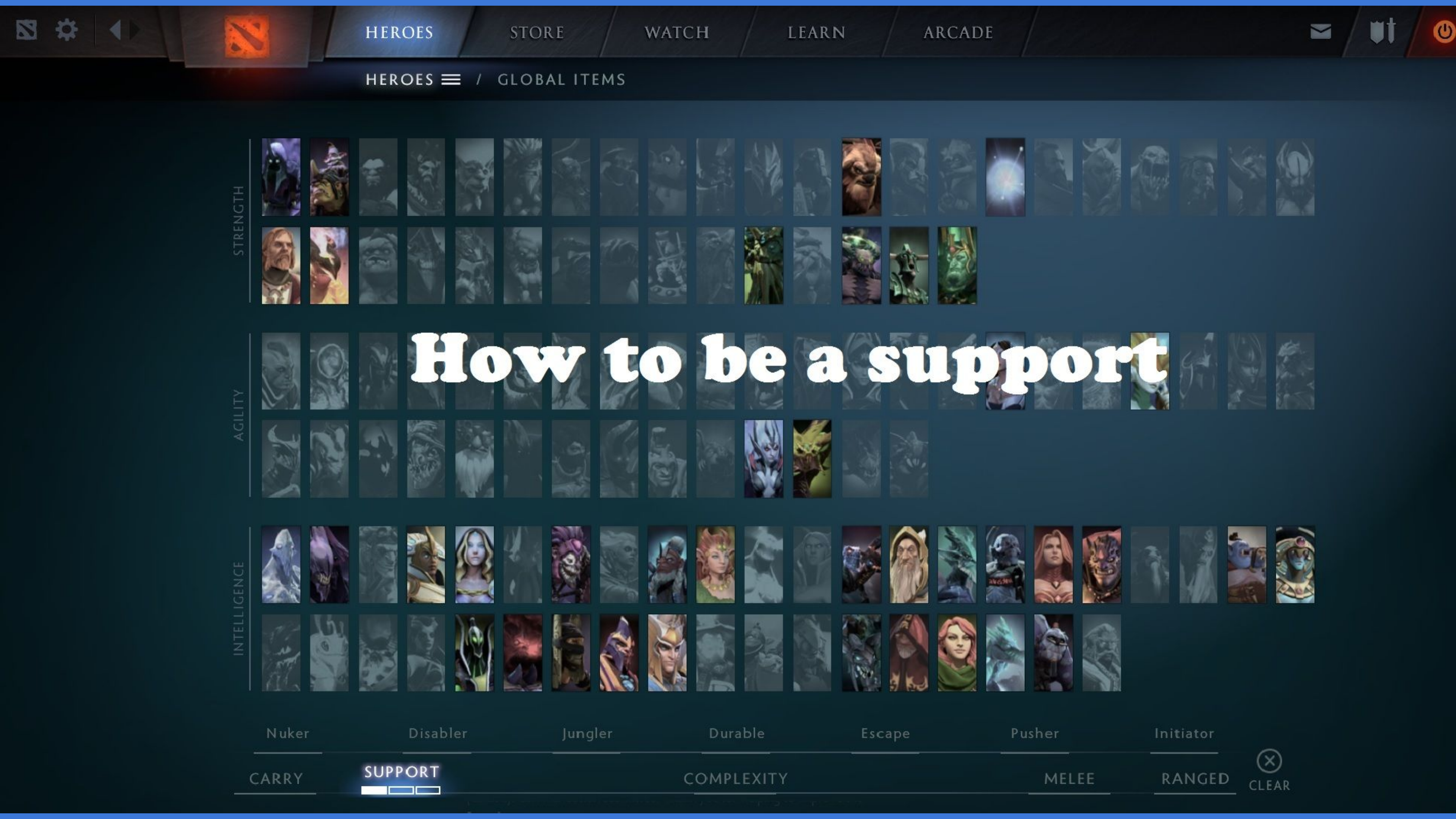
CS 129.18

Support Vector Machines


HEROES
STORE
WATCH
LEARN
ARCADE
HEROES
GLOBAL ITEMS
STRENGTH
AGILITY
INTELLIGENCE
How to be a support
Nuker
Disabler
Jungler
Durable
Escape
Pusher
Initiator
CARRY
SUPPORT
COMPLEXITY
MELEE
RANGED
CLEAR

Support Vector Machines

The support vector machine is a generalization of the maximal margin classifier.

Maximal Margin Classifier

Algorithm that seeks to determine a **linear** boundary between classes of data

Support Vector Machines

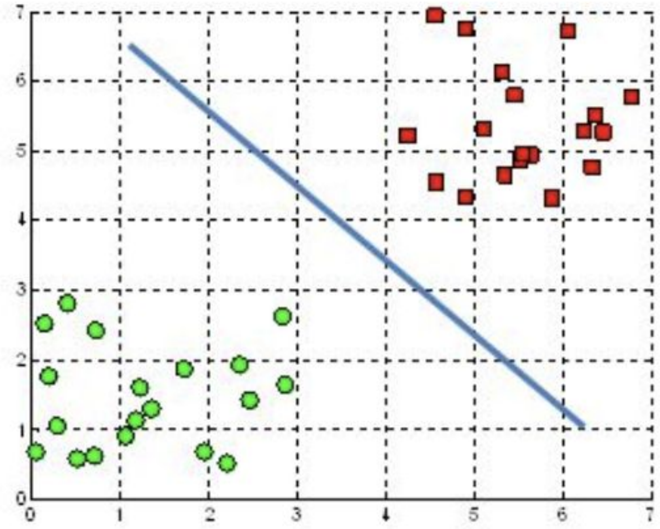
Not limited to linear separations of data, and can be used for binary or multiclass classification

