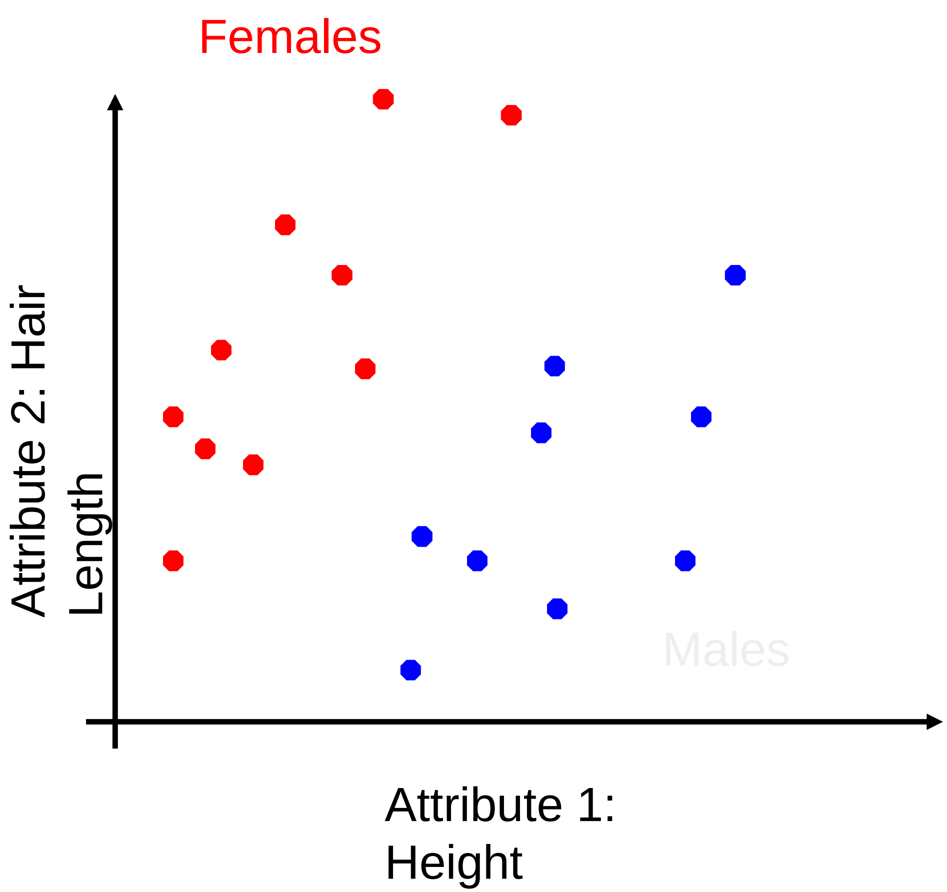
Support Vector Machines

Prof. Ankur Sinha

Indian Institute of Management Ahmedabad

Gujarat India

Support Vector Machines (SVM)



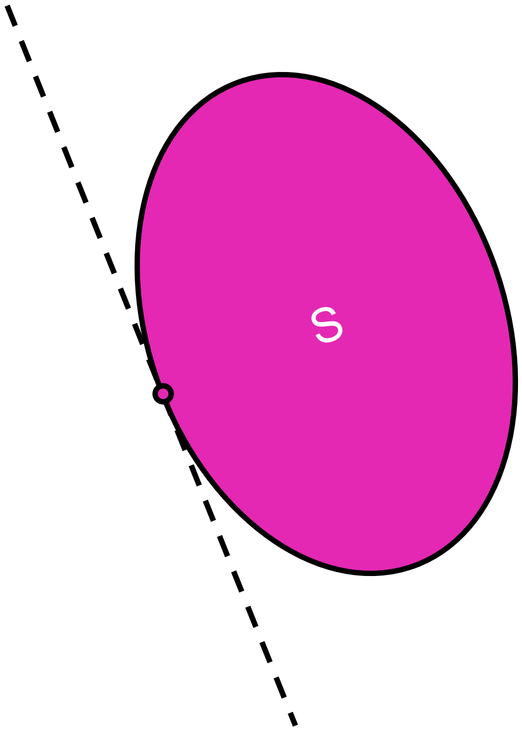
I have some objects with two attributes

It could be people with their heights and hair length

There are some points where you know the males and females

You want to train a robot based on these available points to classify people based on these attributes

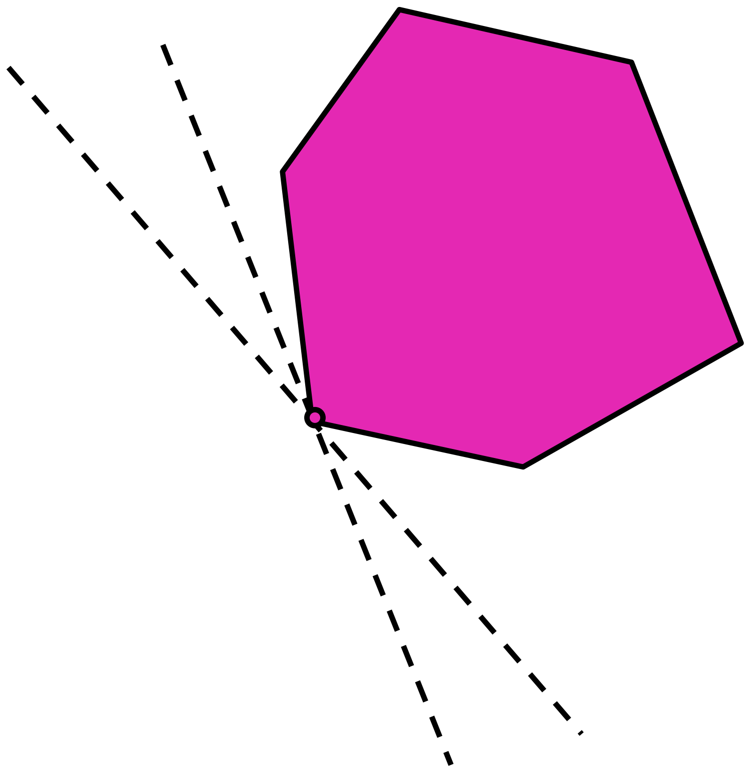
Supporting Hyperplane



A supporting hyperplane of a set S that has both of the following two properties:

