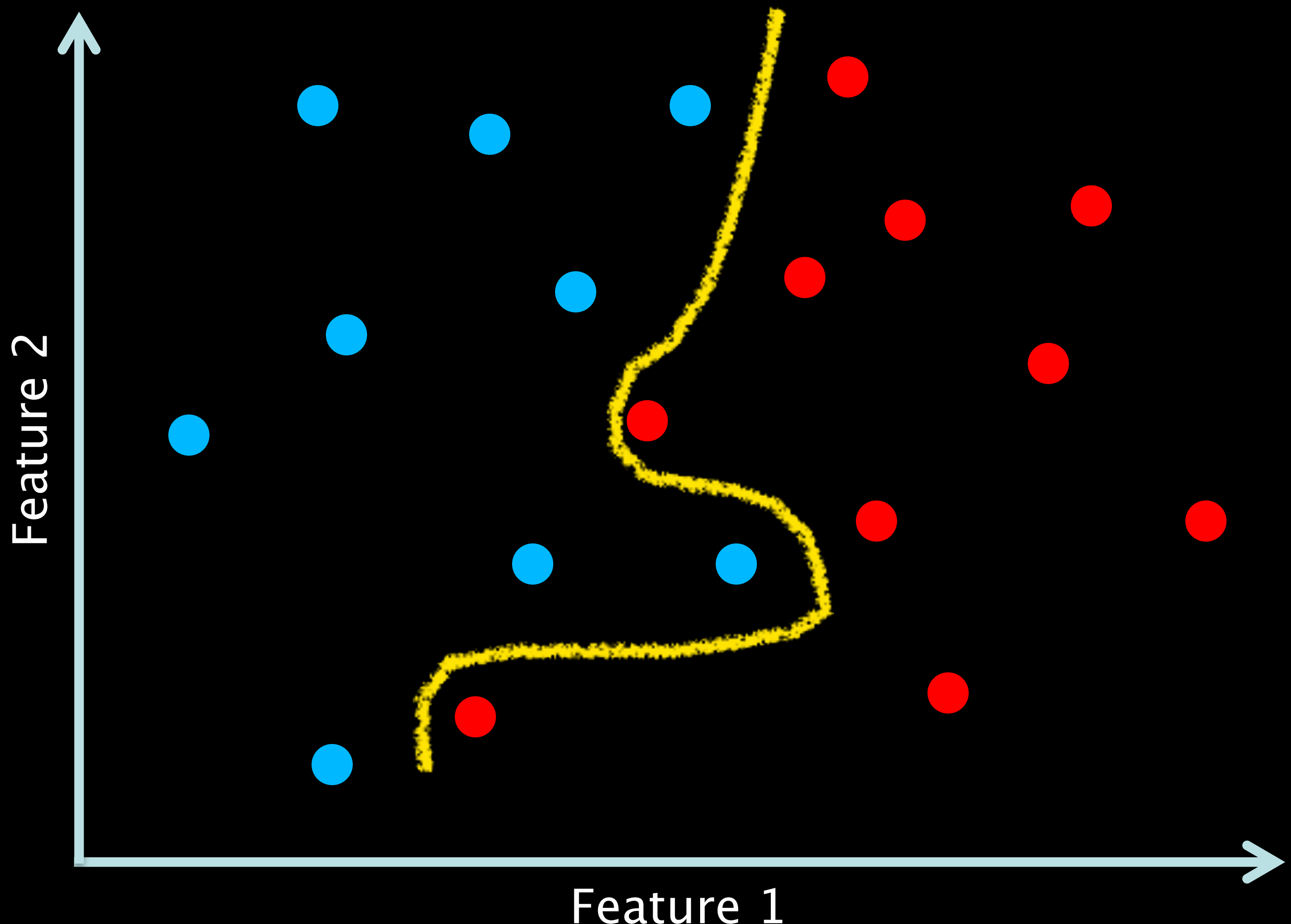
Two Classes (blue & red)



Separation by a simple model (— =decision boundary)

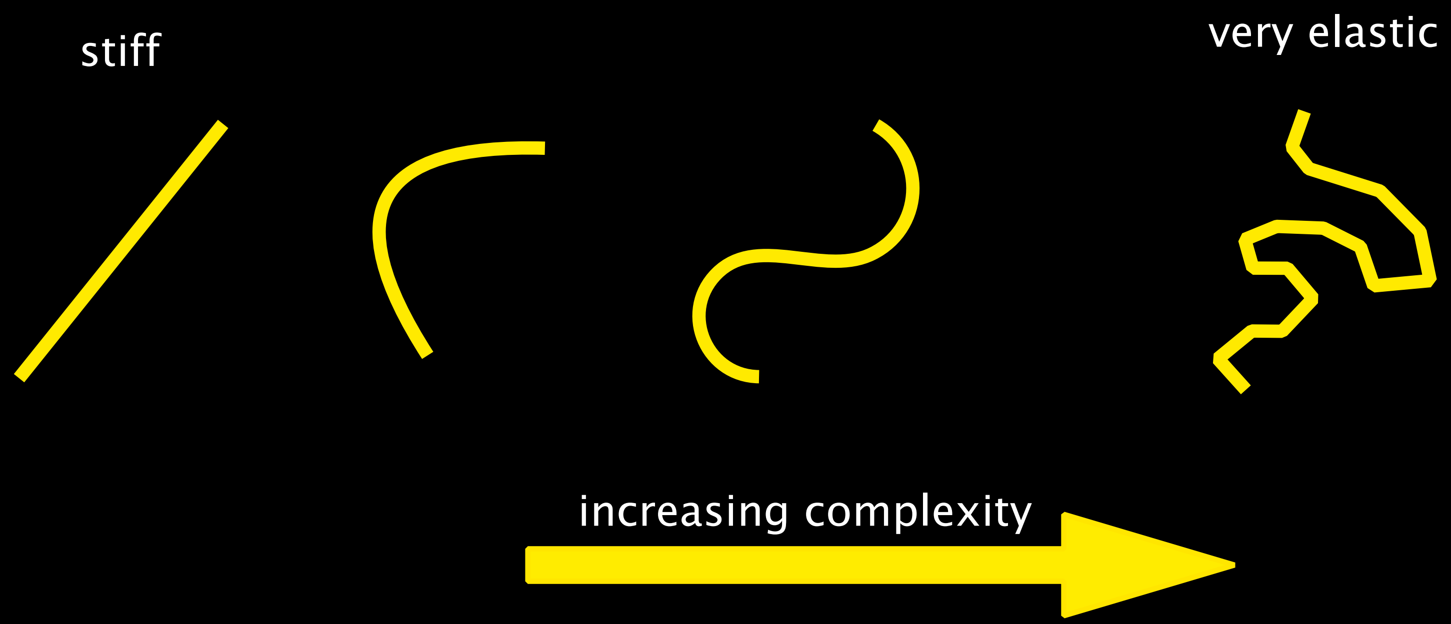


Separation by a complex model (– =decision boundary)