Hyperplane

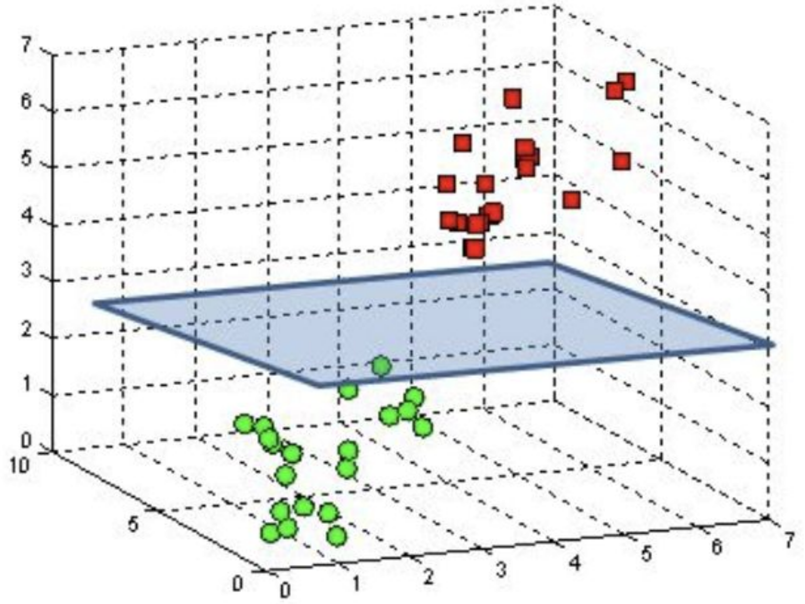
In an *n-dimensional* space, a hyperplane is a flat affine subspace of dimension n -1.

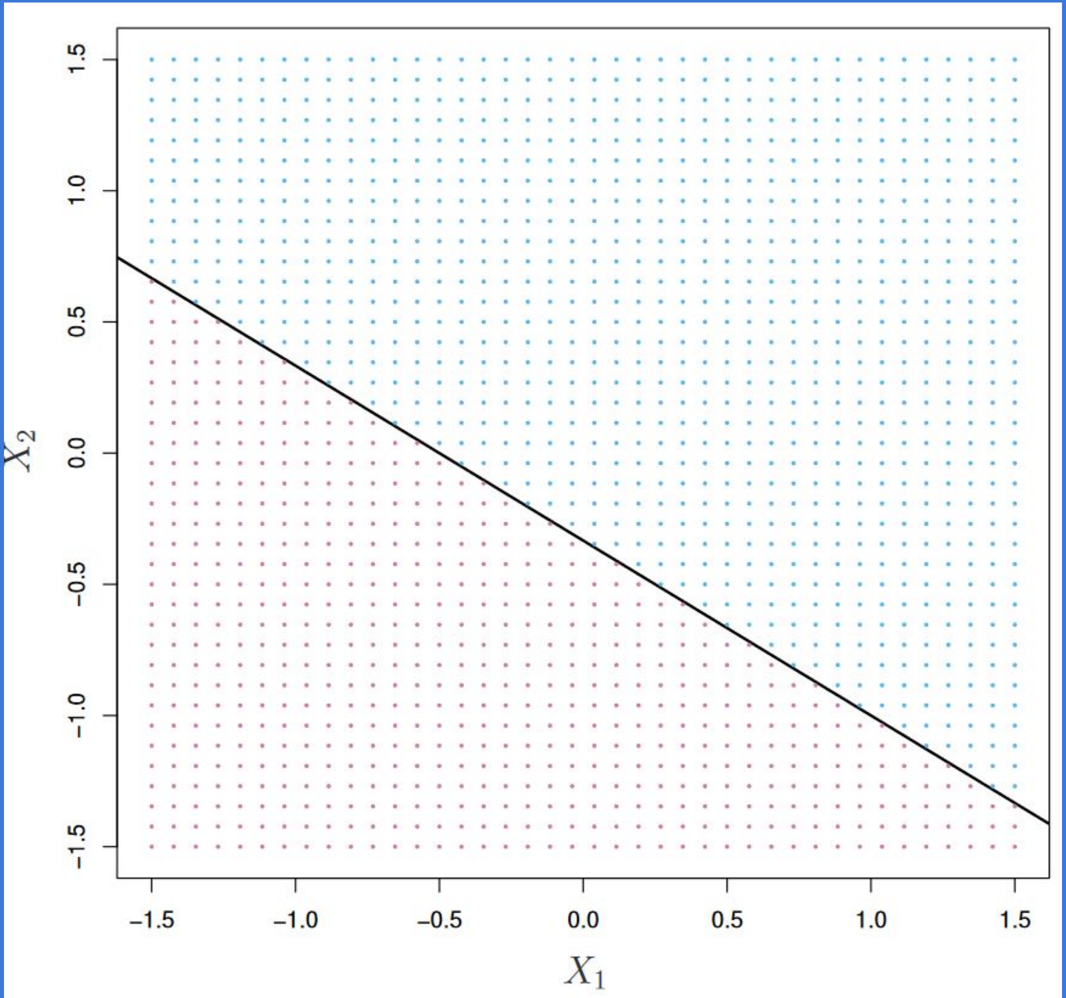
Geometrically, in a 2D space, the hyperplane will be a line, and in a 3D space, it will be a flat plane.

