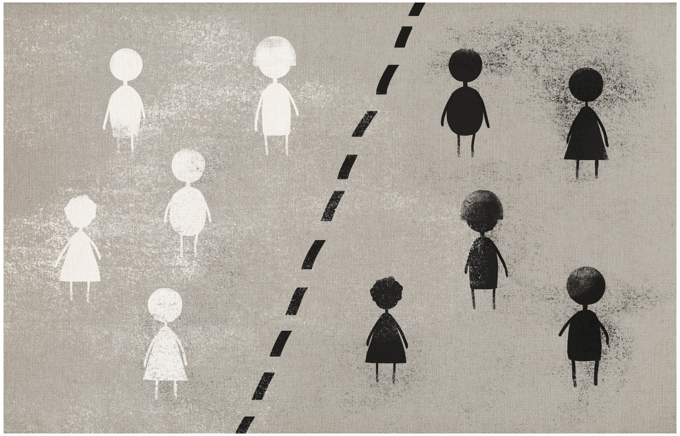
Support Vector Machines



Lluís Talavera, 2022

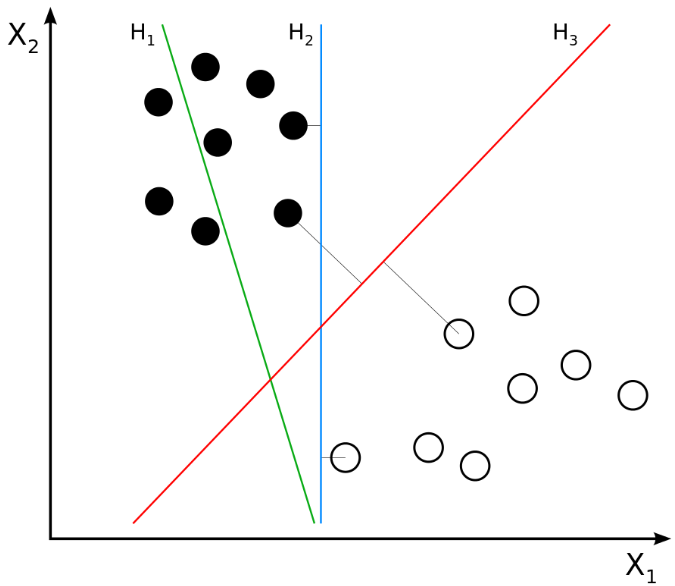
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

ETSEIB

Support Vector Machines (SVM)

Goal: find a hyperplane that separates two classes (linear separability is assumed).

There are an infinite number of hyperplanes but only some of them create a boundary between the classes. Which one is the best?



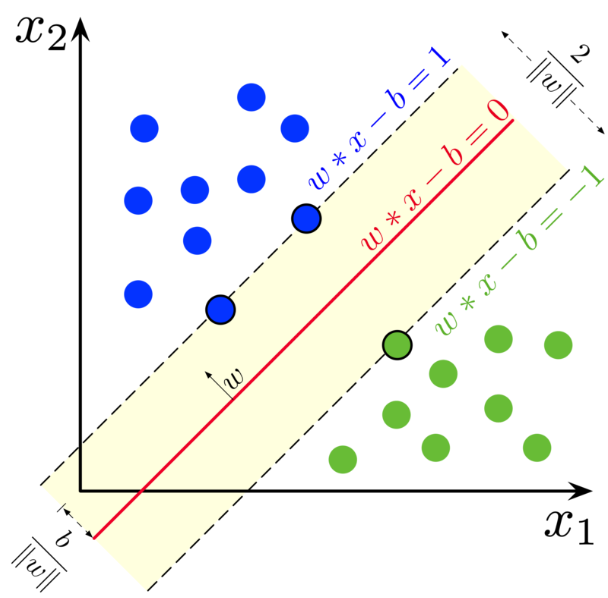
source

Hard margin classification

Support vectors: points from both classes closer to the hyperplane.