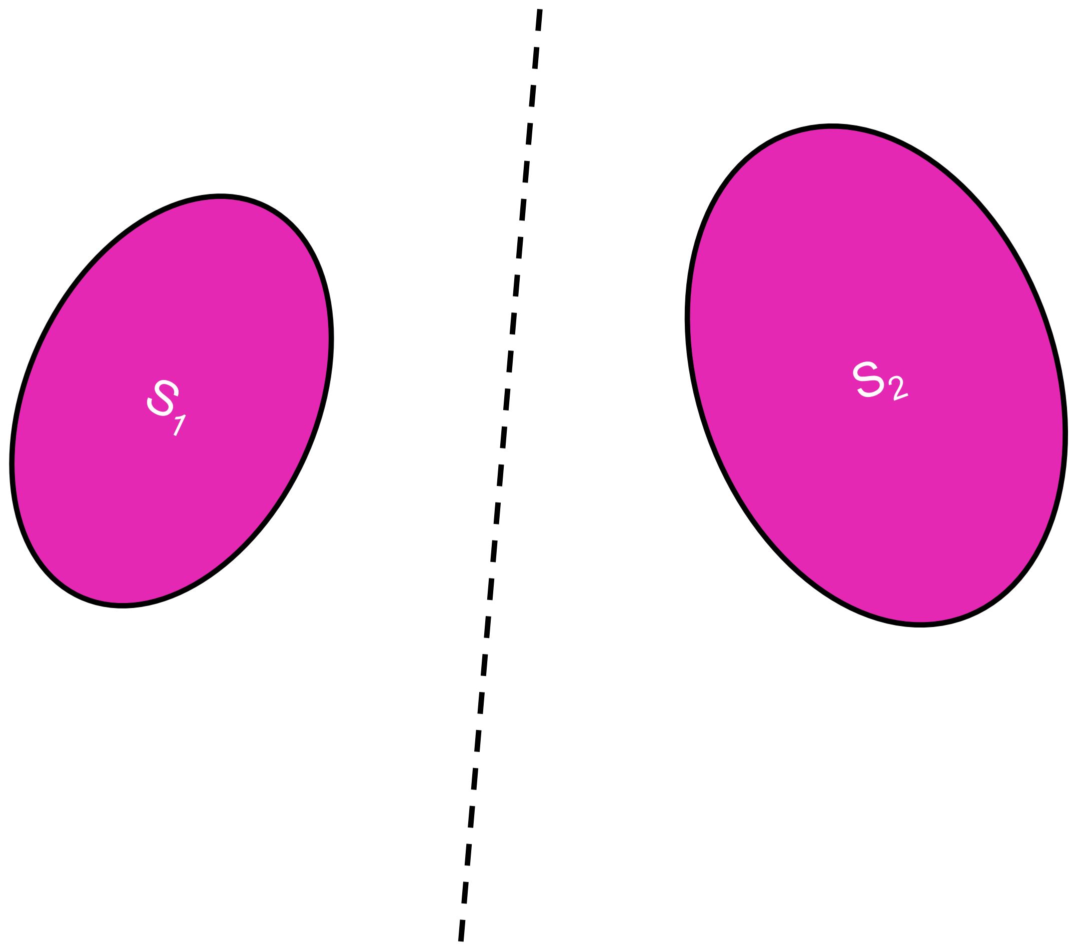
- S is entirely contained in one of the half-spaces
- S has at least one boundary-point on the hyperplane

Supporting Hyperplane

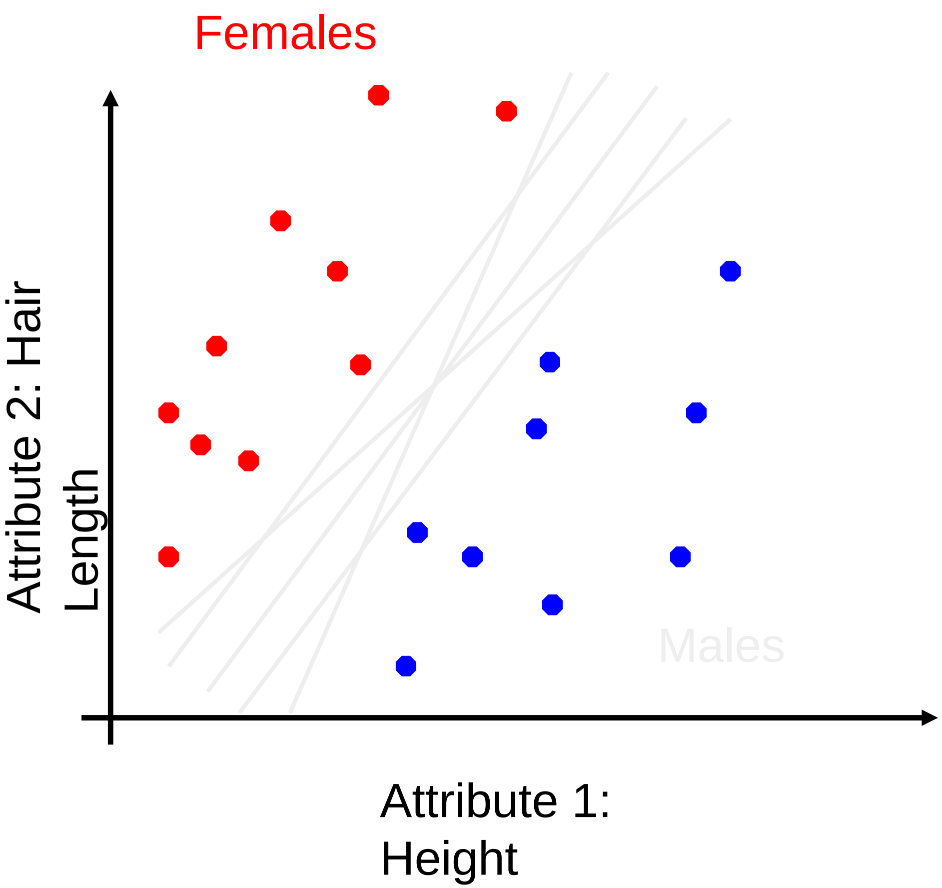


More than one hyperplane may be possible that meets the supporting hyperplane criteria

