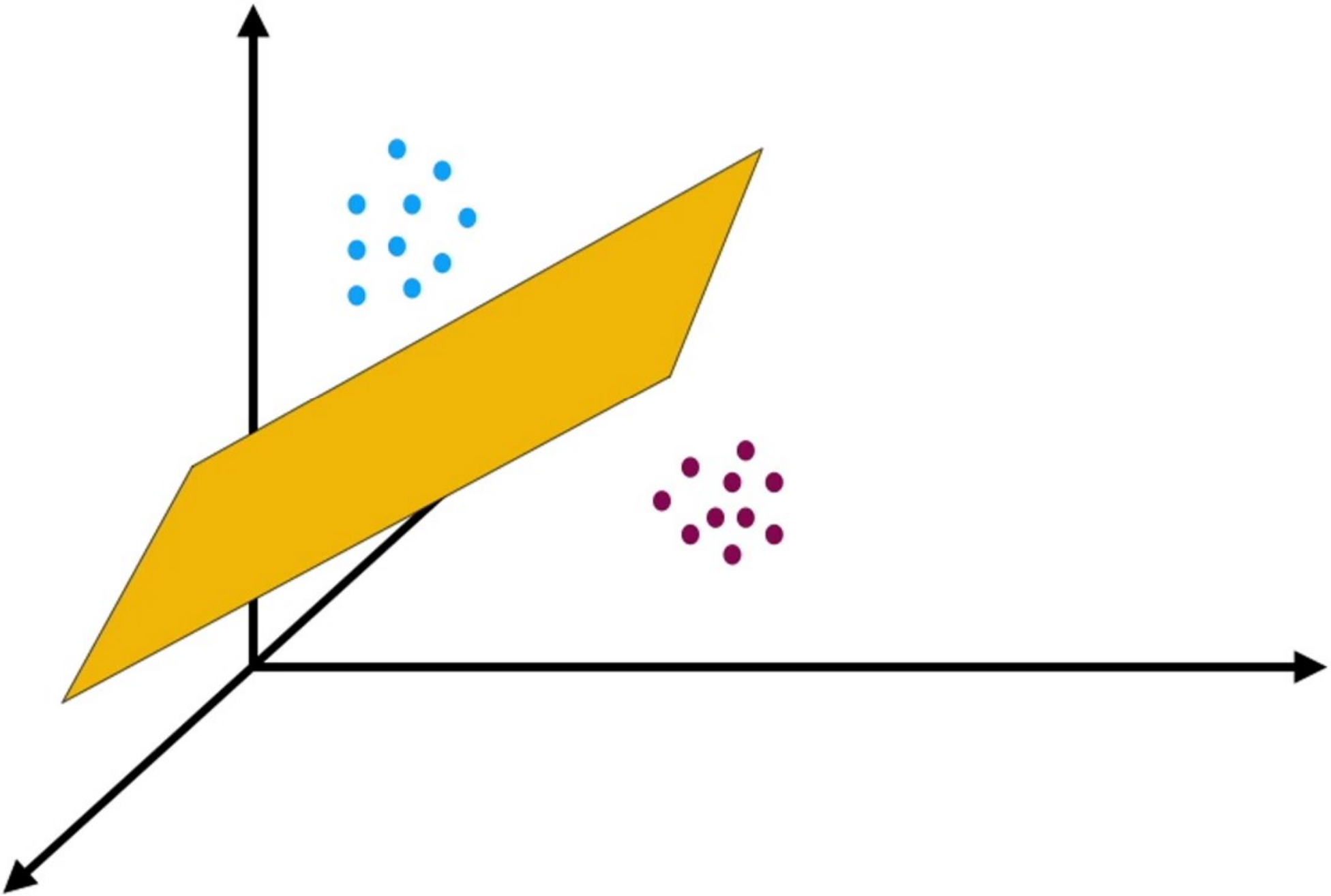
Support Vector Machines

What is an SVM

- SVMs are used to build binary classifiers
- Make classification decision on basis of a "linear function" of a point's coordinates
- Does not require prior knowledge of probability distribution of the points
- Involves an explicit training stage

Hyperplane



How does an SVM Work?