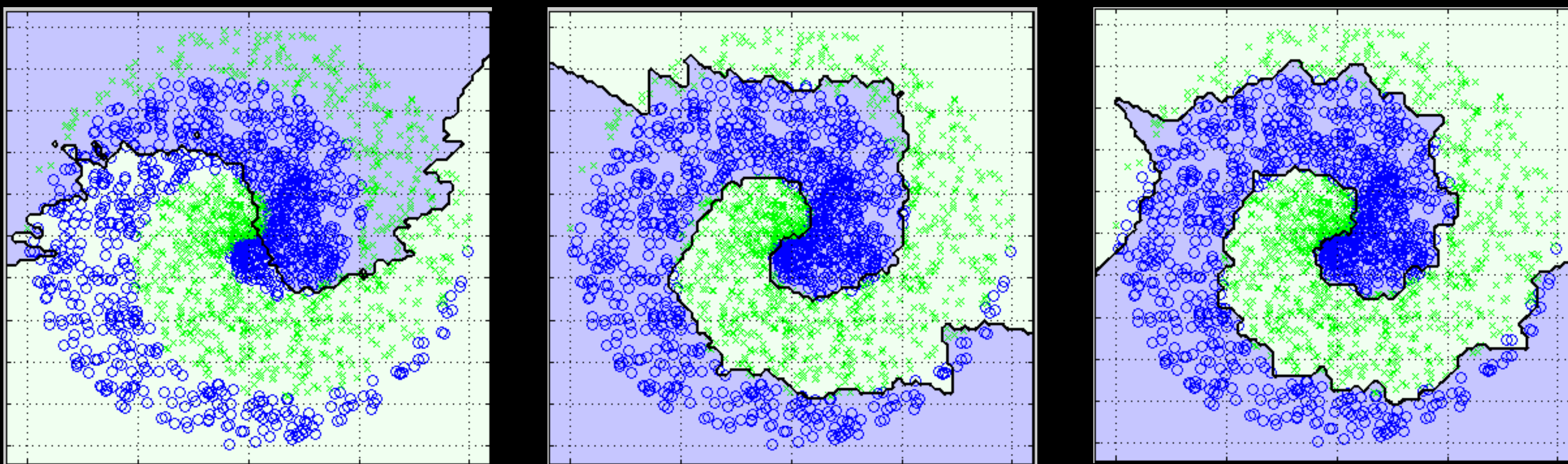
In classification…

- The complexity of a model is the flexibility of the decision boundary



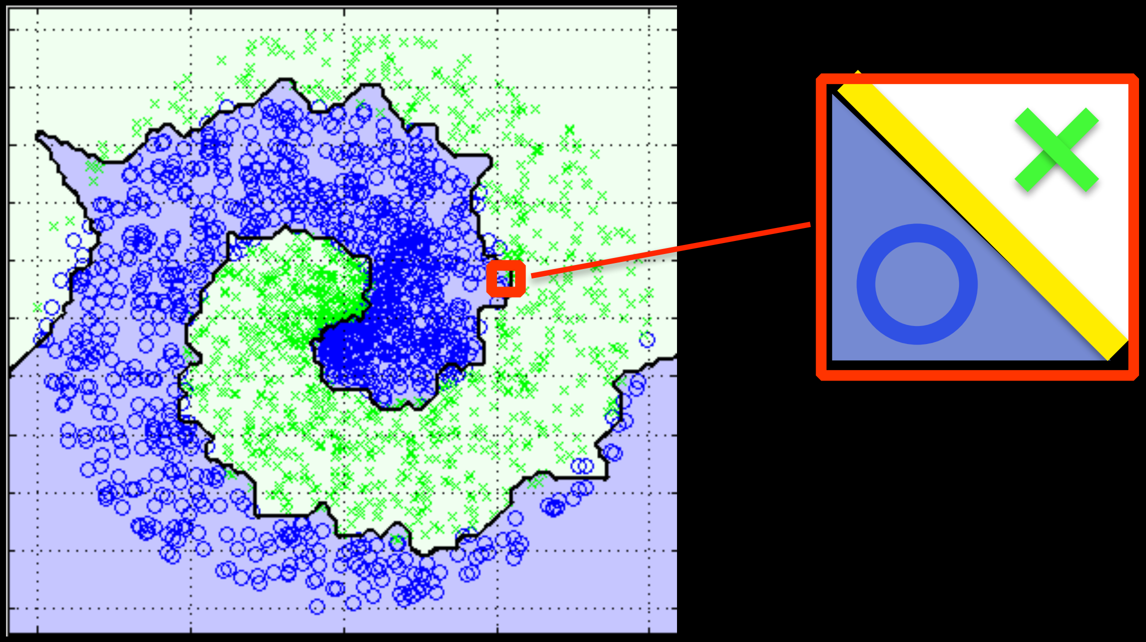
How to determine model complexity?

- Depends on complexity of the separation between the classes
- Start with the simplest model (large k in kNN), and increase complexity (smaller k)



k = 100 k = 10 k = 1

Increasing model complexity

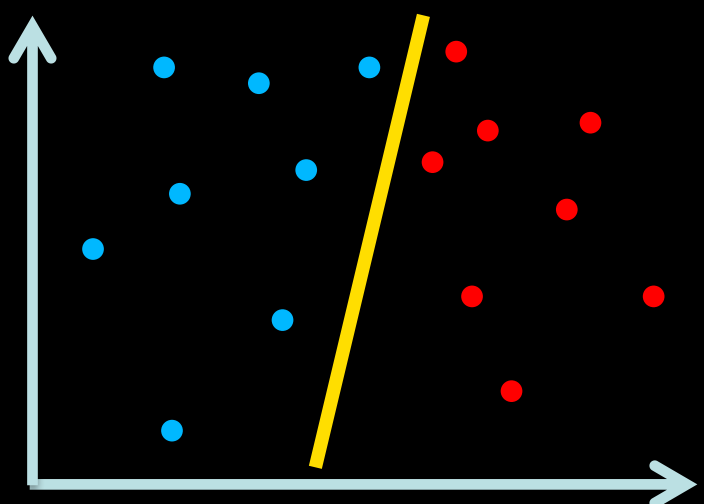


In kNN, the model (decision boundary) is defined by all instances in the training set. That is quite expensive. Can't we represent the boundary by a few instances only?