Separating Hyperplane



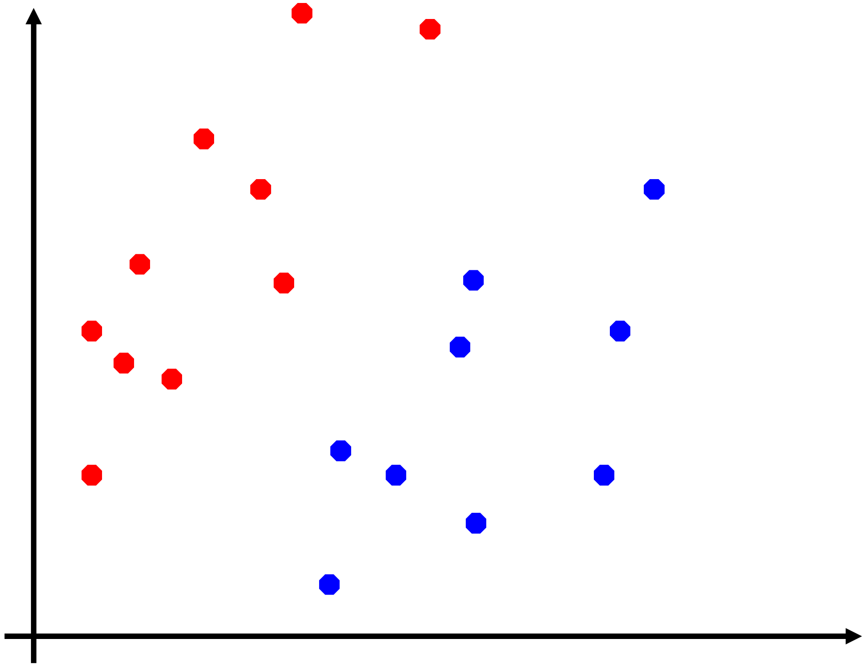
Support Vector Machines (SVM)



Multiple planes are possible

Which one to choose?

Support Vector Machines (SVM)

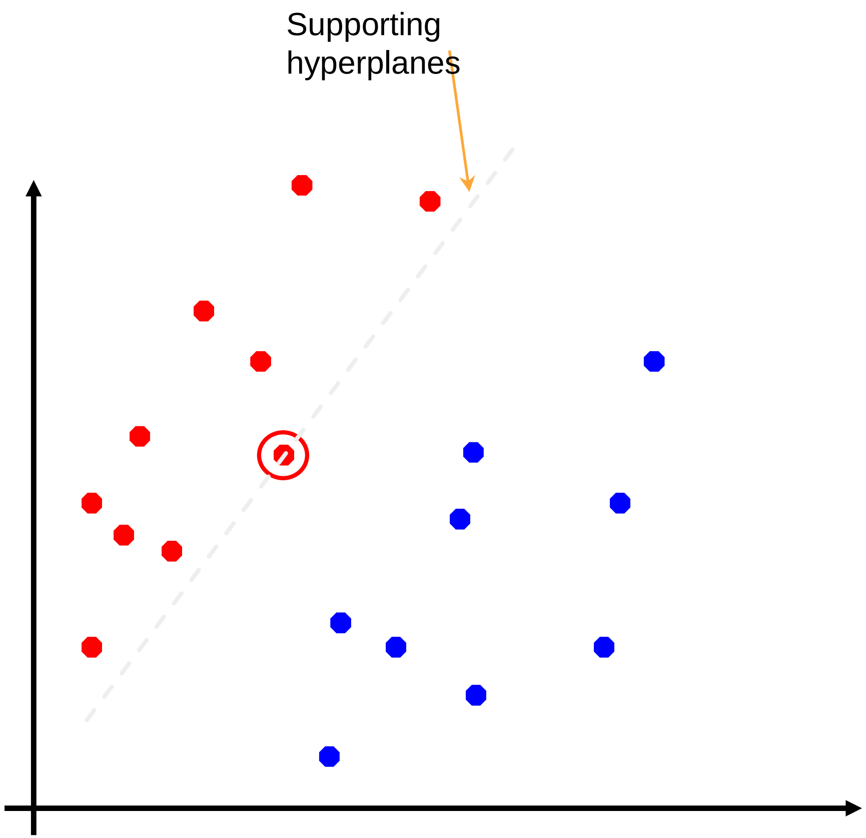


Choose the maximum margin hyperplane: Idea behind SVM

A maximum margin hyperplane separates the data points maximally

SVM tries to find the maximally separating plane

Support Vector Machines (SVM)

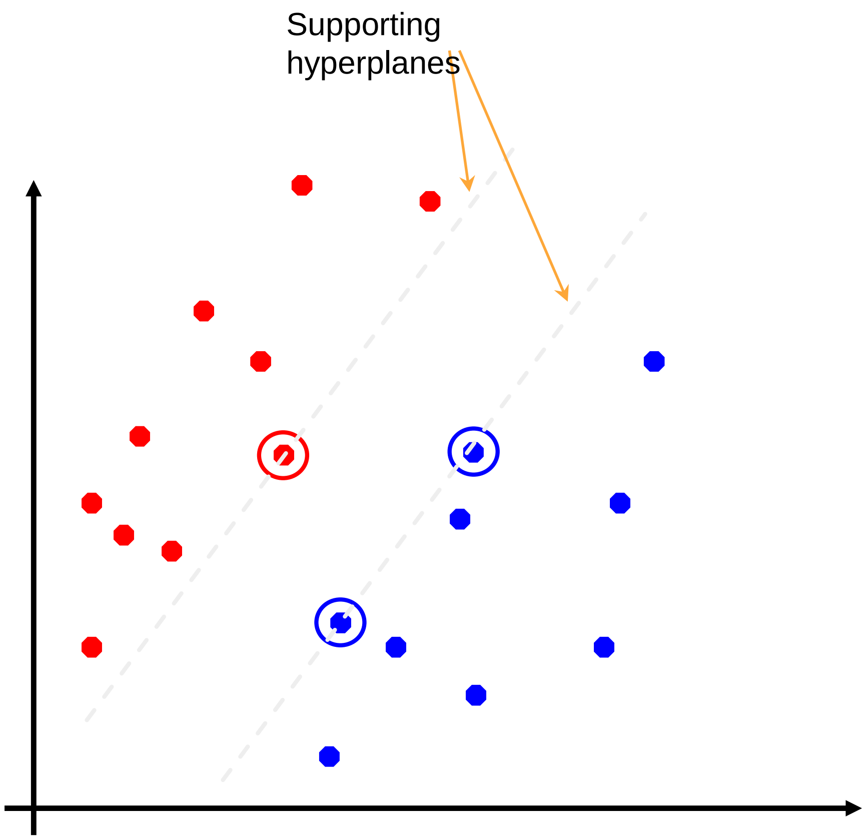


Choose the maximum margin hyperplane: Idea behind SVM

A maximum margin hyperplane separates the data points maximally

SVM tries to find the maximally separating plane

Support Vector Machines (SVM)

