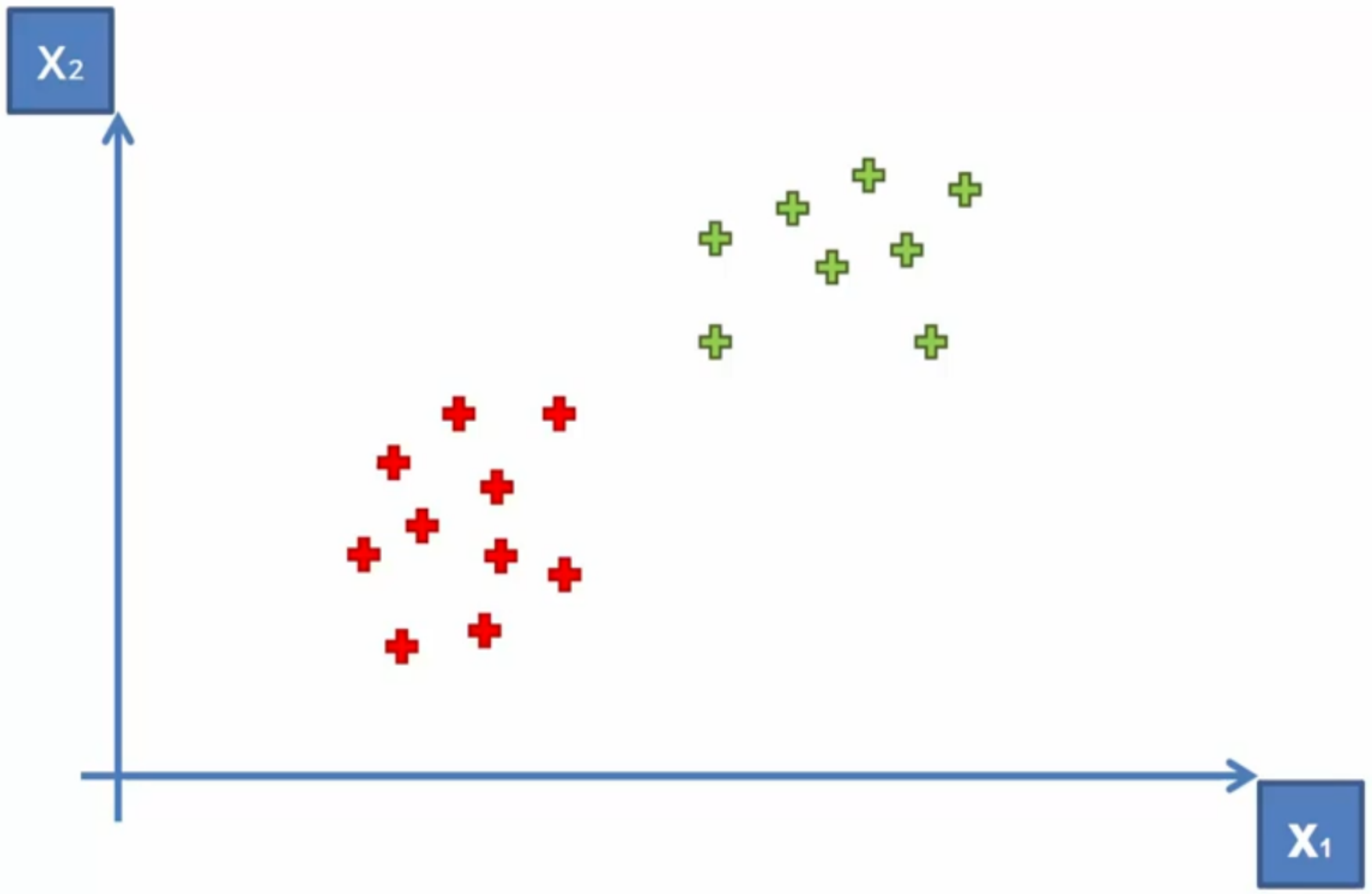
Support Vector Machine

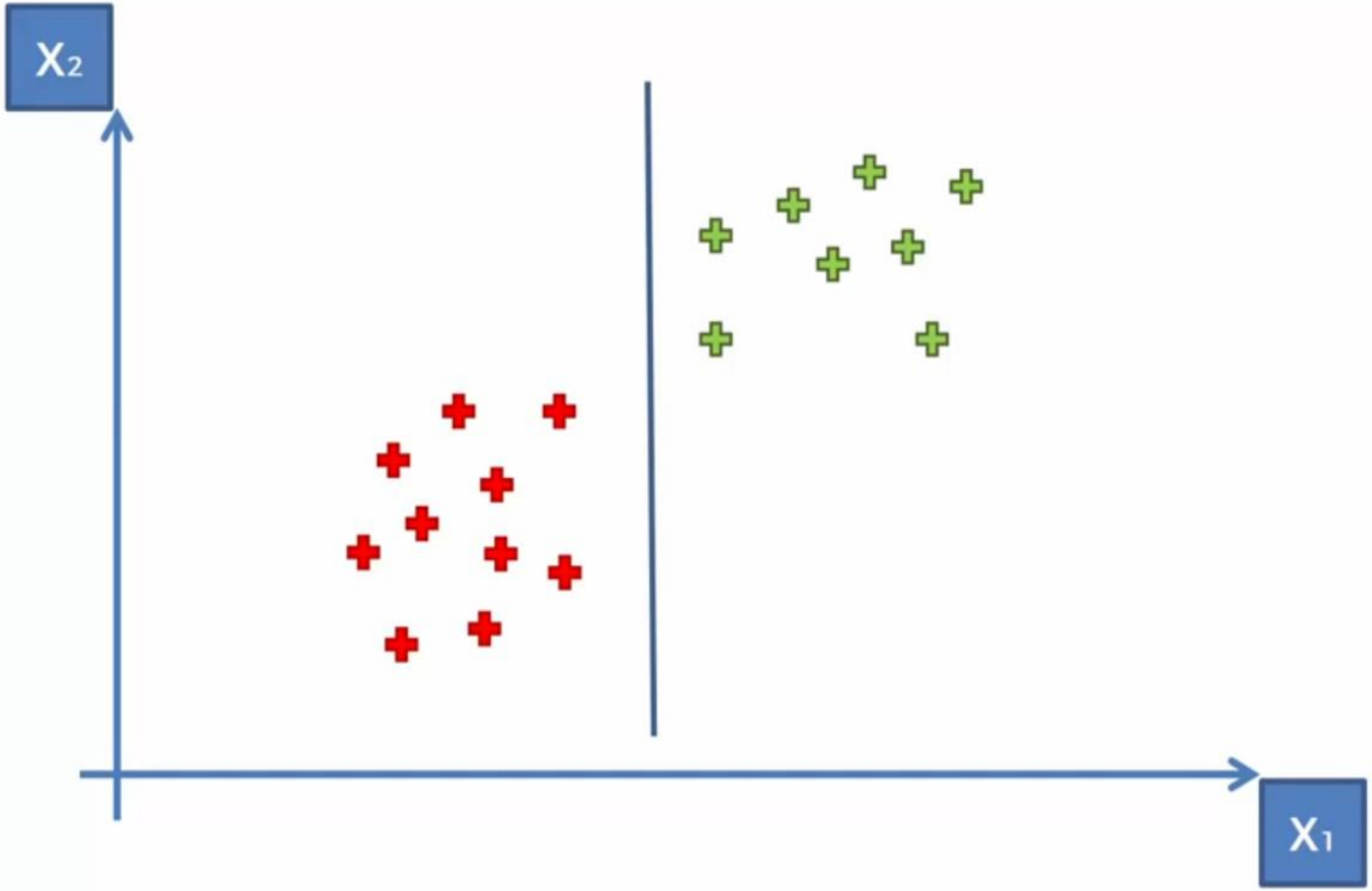
Content