Margin: the distance between the support vectors and the hyperplane.

The optimal hyperplane is the one that classifies all the points correctly and maximizes the margin.



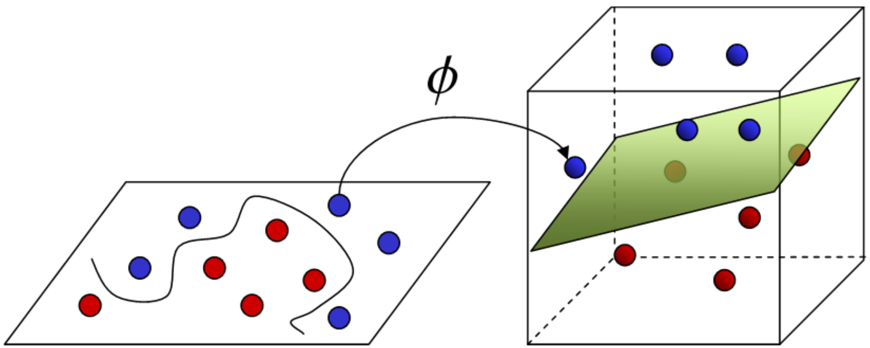
source

Soft margin classification

- The data may not be completely linearly separable or the margin can be too small and the model become prone to overfitting.
- We can add some flexibility by allowing a small number of points to be missclassified.
- A hyperparameter C controls the trade-off between maximizing the margin and minimizing the error. It is a penalty associated with making an error. The higher the value of C is the less likely the algorithm missclassifies a point.
- Very high values of C give a hard margin classifier.

The kernel trick

If data is not linearly separable, we can choose a mapping ϕ that projects the data to a higher dimensional space and try to find an optimal hyperplane in the new space.



The exact algebraic machinery is outside the scope of this course, but SVM only needs the dot product of each pair of points (vectors) to find the hyperplane. A kernel give us the dot product in a projected space without *actually projecting the points* (it is computationally cheaper!).

$K(x_i,x_j) = \left<\phi(x_i), \phi(x_j)\right>$

Some kernels

Linear

$K(\mathbf{x},\mathbf{y}) = \left< \mathbf{x},\mathbf{y} \right>$

Polynomial

$K(\mathbf{x},\mathbf{y}) = (\gamma \left< \mathbf{x},\mathbf{y}\right> + r)^{d}$

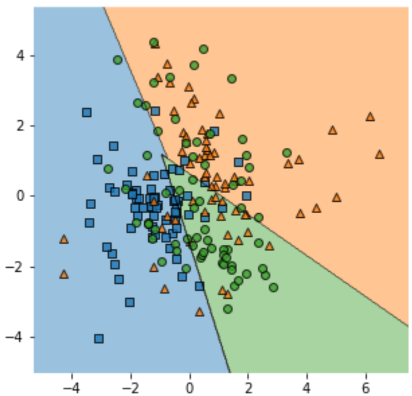
Radial Basis Function (RBF)

$K(\mathbf{x},\mathbf{y}) = \large e^{-\gamma ||\mathbf{x}-\mathbf{y}||^{2}}$

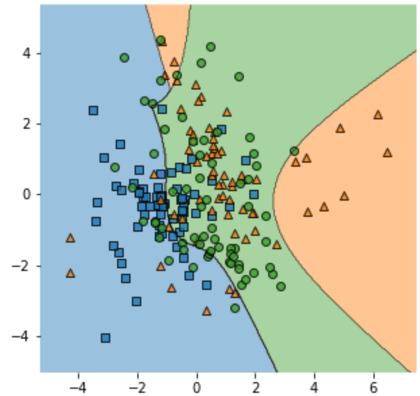
Sigmoid

$K(\mathbf{x},\mathbf{y}) = tanh(\gamma \left< \mathbf{x},\mathbf{y} \right> + r)$