A hyperplane in $\mathbb{R}^2$ is a line



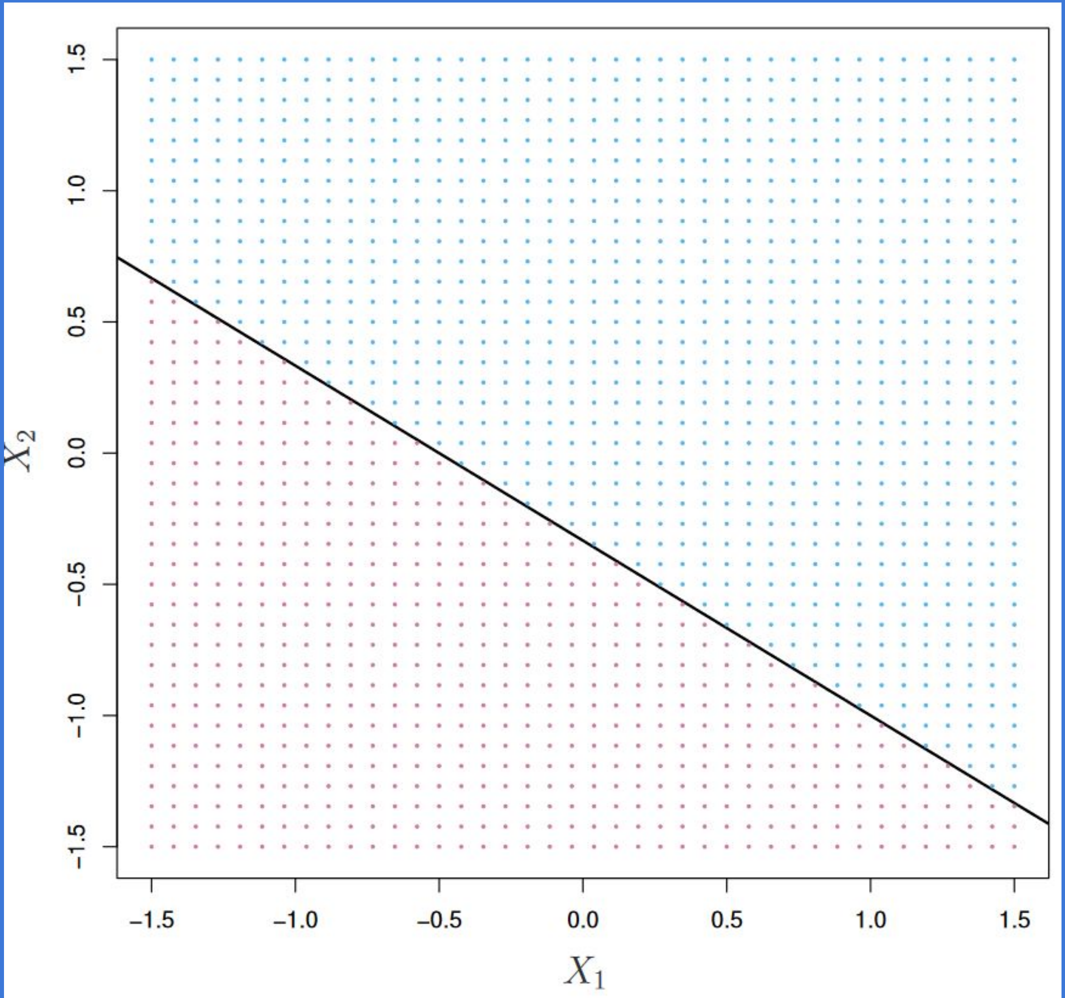
A hyperplane in $\mathbb{R}^3$ is a plane





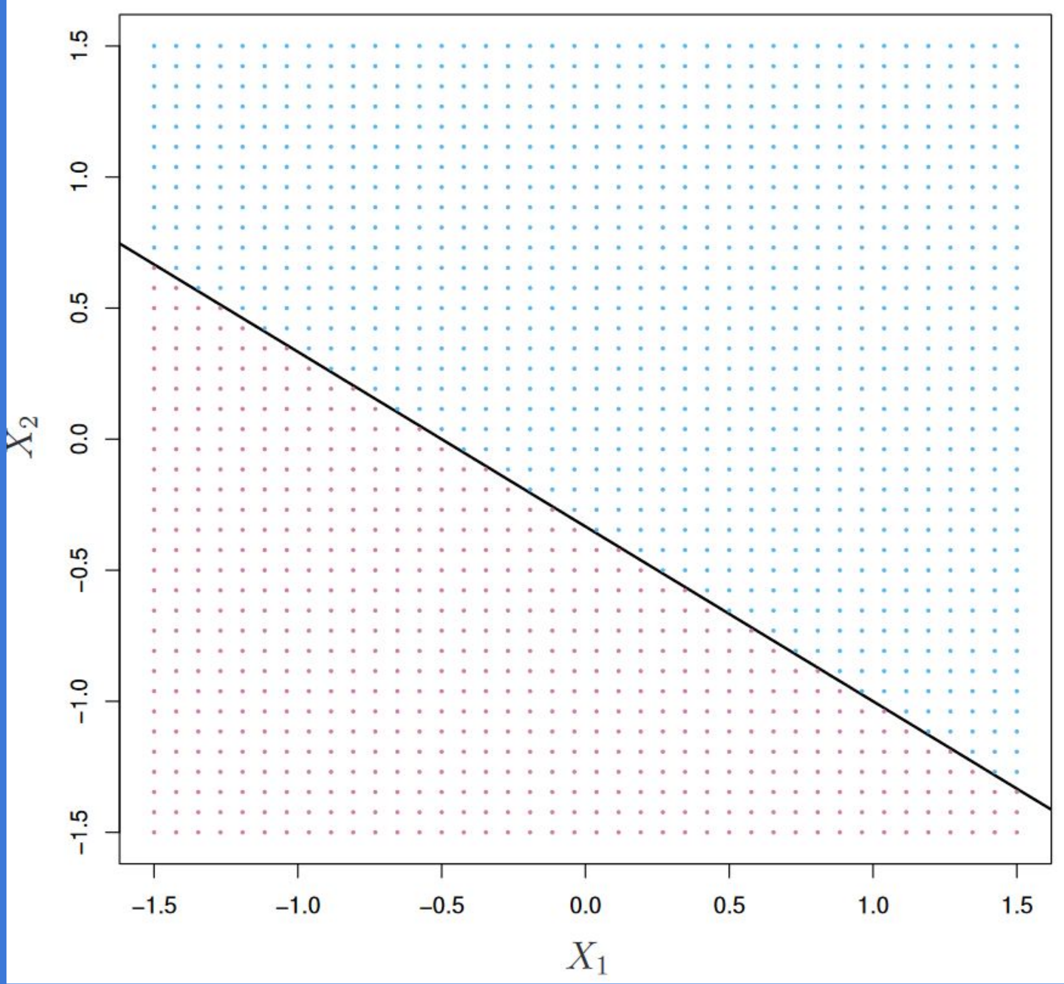
Intuition

If data points do not lie on the hyperplane, then they are one either side of the hyperplane.



Intuition

In general, if the data can be perfectly separated using a hyperplane, then there is an **infinite number** of hyperplanes between the data points.