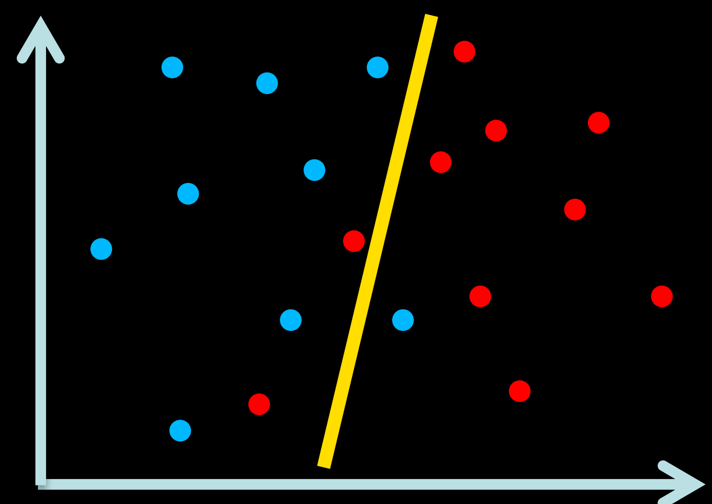
Support Vector Machines
a.k.a. Kernel Machines

Simplest SVM

- Linear SVM
- Places a straight line between the classes (simplest model)



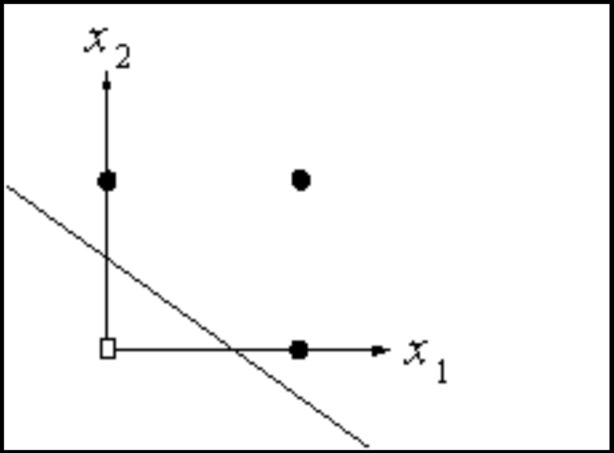
the instances are linearly separable



the instances are **not** linearly separable

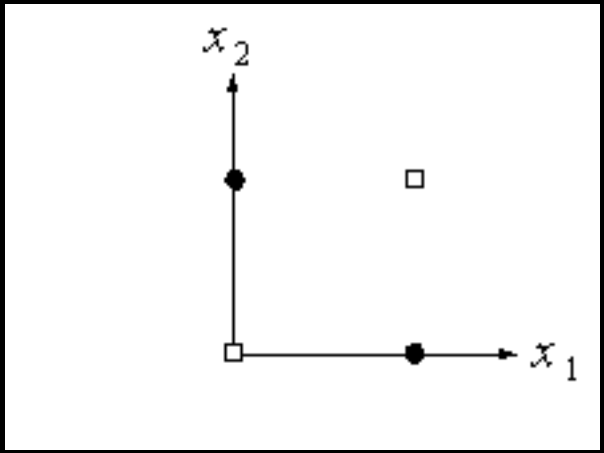
Linearly separable

The instances of the two classes can be separated by a straight decision boundary (OR problem)