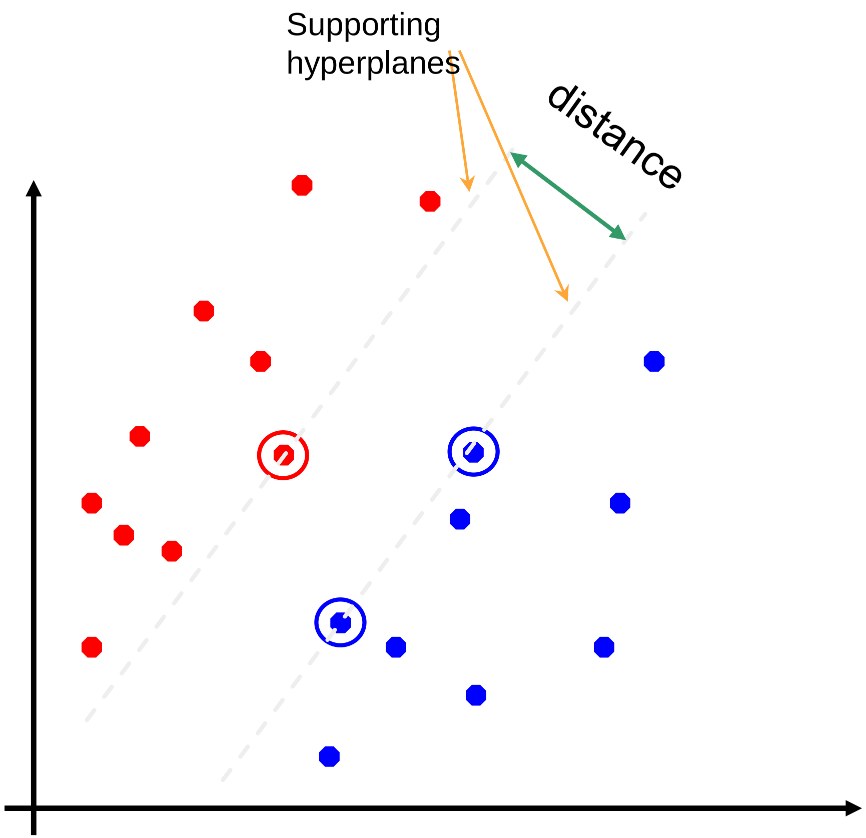


Choose the maximum margin hyperplane: Idea behind SVM

A maximum margin hyperplane separates the data points maximally

SVM tries to find the maximally separating plane

Support Vector Machines (SVM)

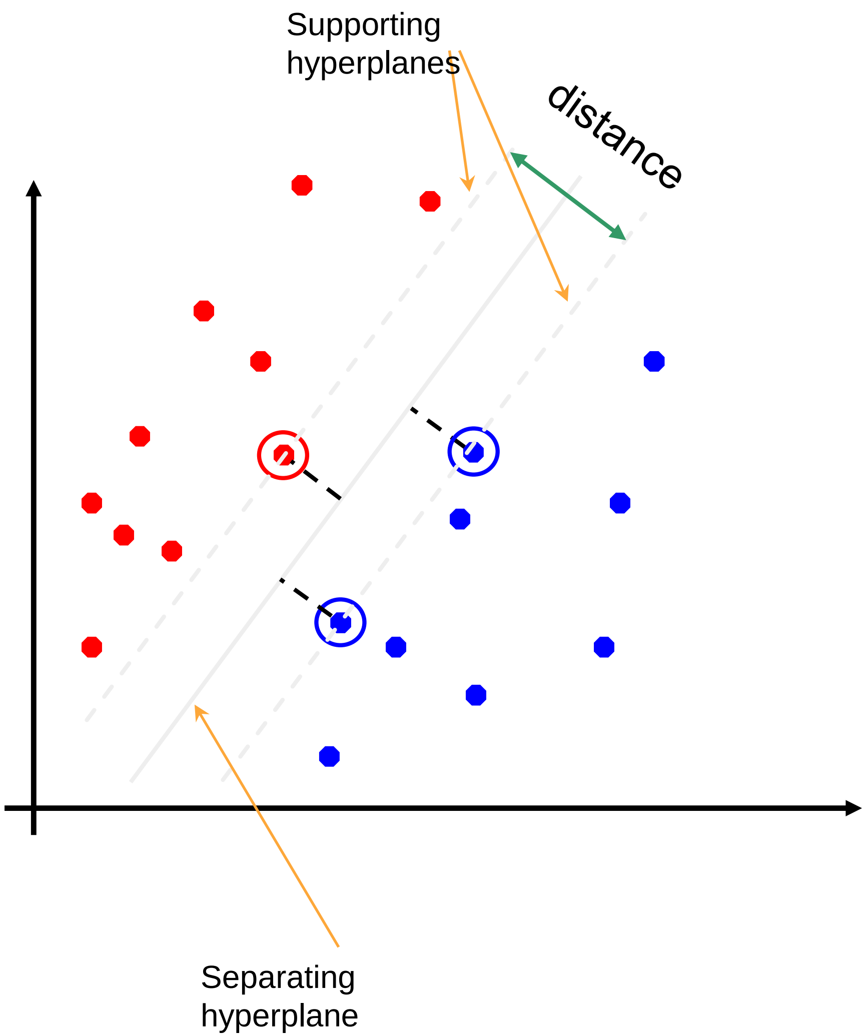


Choose the maximum margin hyperplane: Idea behind SVM

A maximum margin hyperplane separates the data points maximally

SVM tries to find the maximally separating plane

Support Vector Machines (SVM)