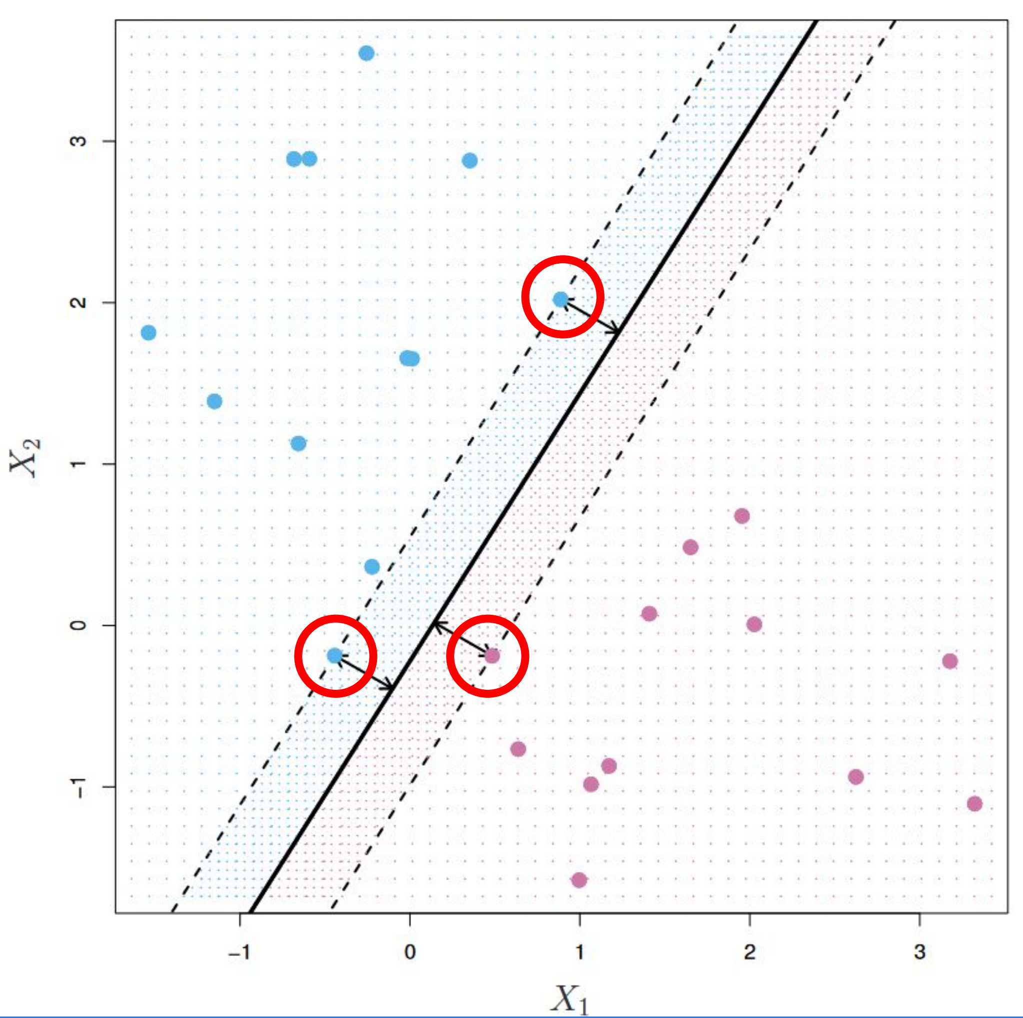
Intuition

In general, if the data can be perfectly separated using a hyperplane, then there is an **infinite number** of hyperplanes between the data points.

Maximal Margin Hyperplane

The separating hyperplane that is farthest from the observations.

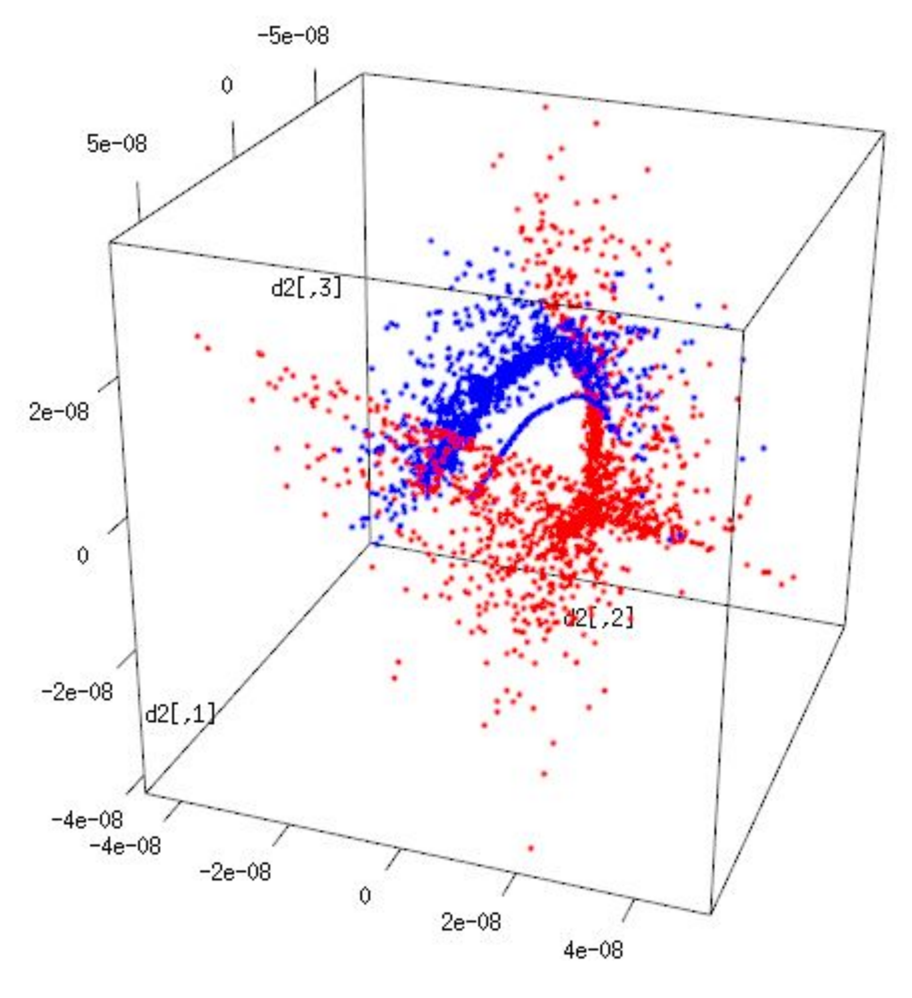
Perpendicular distance from each data point contributes to the margin. The maximal margin hyperplane is essentially just maximizes the distance.