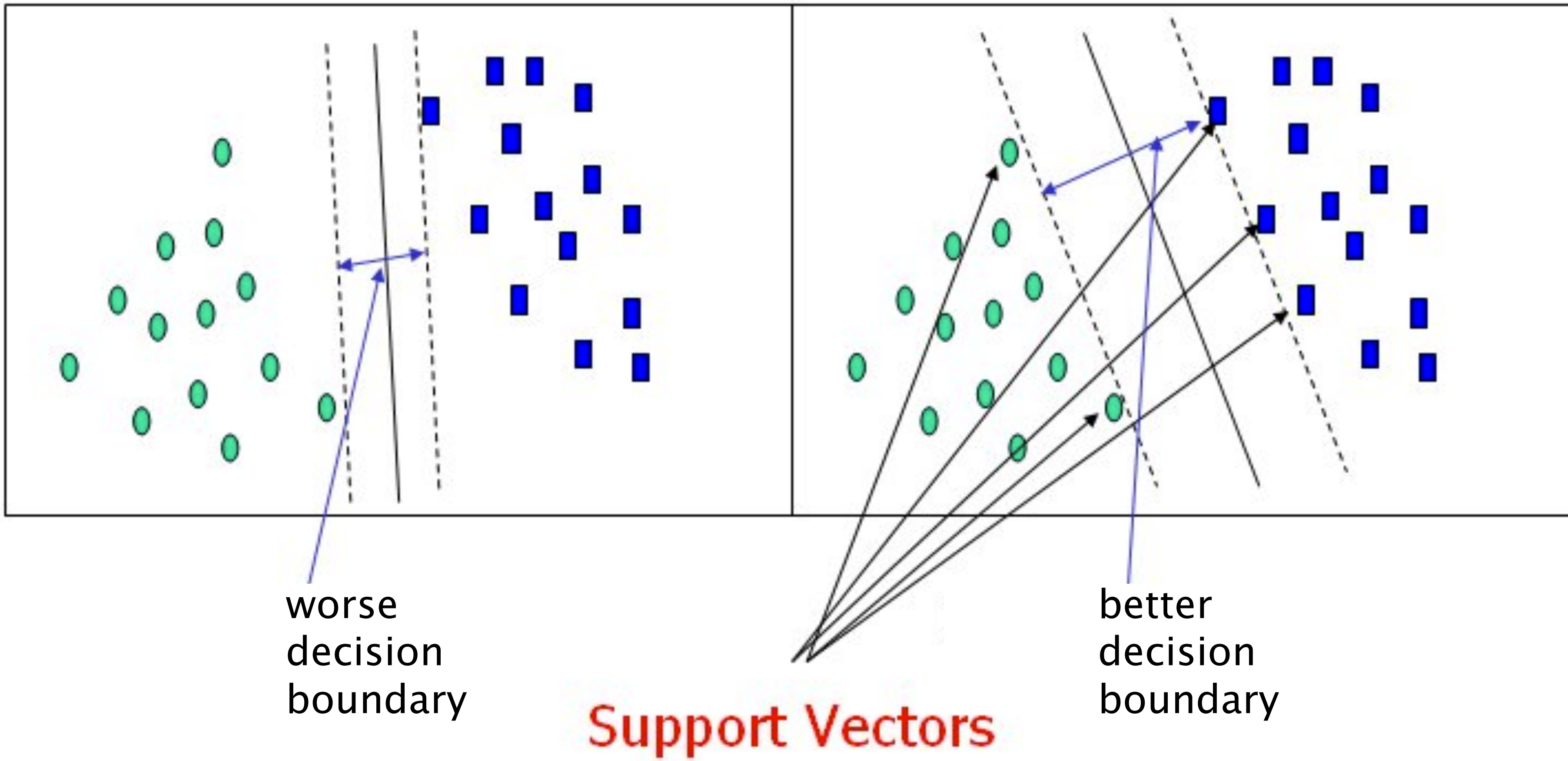


Linearly inseparable

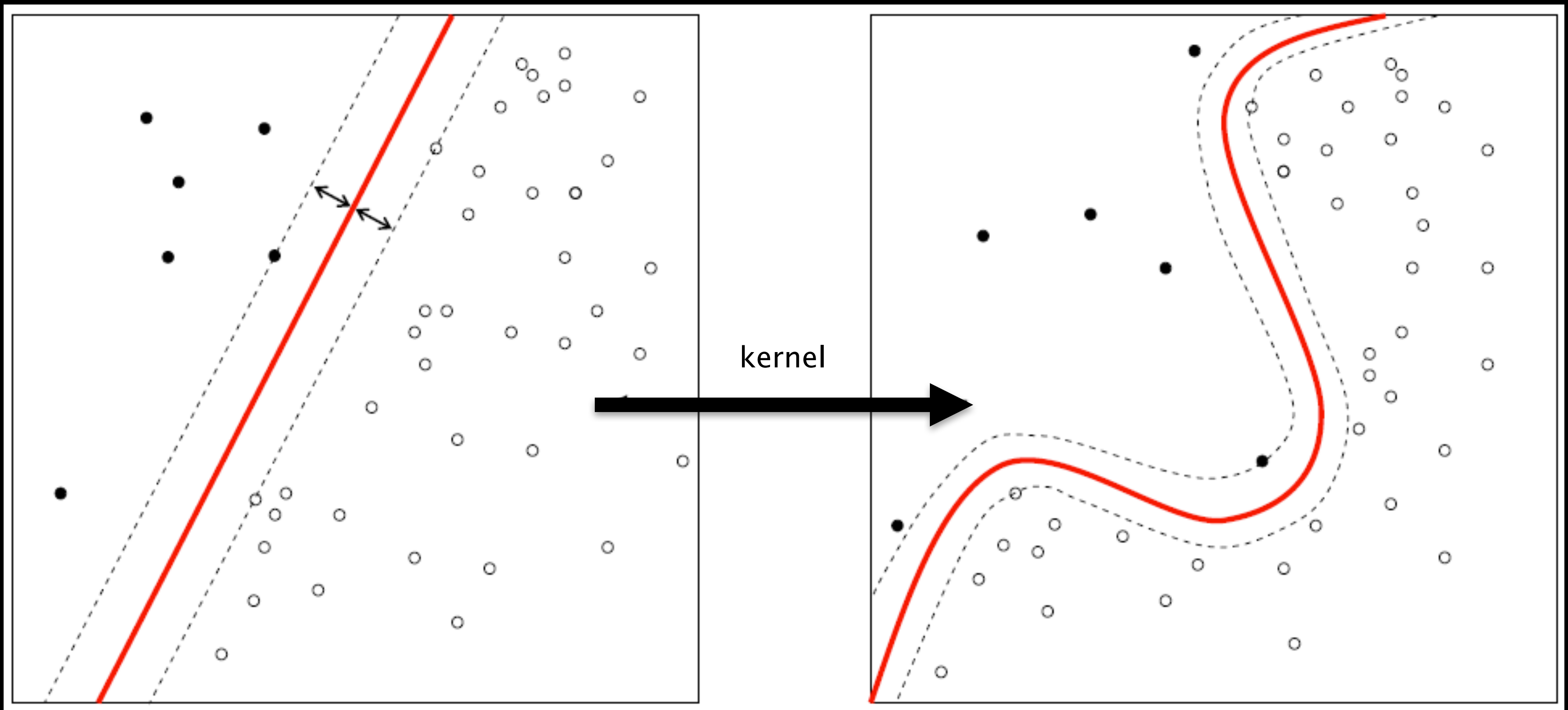
The instances of the two classes can **not** be separated by a straight decision boundary (XOR problem)



Why “Support Vector Machine”?

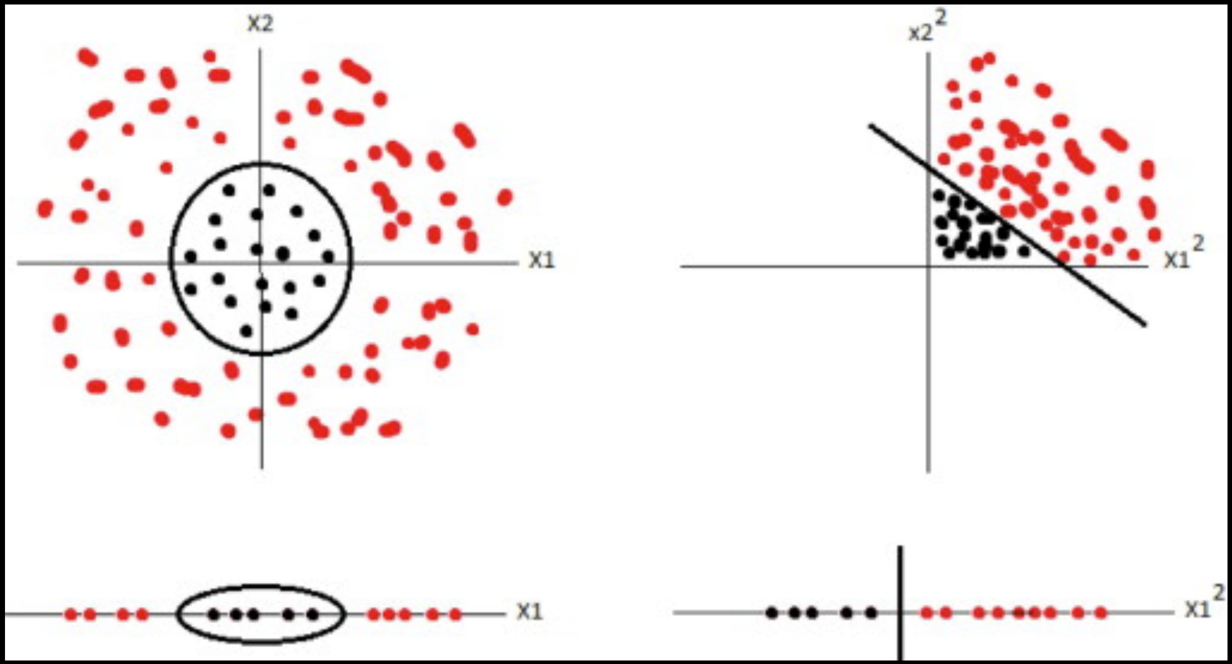


SVMs can make more complex decision boundaries by using kernels



Kernel

- A kernel performs a mathematical operation on instances in two classes that are **not** linearly separable, so that they do become linearly separable