The Support Vector Machine finds an (n-1)-dimensional hyperplane in an n-dimensional space to separate points into two categories

Hyperplane

In a vector space of n dimensions, a hyperplane is a geometric shape with (n-1) dimensions and zero thickness in one dimension

Hyperplane

In our example: $Ax + By + Cz = D$

Points on one side of hyperplane:

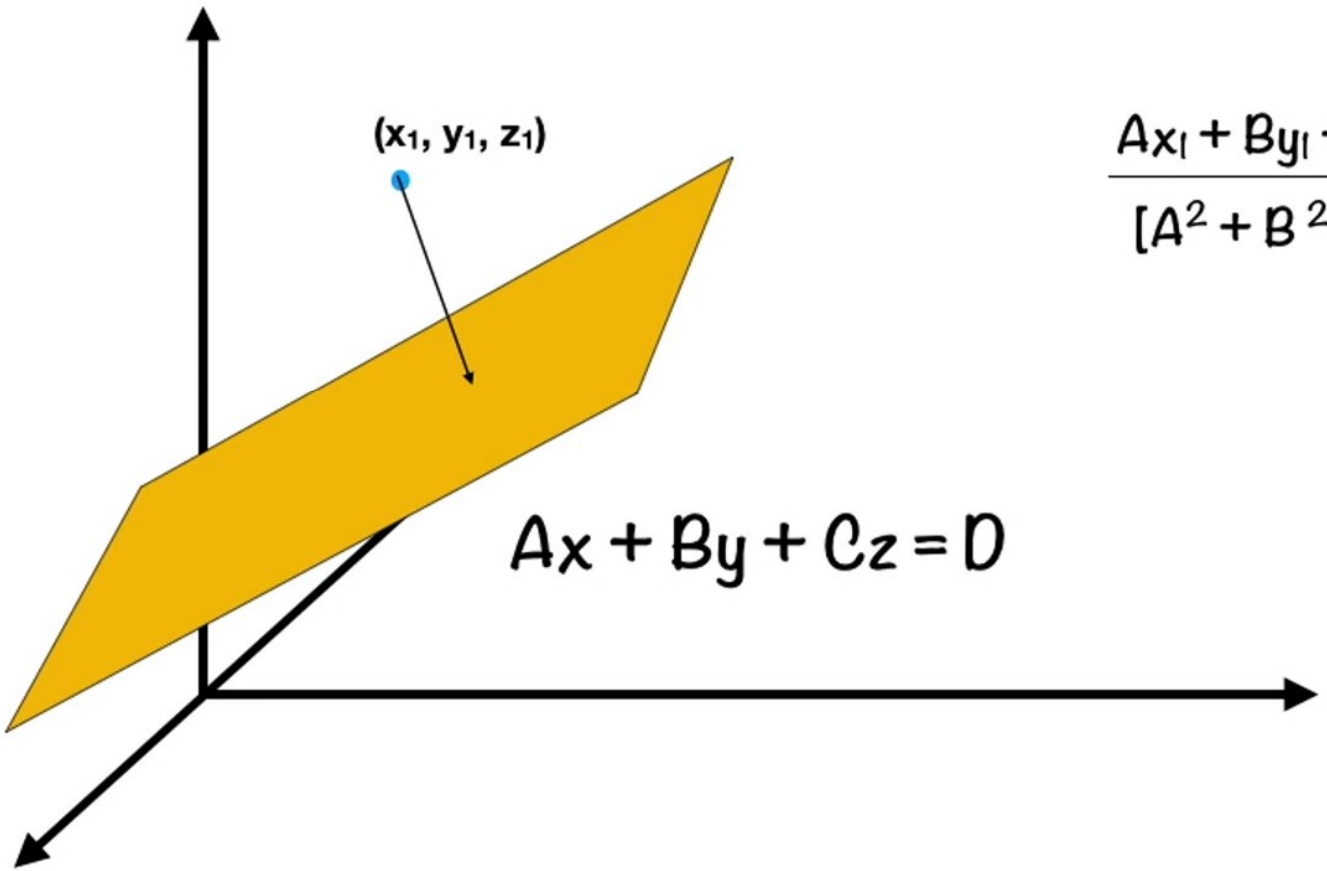
$Ax + By + Cz > D$

Points on other side of hyperplane:

$Ax + By + Cz < D$

SVM tries to find such a linear equation

Hyperplane

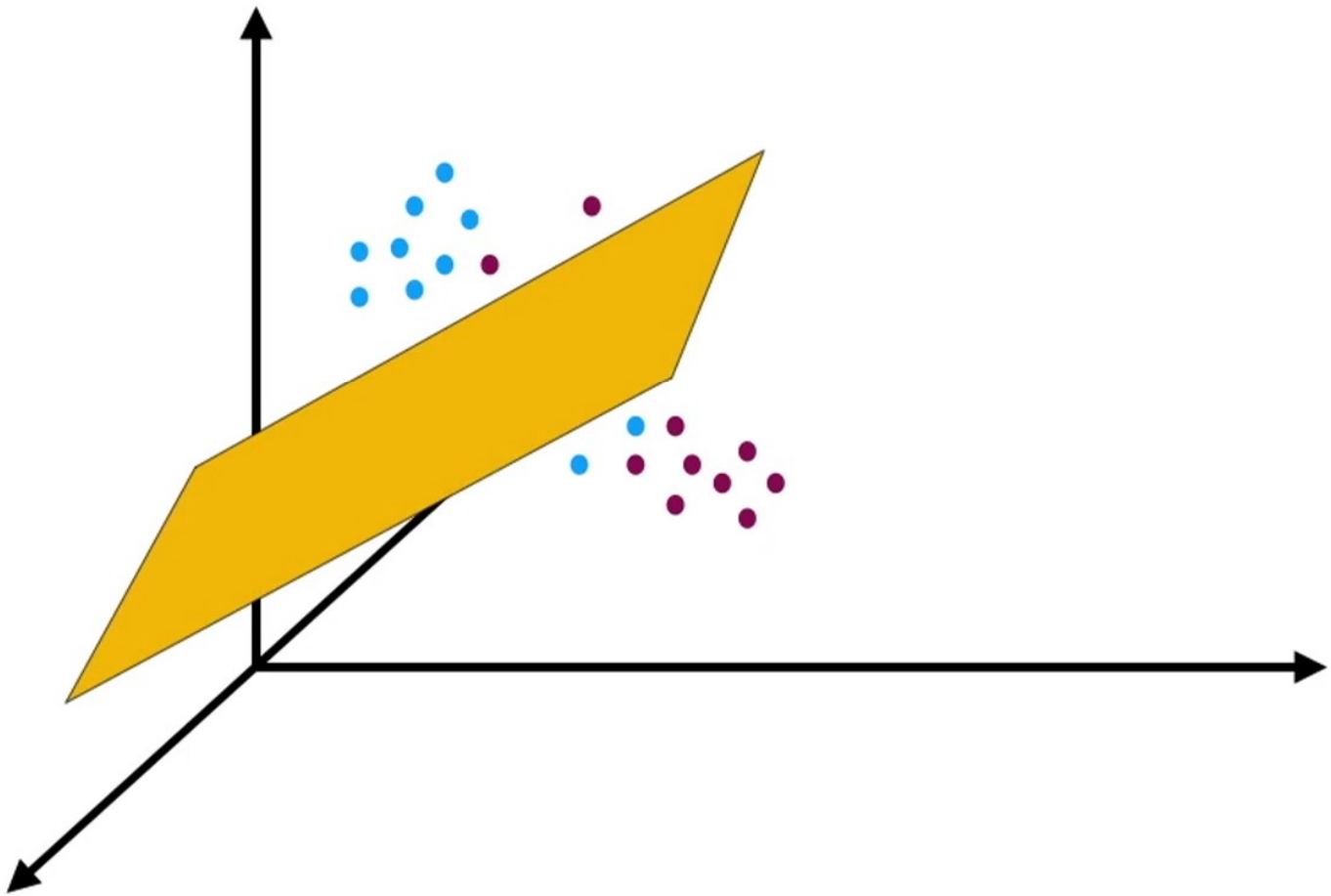


$$\frac{Ax_1 + By_1 + Cz_1 - D}{[A^2 + B^2 + C^2]^{1/2}}$$

Hyperplane

The "best" hyperplane is one which maximises the sum of the distances of the nearest points on either side of the hyperplane

What if points are not linearly separable?