Support Vectors

The three equidistant closest points are the support vectors.

If their position changes, the hyperplane also changes. This means that the hyperplane is only dependent on the support vectors, and not other points.



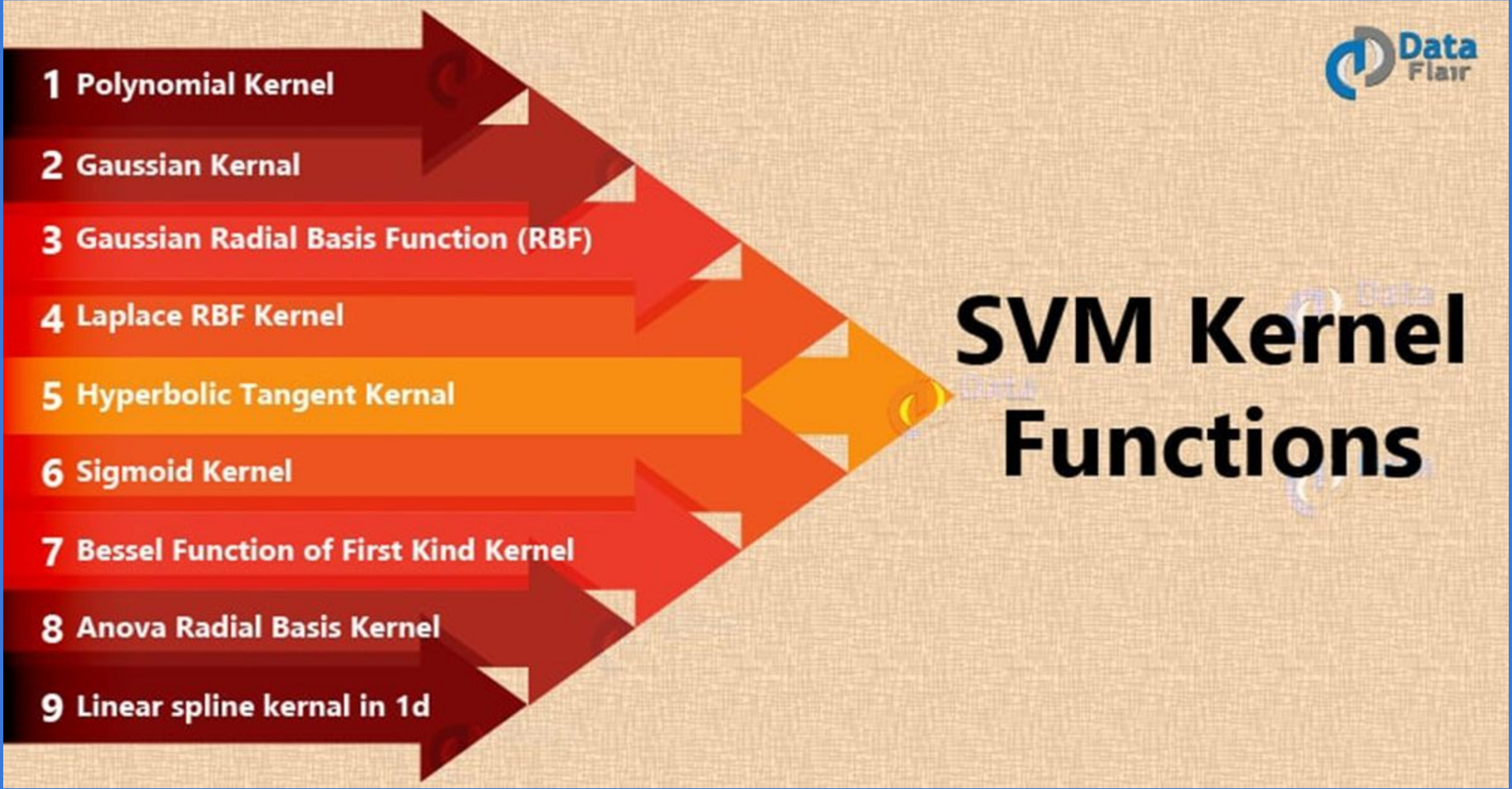
Non-Linearity

In this case, there is no maximal margin classifier.

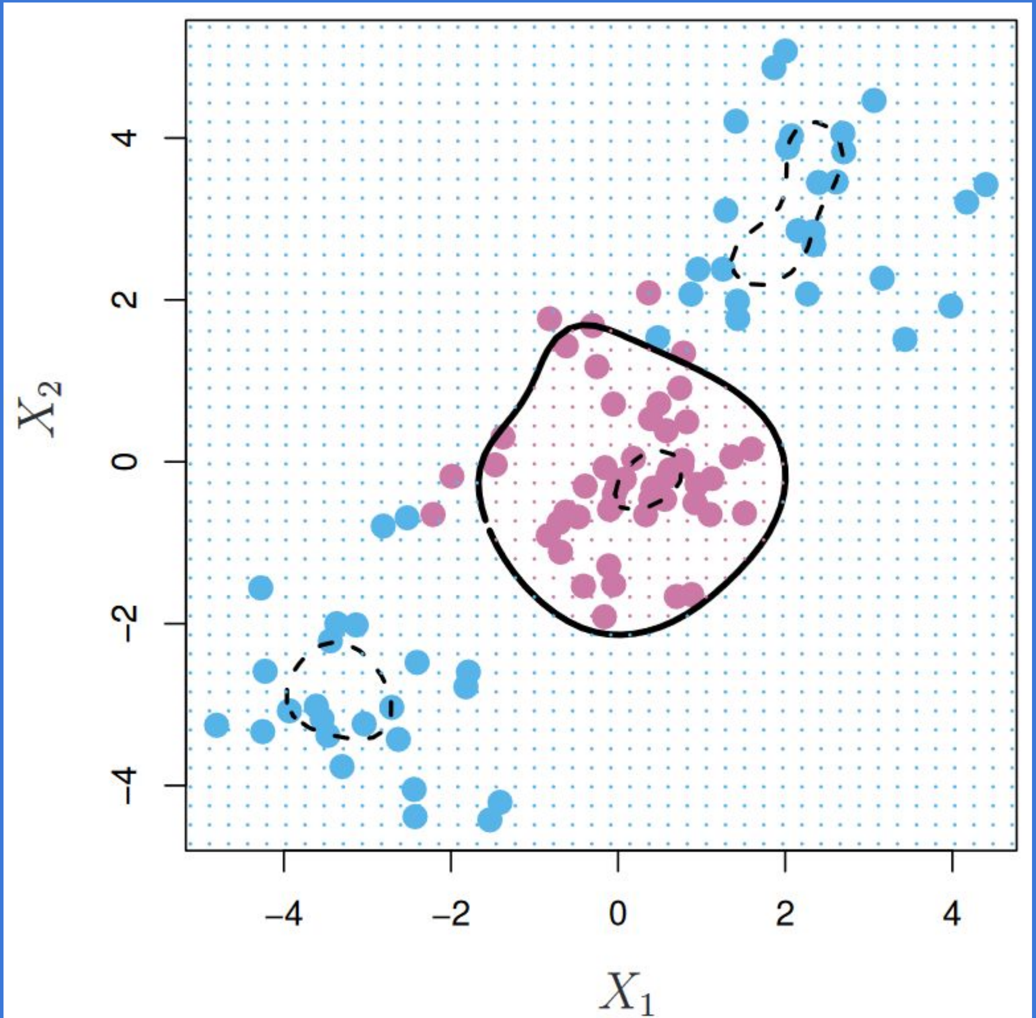
Instead of a hyperplane, use a support vector classifier that can *almost* separate the classes using a softer margin called support vector classifier.