Kernel = Similarity Measure

- A kernel takes pairs of instances and determines their similarity\
- There are many kernels
- The proper choice of a kernel depends on the domain
- The "linear kernel" (actually no kernel) is the simplest version
- The linear kernel should be used especially when you have many features (recall the previous lecture)

Kernels

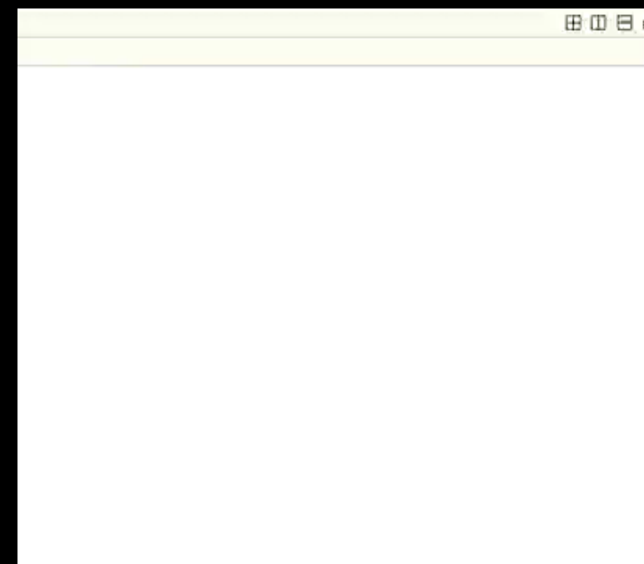
- Linear
- Polynomial
- Radial Basis Function
- Sigmoid