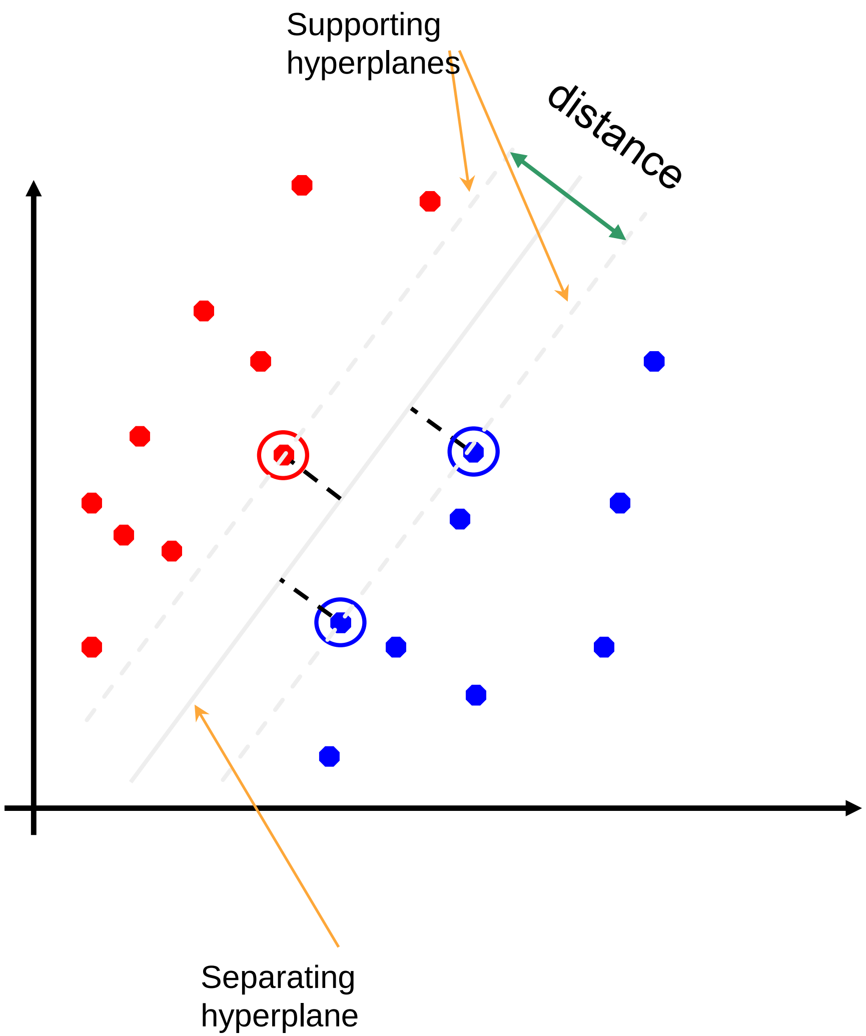


Choose the maximum margin hyperplane: Idea behind SVM

A maximum margin hyperplane separates the data points maximally

SVM tries to find the maximally separating plane

Support Vector Machines (SVM)

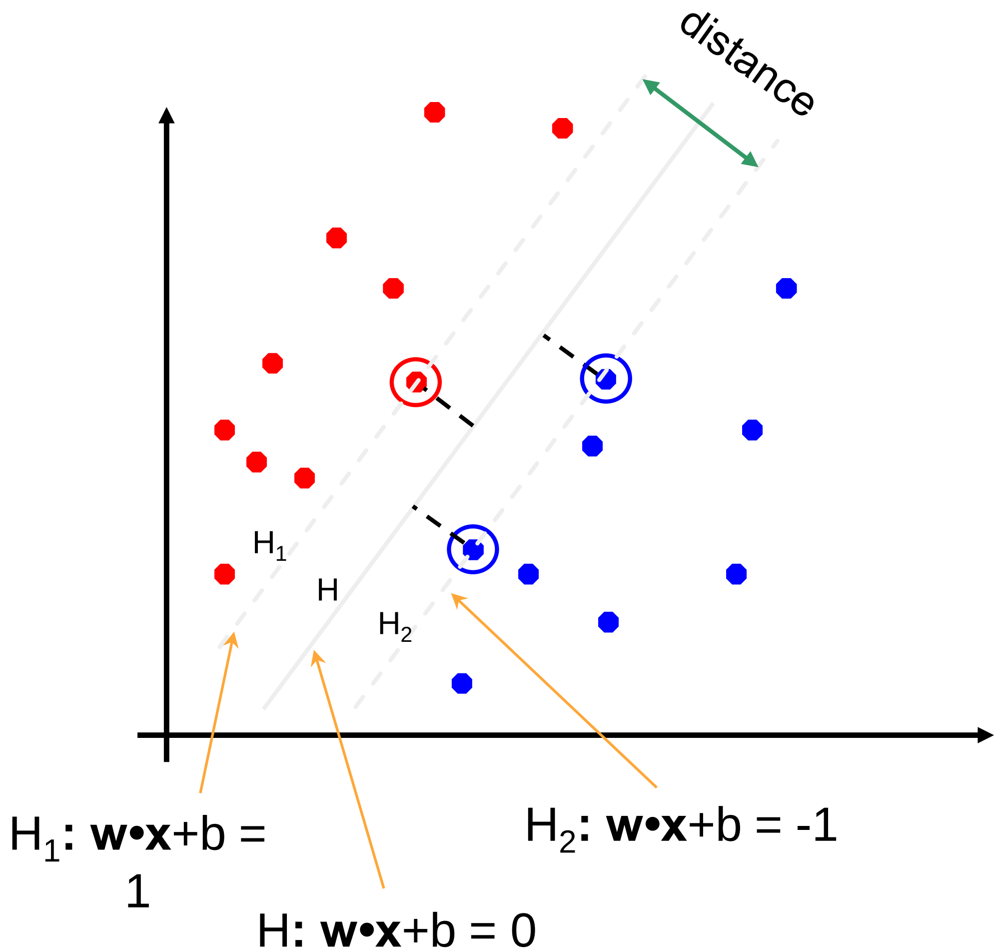


Choose the maximum margin hyperplane: Idea behind SVM

A maximum margin hyperplane separates the data points maximally

SVM tries to find the maximally separating plane

Support Vector Machines (SVM)



The distance from a point $\mathbf{x_0} = (x_0, y_0)$ to a line:
$Ax+By+c = 0$ is given as:
$|A x_0 + B y_0 + c|/\text{sqrt}(A^2+B^2)$