Kernels

A kernel function measures the similarity between two data points. The idea of similarity is task-dependent. If your task is object recognition, a good kernel will assign a high score to a pair of images that contain the same objects, and a low score to a pair of images with different objects.

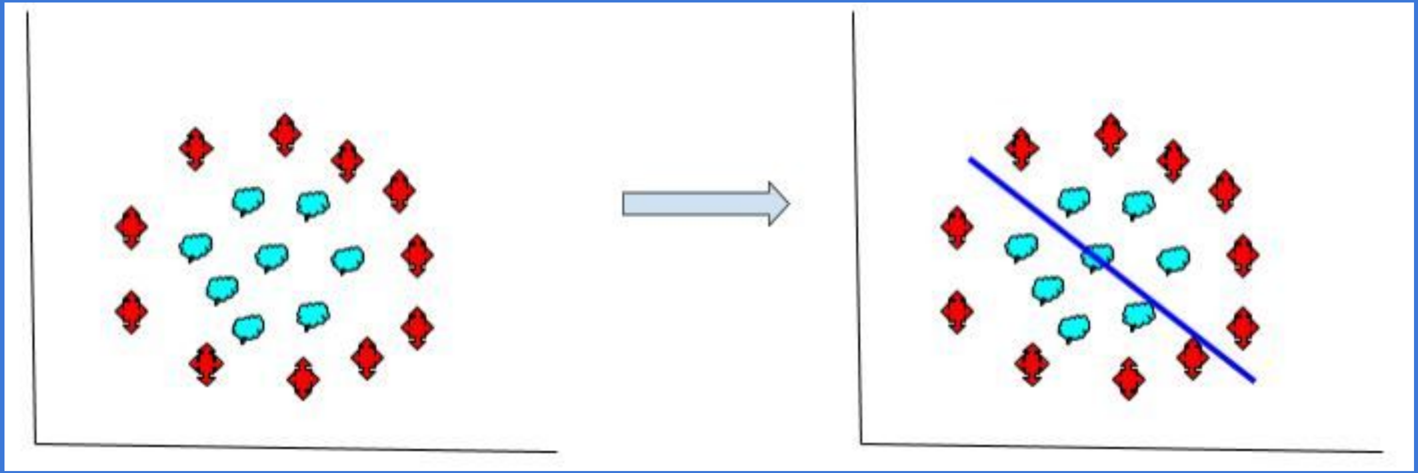


1 Polynomial Kernel
2 Gaussian Kernal
3 Gaussian Radial Basis Function (RBF)
4 Laplace RBF Kernel
5 Hyperbolic Tangent Kernal
6 Sigmoid Kernel
7 Bessel Function of First Kind Kernel
8 Anova Radial Basis Kernel
9 Linear spline kernal in 1d
SVM Kernel Functions
Data Flair