SVM with a polynomial
Kernel visualization

Created by:
Udi Aharoni

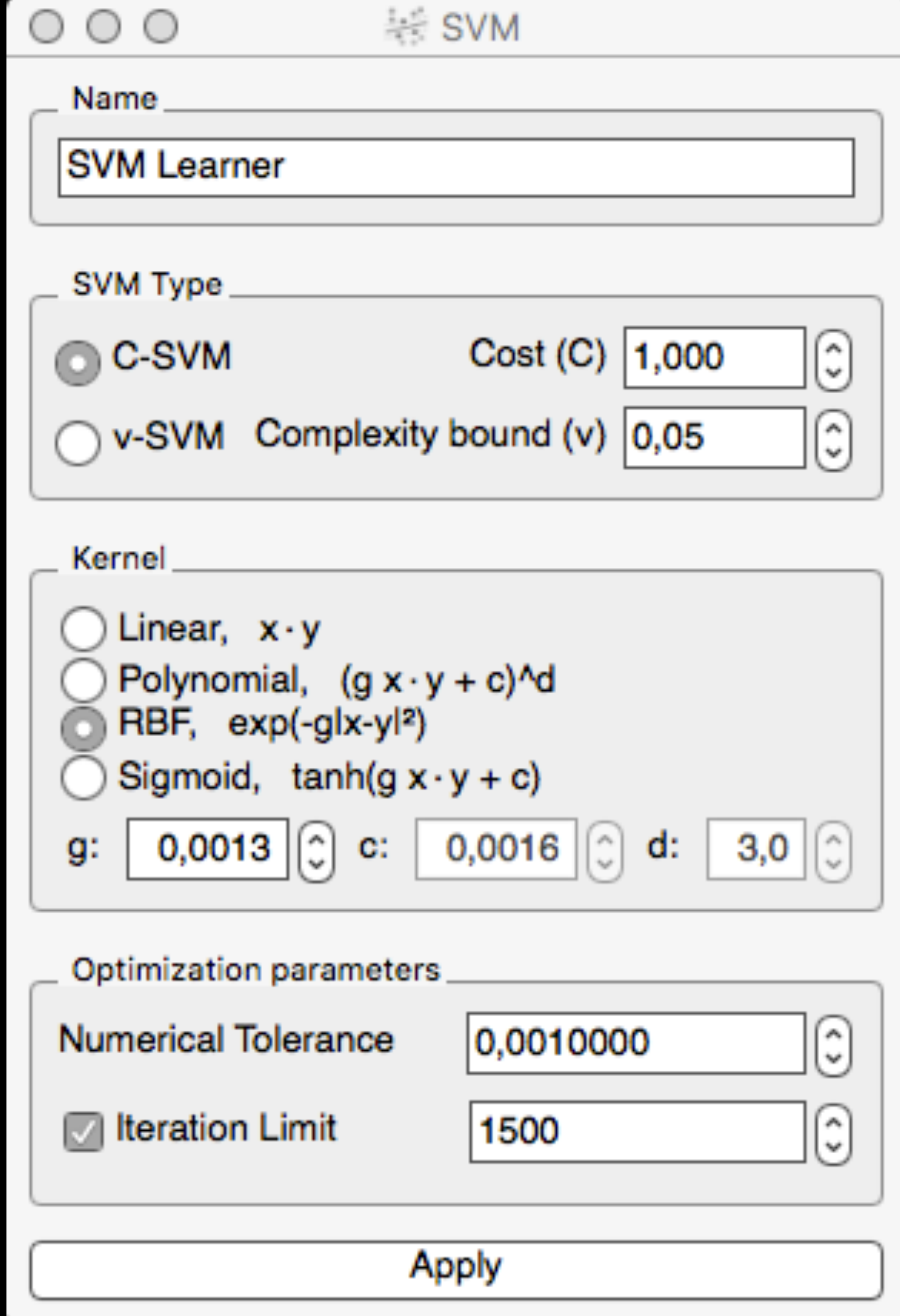
Kernels

- Linear
- Polynomial
- Radial Basis Function
- Sigmoid



Orange

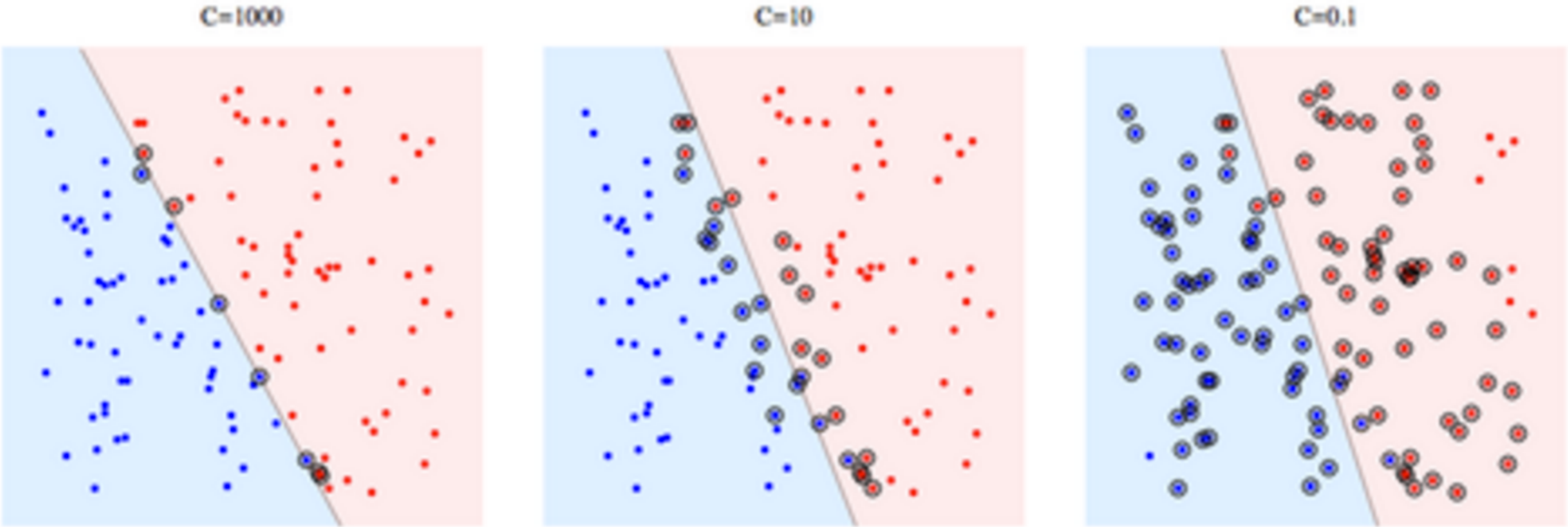
TILBURG UNIVERSITY



Cost of misclassification

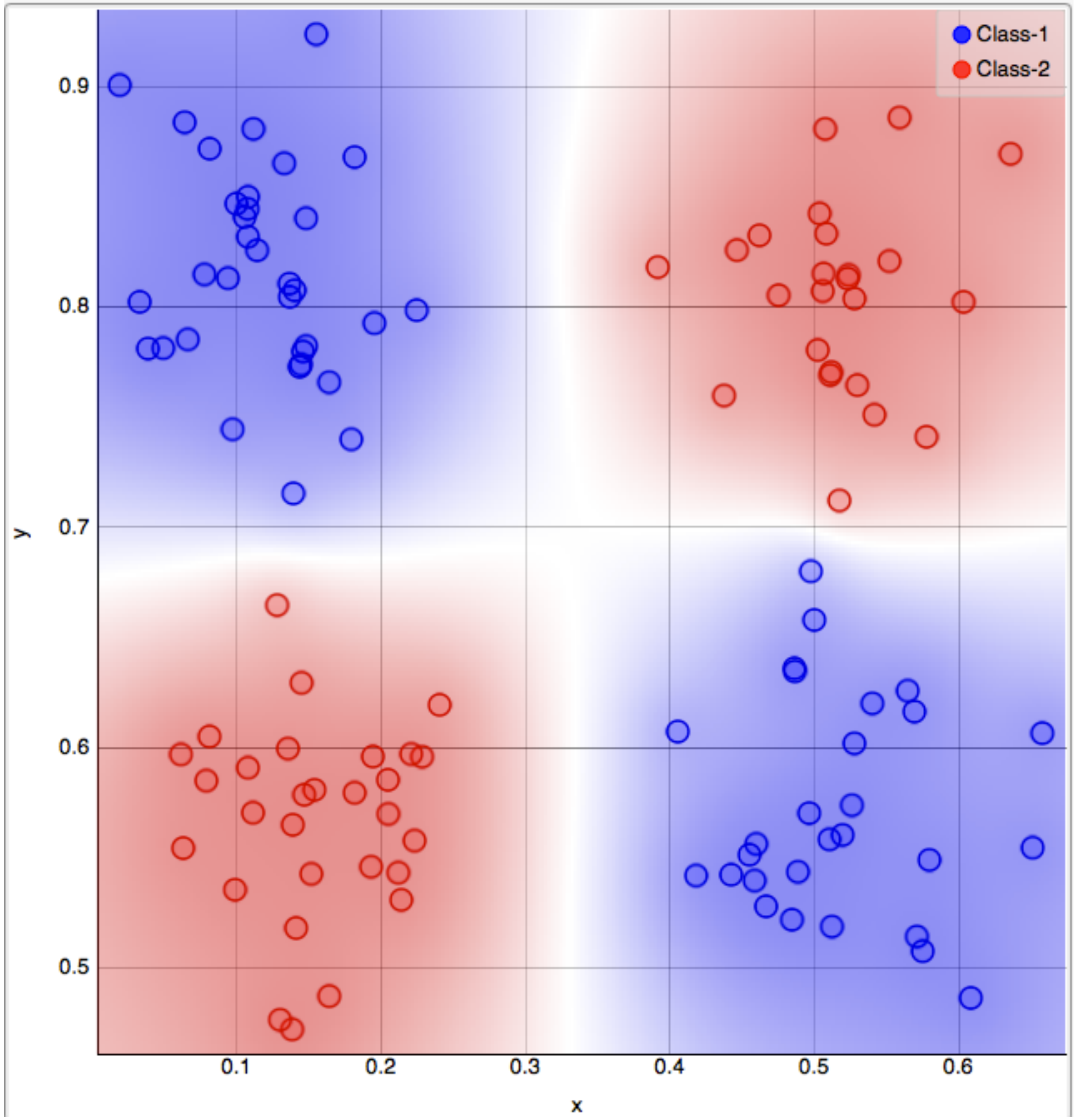
Number of support vectors

C = Cost parameter (misclassification cost)