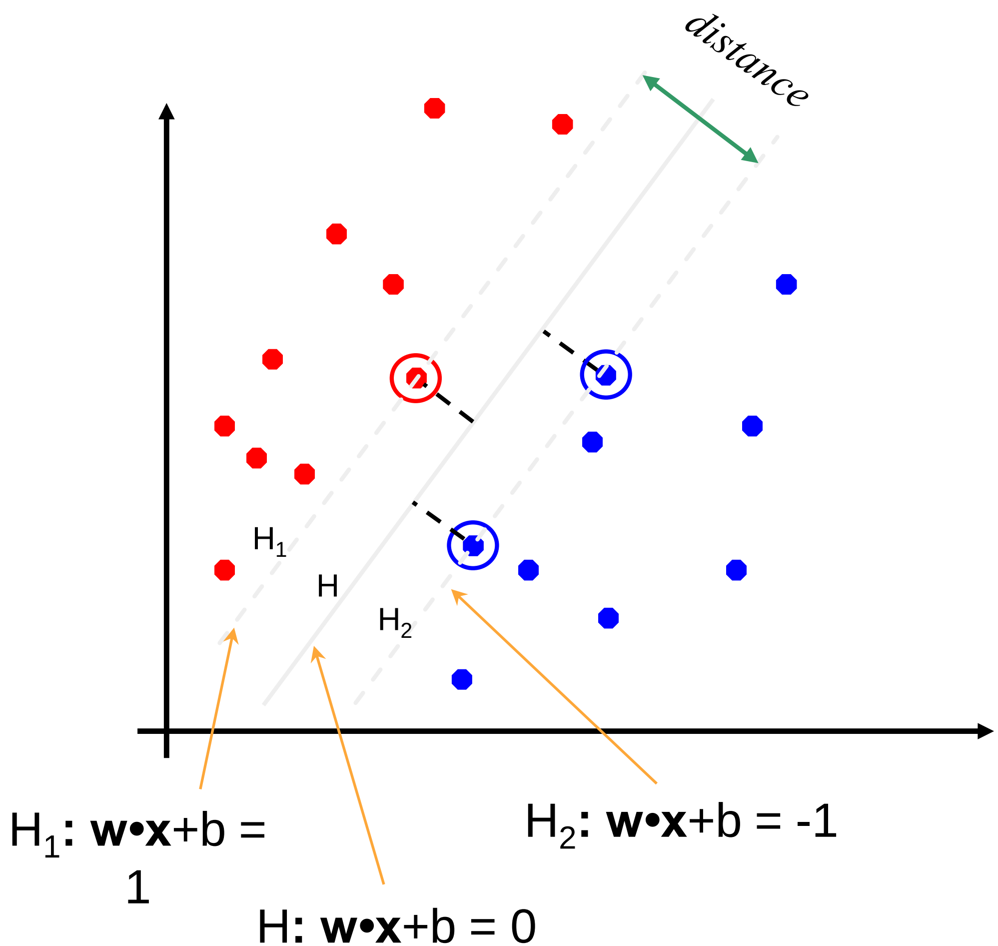
The distance between H and H_1 is:
$1/||\mathbf{w}||$

The distance between H_1 and H_2 is:
$2/||\mathbf{w}||$

Maximizing the margin is same as minimizing $||w||$ with the following condition:
$\mathbf{x}\bullet\mathbf{w}+b \geq +1$ when point lies in blue class
$\mathbf{x}\bullet\mathbf{w}+b \leq -1$ when point lies in red class

An optimization is performed to find appropriate values of w and b.

Support Vector Machines (SVM)



A more concise notation is as follows:

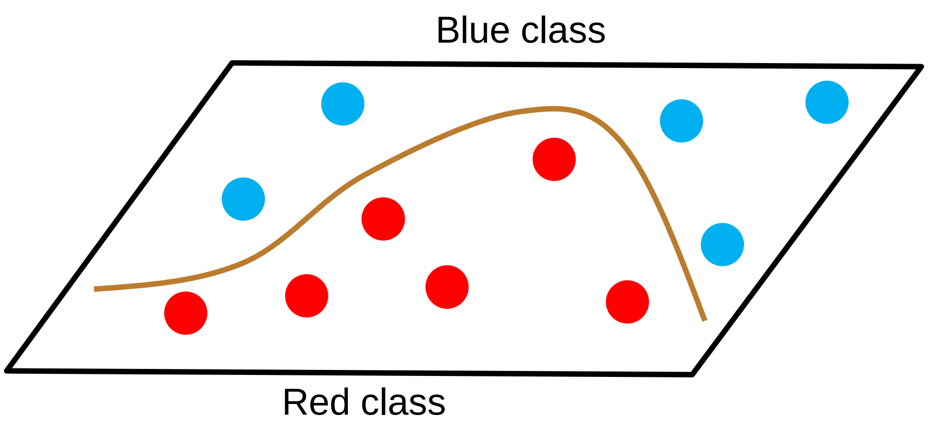
Find $\mathbf{w}$ and b such that

$\frac{1}{2}\, \mathbf{w}^T\mathbf{w}$ is minimized;

and for all $\{(\mathbf{x}_i, y_i)\}$: $y_i (\mathbf{w}^T\mathbf{x}_i + b) \geq 1$

Where x_i represents an object, and y_i represents its class

How about the following case?