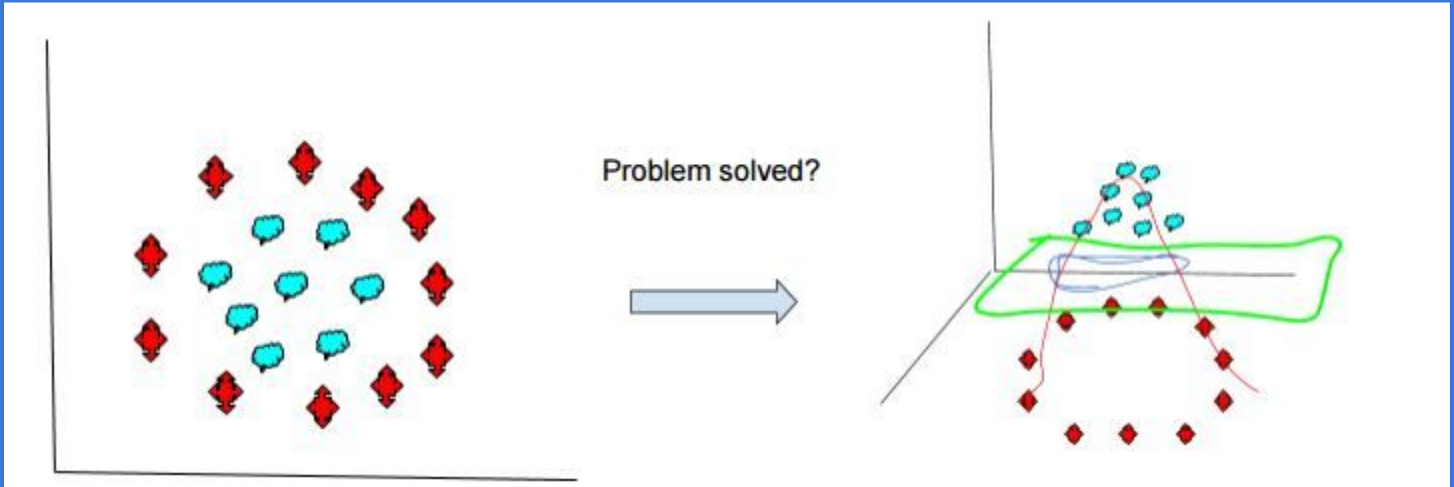


SVM with Radial Kernel

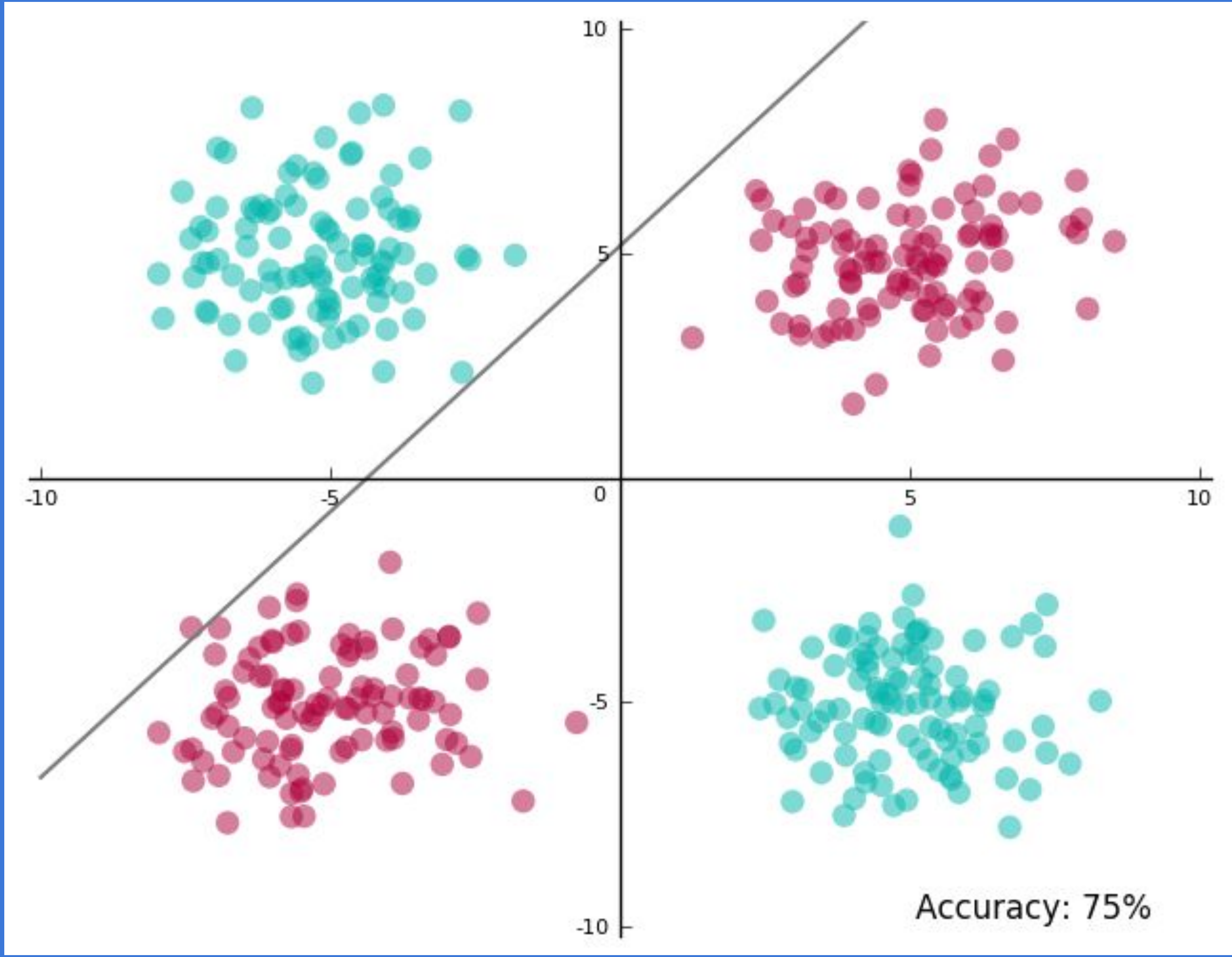
Kernel: A function that maps the data to a higher dimension where the data is separable



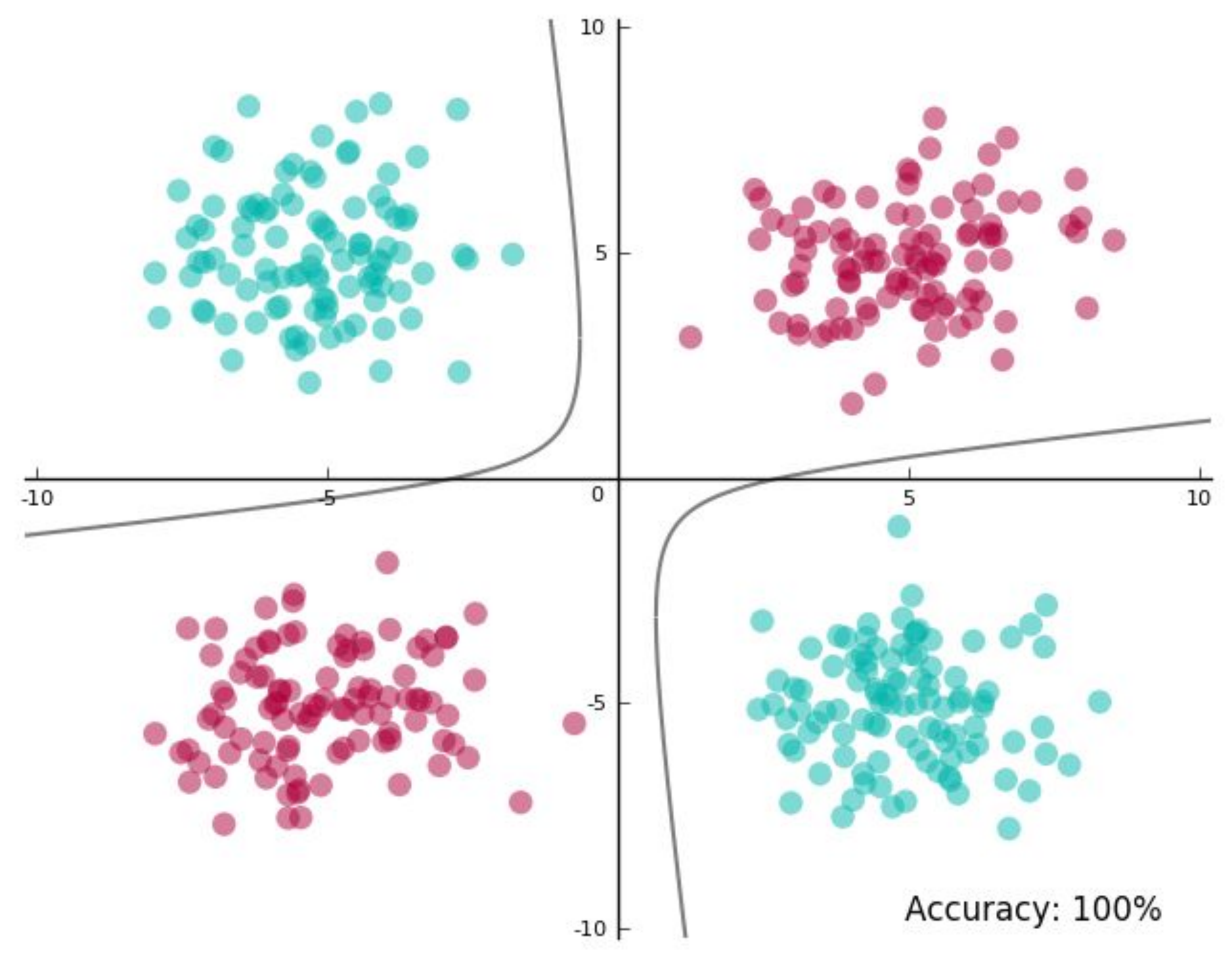
Non-linearly separable data, hyperplane fails!