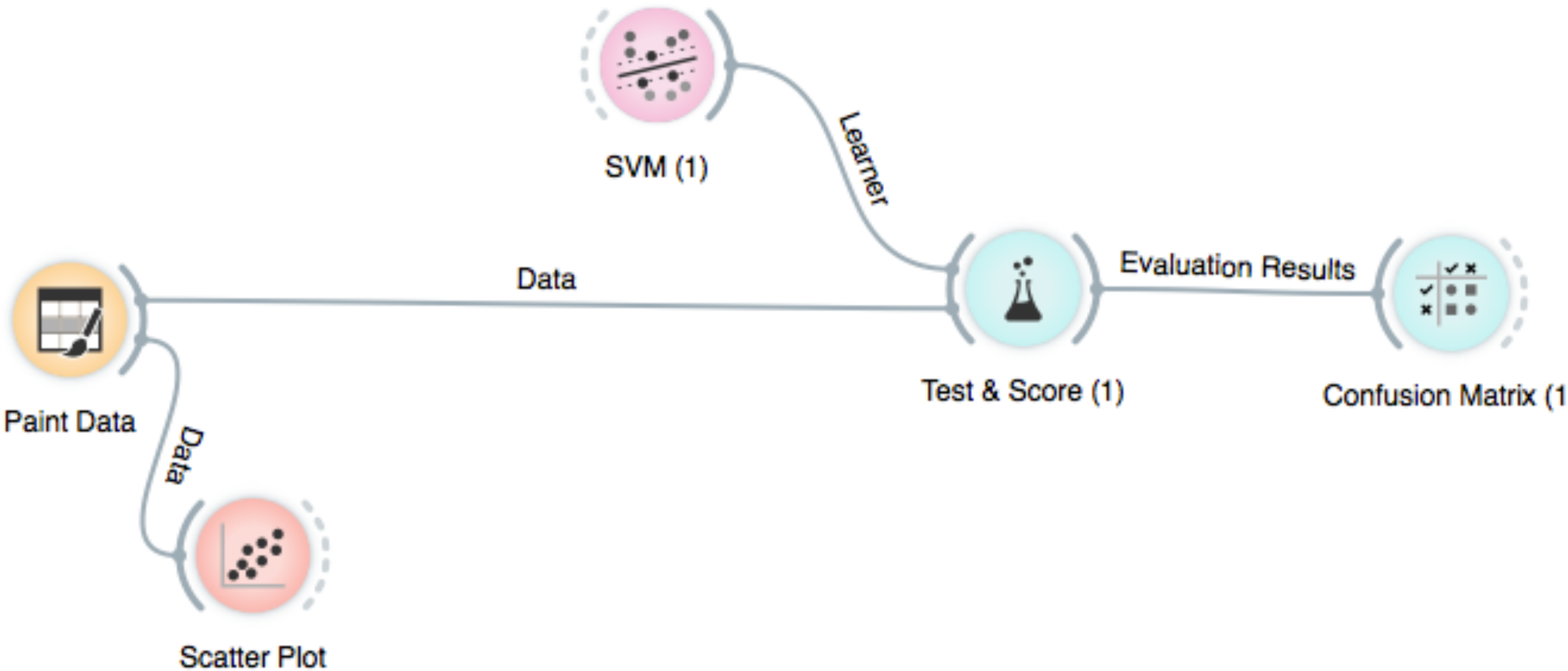
https://www.quora.com/What-are-C-and-gamma-with-regards-to-a-support-vector-machine

XOR dataset (create using Paint Data)

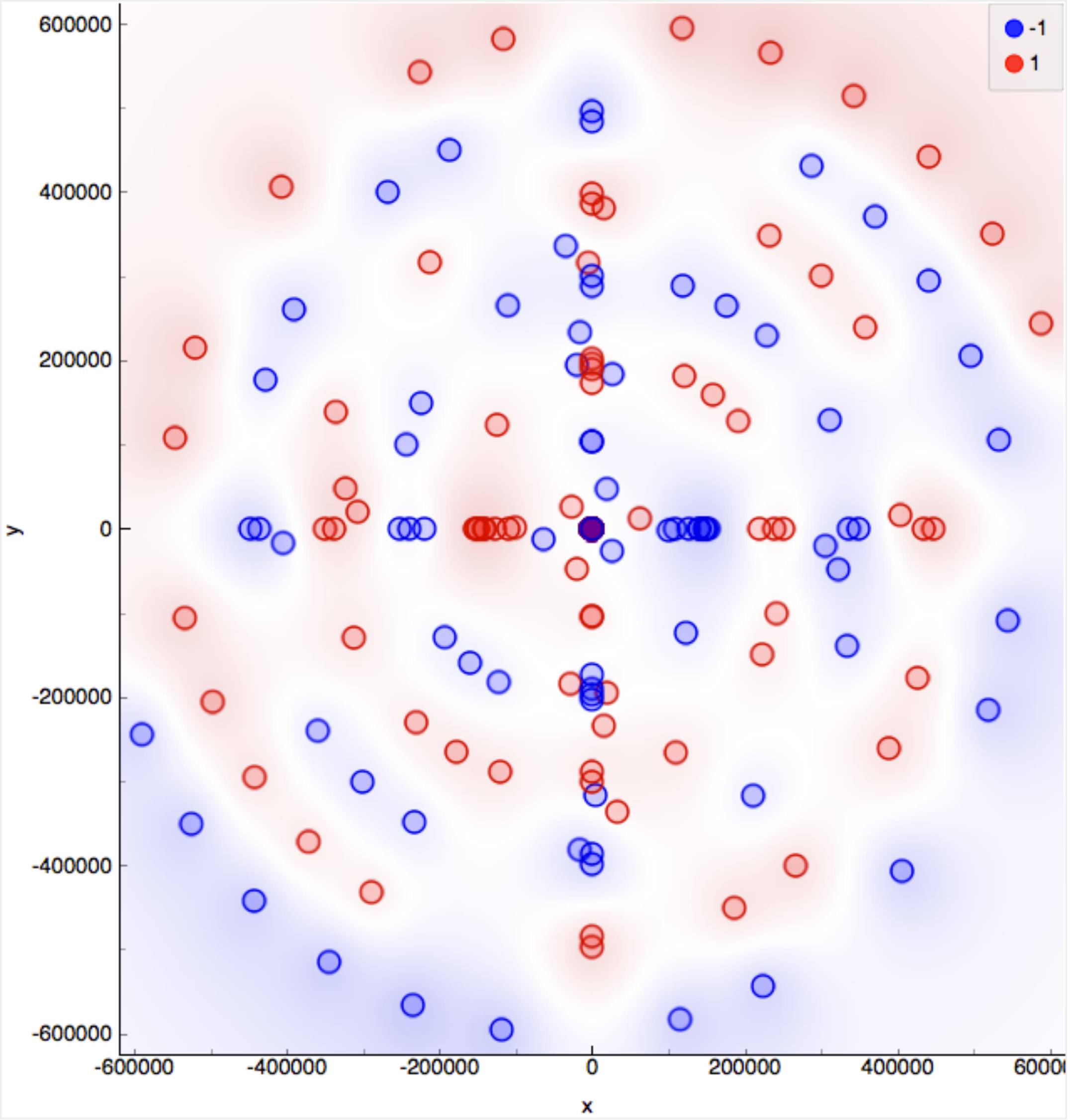


Tilburg University

Experimenting with SVMs (XOR dataset)