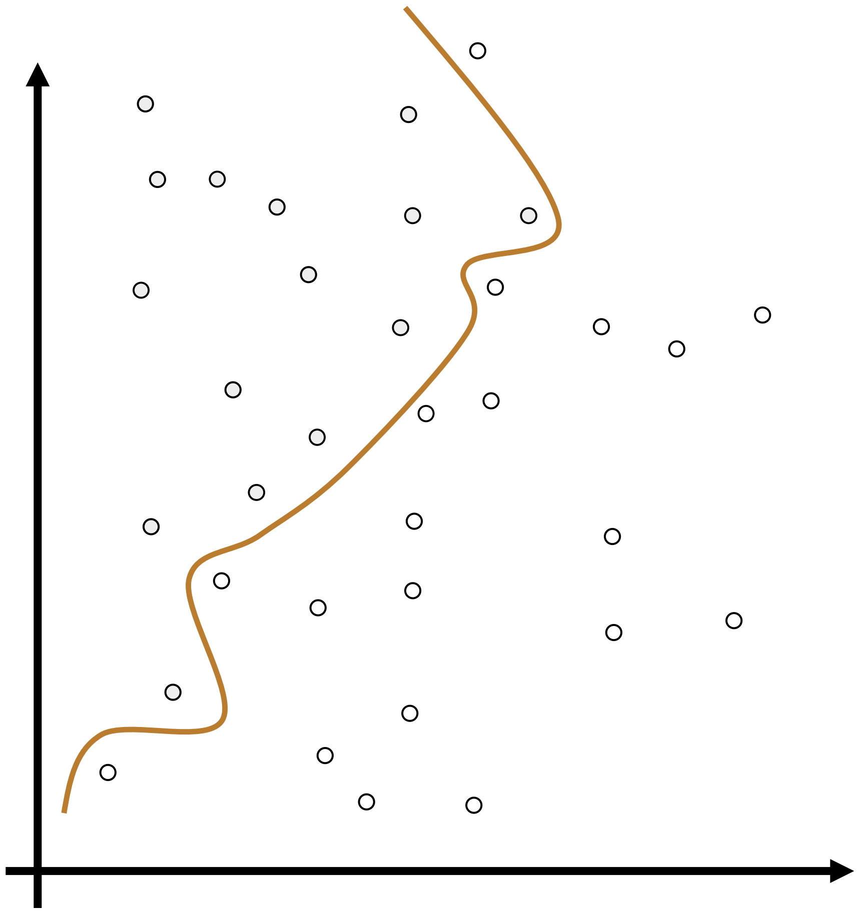
A linear model cannot be fitted

Solution: Use kernel functions to learn a non-linear model

Kernels

- Kernels can be of different types, for example:
  - Linear kernel: $K(w,x) = (w \cdot x)$
  - Polynomial kernel: $K(w,x) = (w \cdot x + 1)^p$, where p is a parameter
  - Gaussian kernel: $K(w,x) = \exp(||w-x||^2/2\sigma^2)$ where σ is a parameter

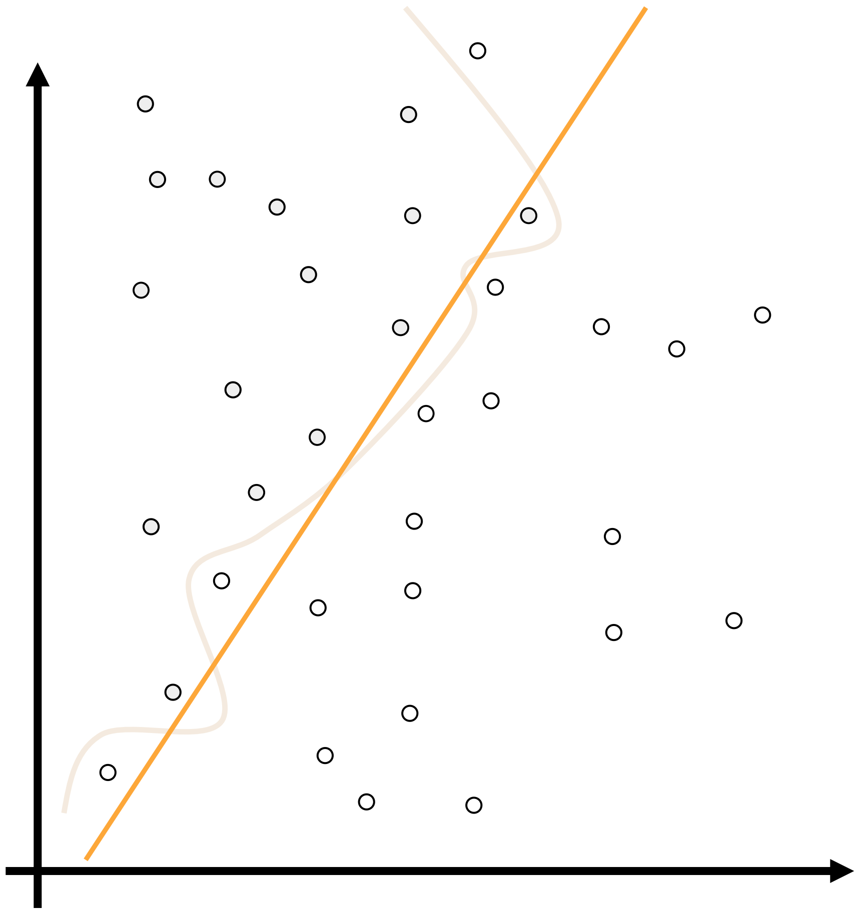
Simple vs Complex Kernels



- Non-linear Kernel: All data points are classified correctly

  - No training error
- Is it a good idea in case of this example?

Simple vs Complex Kernels



- Non-linear Kernel: All data points are classified correctly

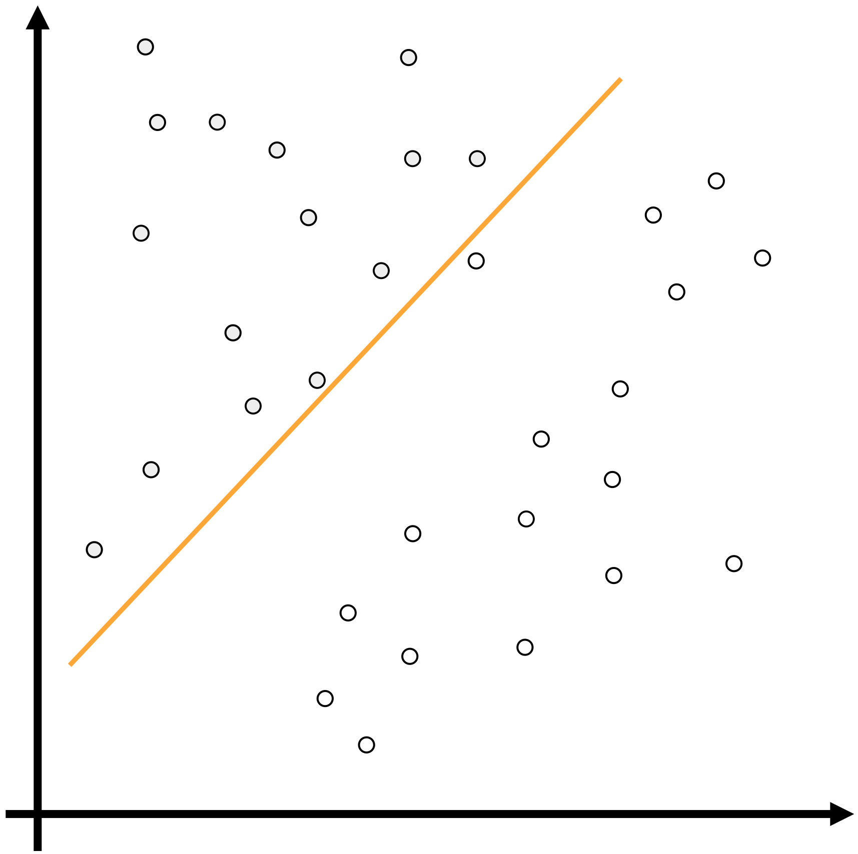
  - No training error
- Is it a good idea in case of this example?
- The straight line serves the purpose of dividing the two classes
- Non-linear Kernels may lead to:

Overfitting!