Project data onto a higher dimension using a kernel function, and there now is a hyperplane!

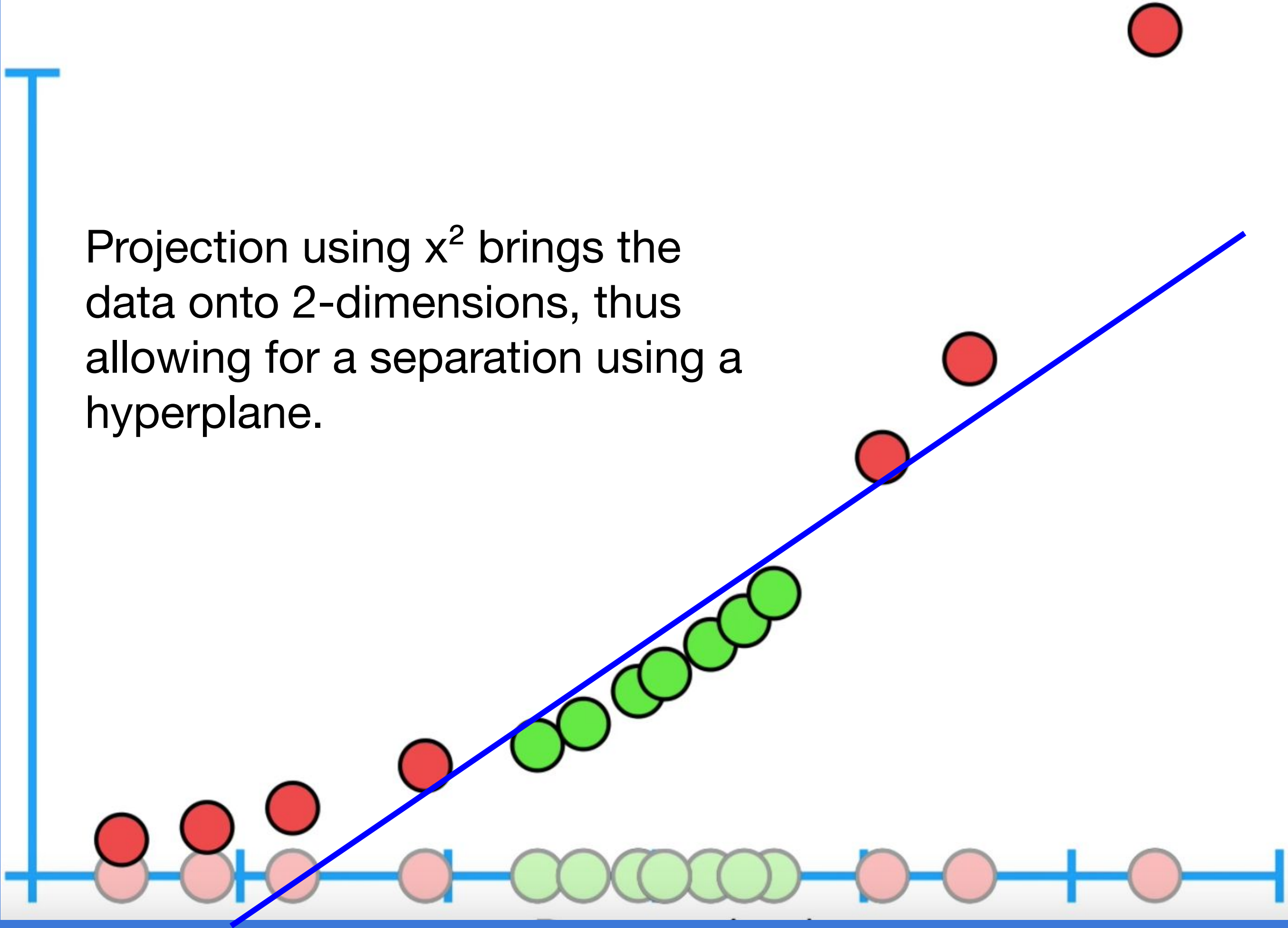


$$X_1 = x_1^2$$
$$X_2 = x_2^2$$
$$X_3 = \sqrt{2}x_1x_2$$



10
5
-10
-5
0
5
10
-5
-10
Accuracy: 100%

Projection using x^2 brings the data onto 2-dimensions, thus allowing for a separation using a hyperplane.



Thank you