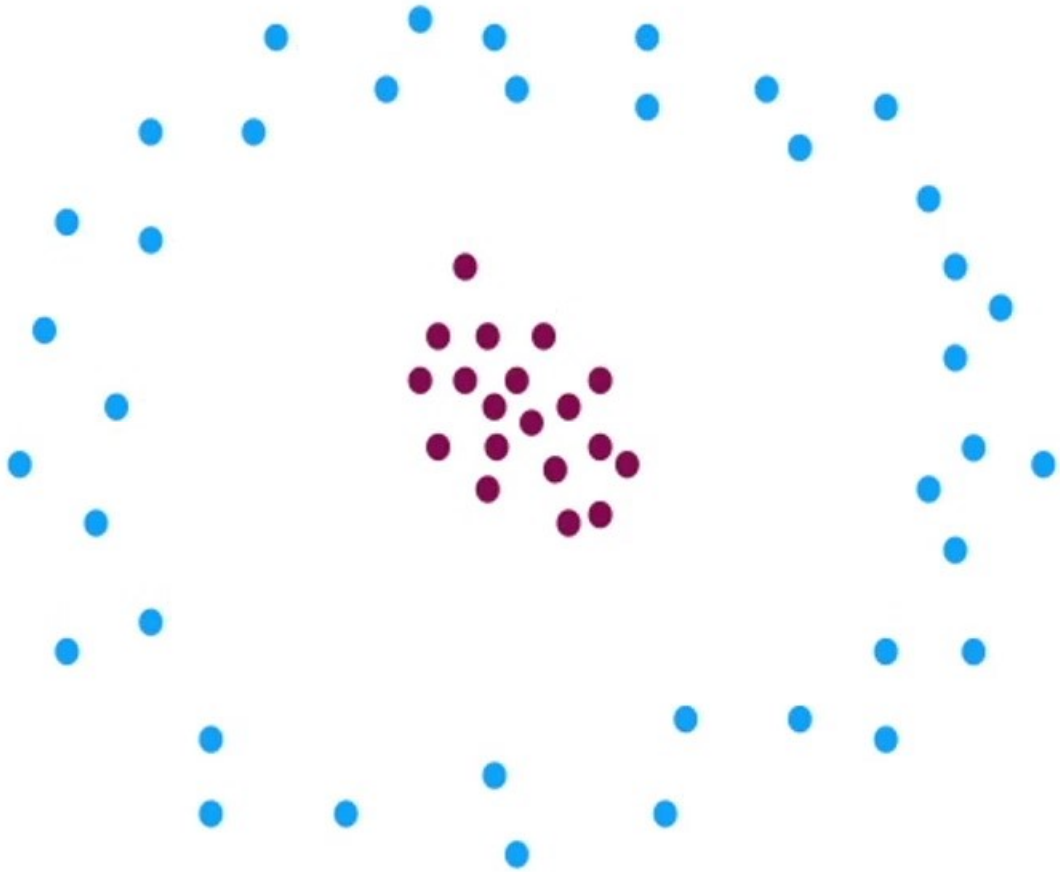
Hyperplane

The soft margin method finds a hyperplane which performs as clean a separation of points as possible

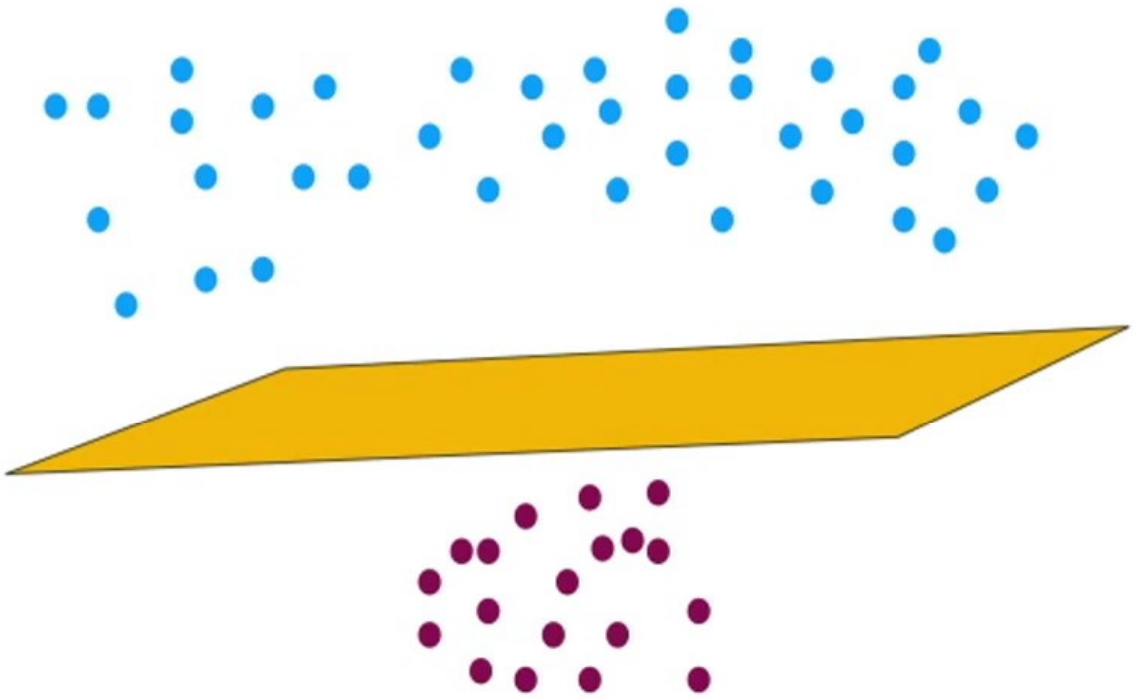
Non-linear separation

- SVM is a linear classifier
- But can be used to perform non-linear classification
- Achieved by using kernel trick

Kernel Trick



Kernel Trick



Kernel Trick

- Linear classification does a dot product of vectors with many elements
- Kernel is a non-linear function which is used instead of a dot product

Kernel Trick

- Kernel functions operate in feature space which may have many more dimensions than original feature space
- Finds the maximum margin hyperplane in the modified feature space

Kernel Trick

Allows a way to solve problems where the data is not linearly separable by projecting such data into a higher dimensional space