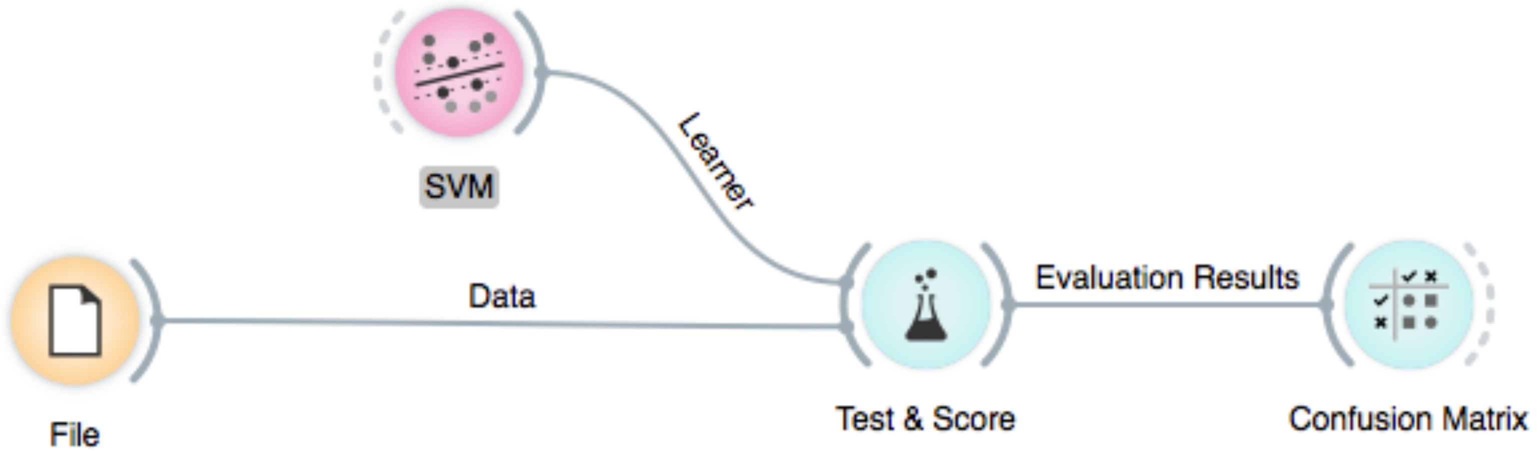


Spirals dataset

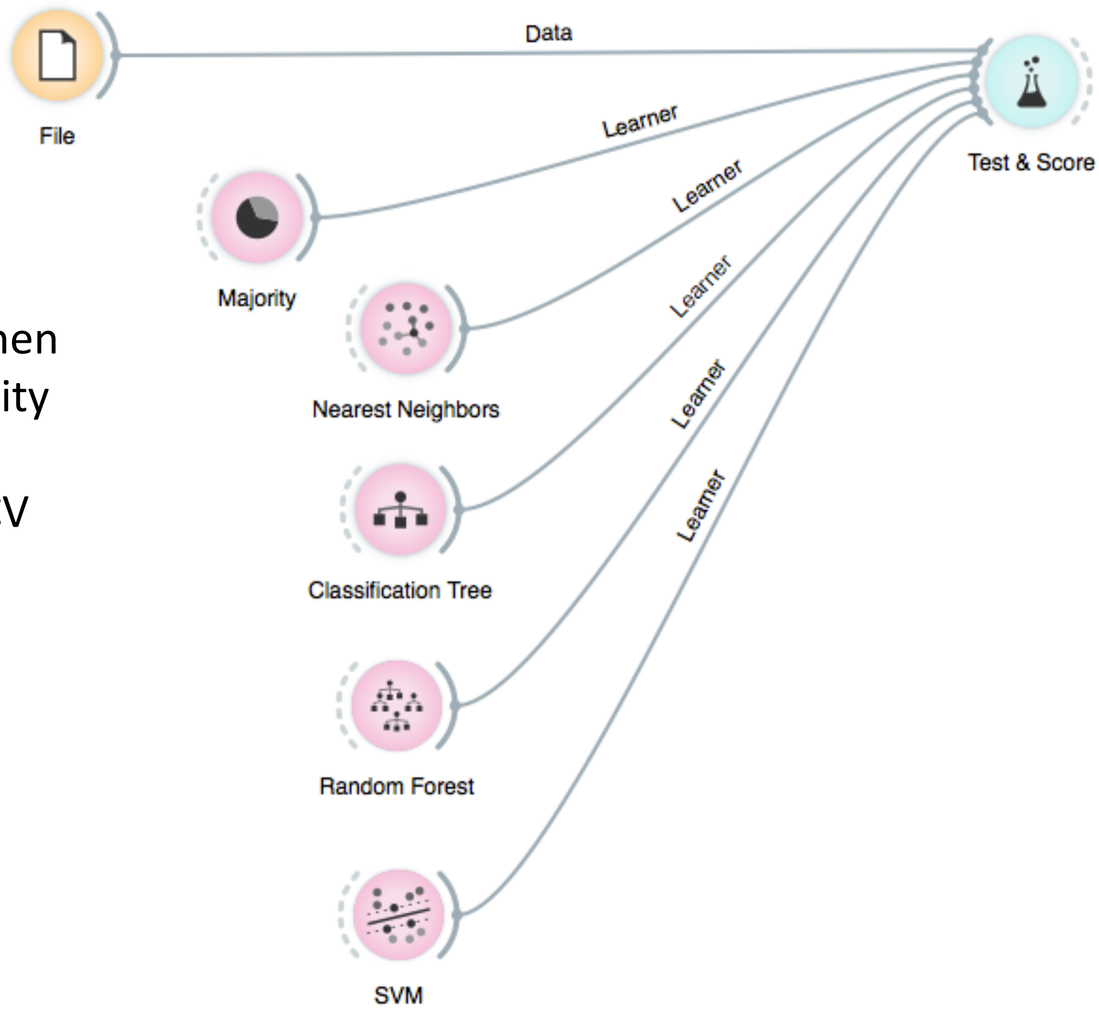


TILBURG UNIVERSITY

Experimenting with SVMs (spirals dataset)



Comparative Evaluation on any dataset



Please note:
Use 5 or 10 fold CV when comparing with Majority classifier!
NOT leaving one out CV