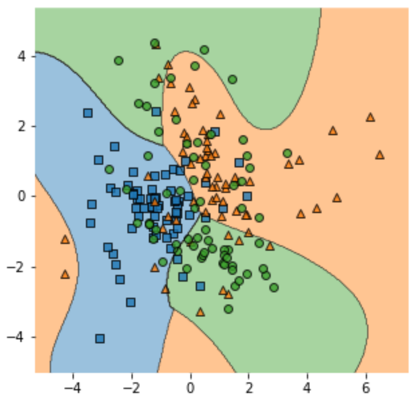
SVM decision boundaries



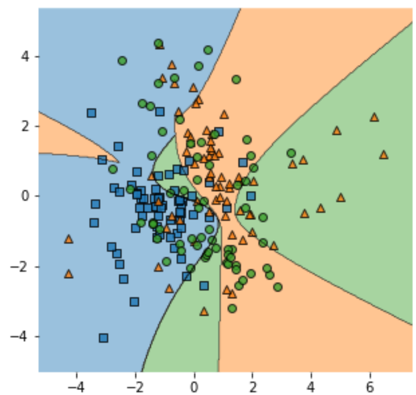
$\scriptsize \text{linear kernel}$



$\scriptsize \text{polynomial kernel degree=3}$



$\scriptsize\text{RBF kernel}$



$\scriptsize \text{sigmoid kernel}$

Some remarks about SVM

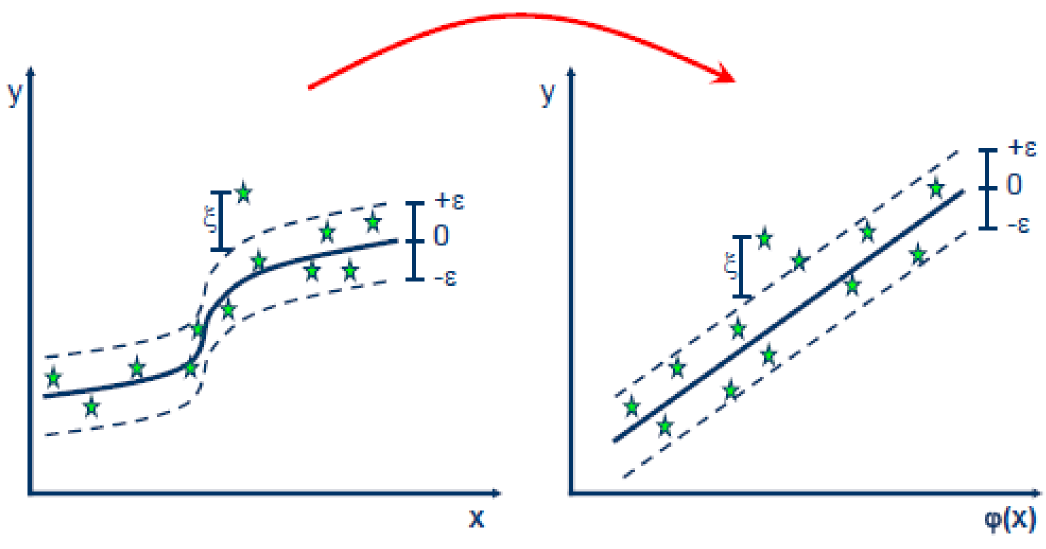
- Can solve complex problems using the appropriate kernel. Finding the rigth kernel may not be easy.
- Kernels can include expert knowledge about the distribution of the data.
- Effective in high dimensional spaces.
- Memory efficient: it does not use all the data, only the support vectors.
- Long training time with large amounts of data.

Support Vector Regression (SVR)

Uses the same principles as the SVM algorithm with few minor differences.

A margin of tolerance (ε) is established so that SVR finds a function with at most ε deviation from the target.

Ignores errors as long as they are less than ε. Only points outside the ε-tube contribute to the optimization.



sklearn imports

Classification

```
from sklearn.svm import SVC
```

Regression

```
from sklearn.svm import SVR
```