Soft Margin vs Hard Margin

- Hard margin classification means that you want to place the hyperplane such that there is complete classification without any errors
- Soft margin classification means that you allow some errors by not caring if some of the points are misclassified

Soft Margin vs Hard Margin

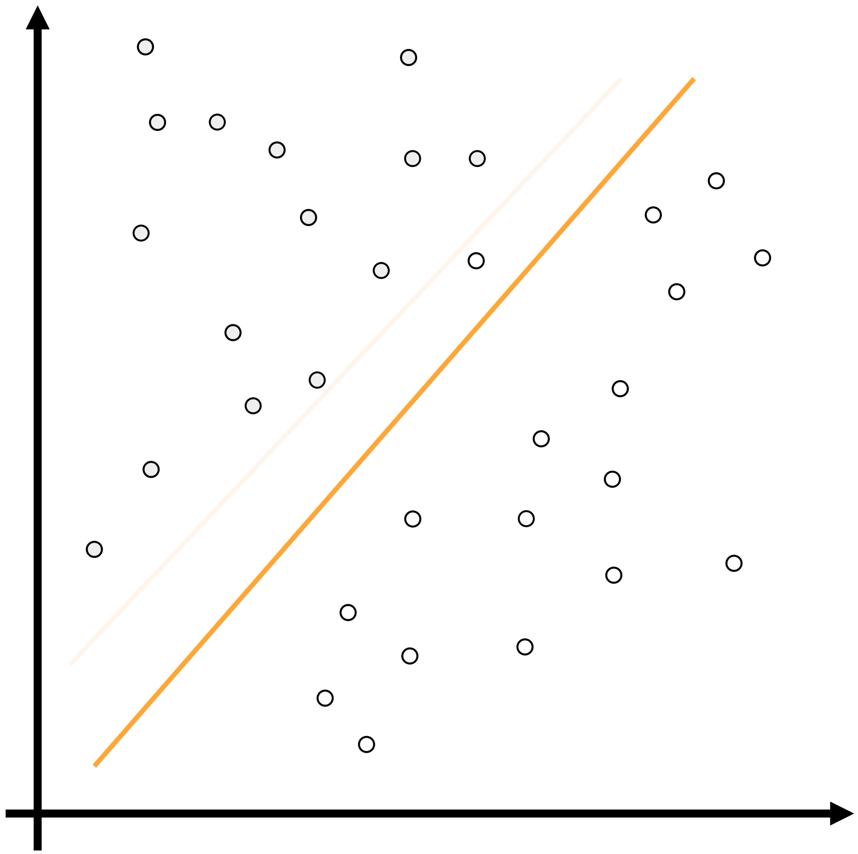


- Hard Margin: All data points are classified correctly

  - No training error

- Is it a good idea in case of this example?

Soft Margin vs Hard Margin



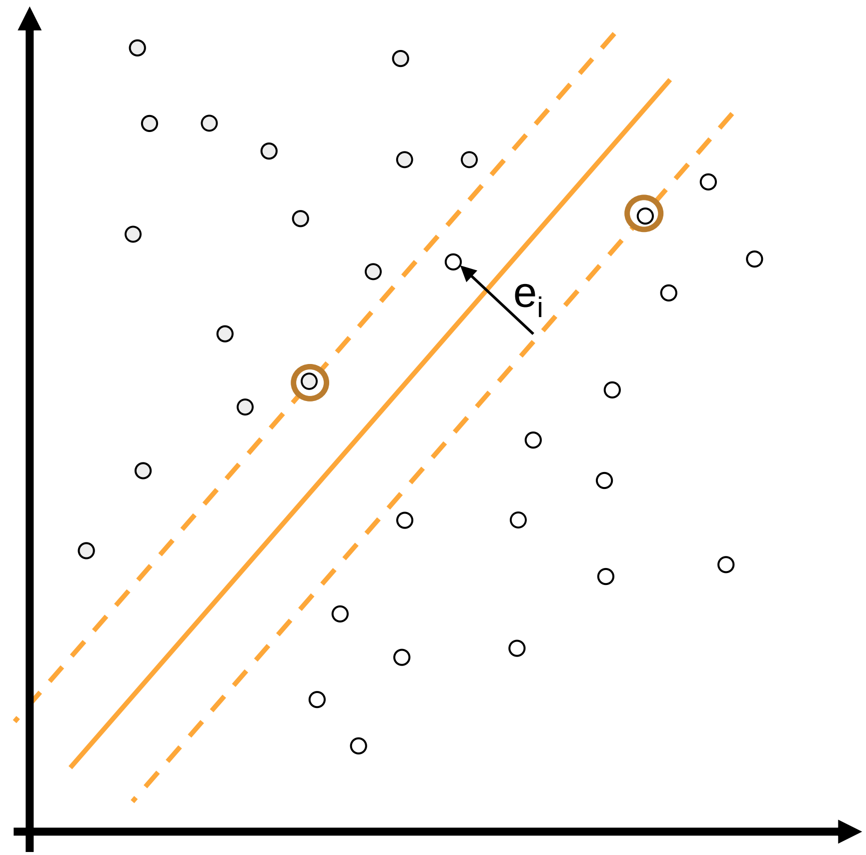
- Hard Margin: All data points are classified correctly

  - No training error
- Is it a good idea in case of this example?
- A soft margin plane may generalize better than the hard margin plane
- A hard margin plane may lead to:

Overfitting!

Soft Margin

- Locate the hyperplane such that it not only maximally separates the two classes but also tries to minimize the errors e_i



Original Formulation:
Find $\mathbf{w}$ and b such that
$\frac{1}{2}\,\mathbf{w}^T\mathbf{w}$ is minimized and for all $\{(\mathbf{x_i}, y_i)\}$
$y_i\,(\mathbf{w}^T\mathbf{x_i} + b) \geq 1$

Soft Margin Formulation
Find $\mathbf{w}$ and b such that
$\frac{1}{2}\,\mathbf{w}^T\mathbf{w} + C\,\Sigma e_i$ is minimized and for all $\{(\mathbf{x_i}, y_i)\}$
$y_i\,(\mathbf{w}^T\mathbf{x_i} + b) \geq 1 - e_i$ and $e_i \geq 0$ for all i

Where C is a parameter

Increasing C leads to a harder margin