- Support Vector Machine
  - Linearly Separable
  - Non-Linearly Separable
  - Hyperplane / Decision Boundary
  - Margin
  - Marginal Distance
  - Support Vectors
  - Kernel SVM

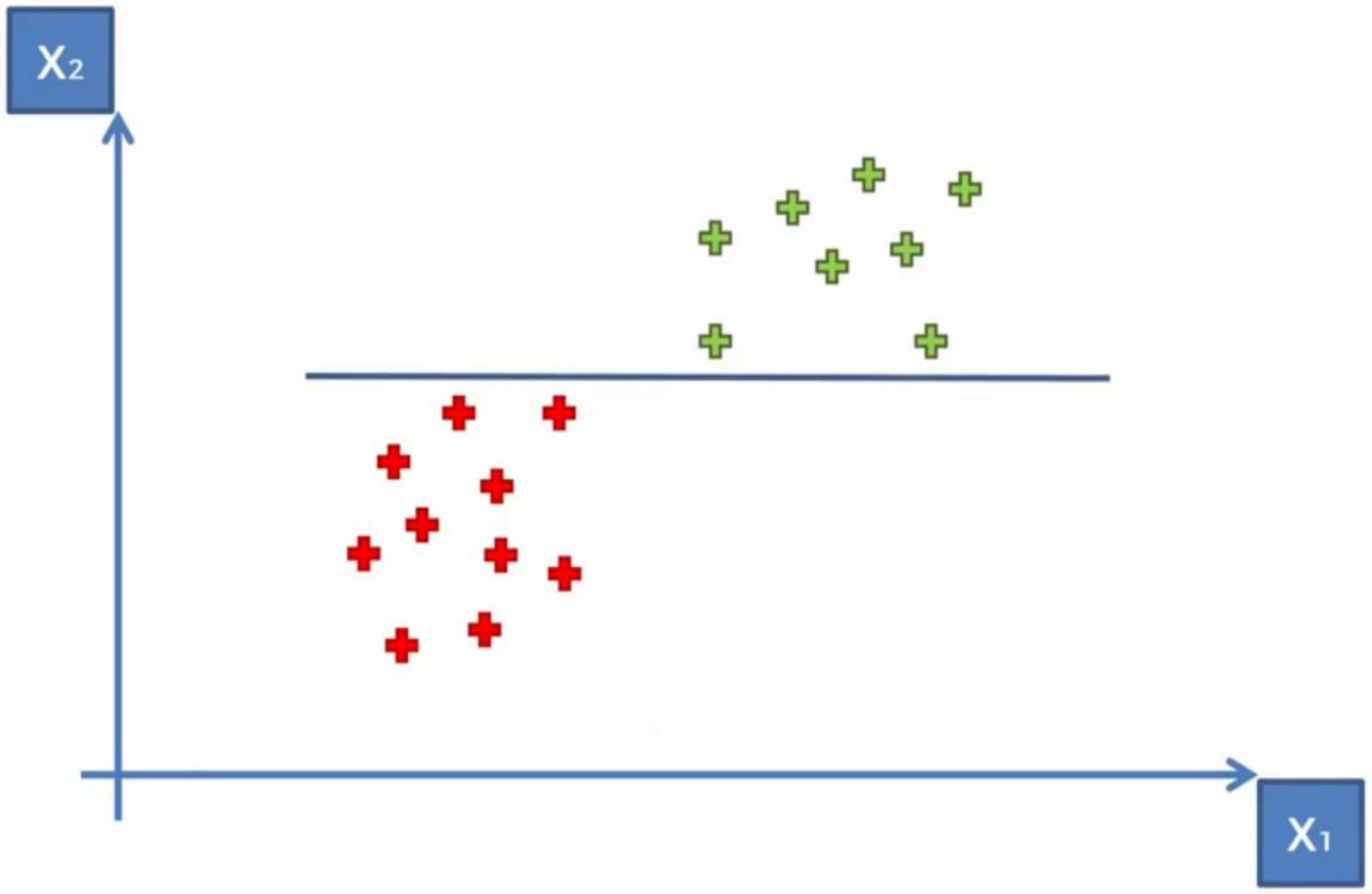
How to separate these points ?



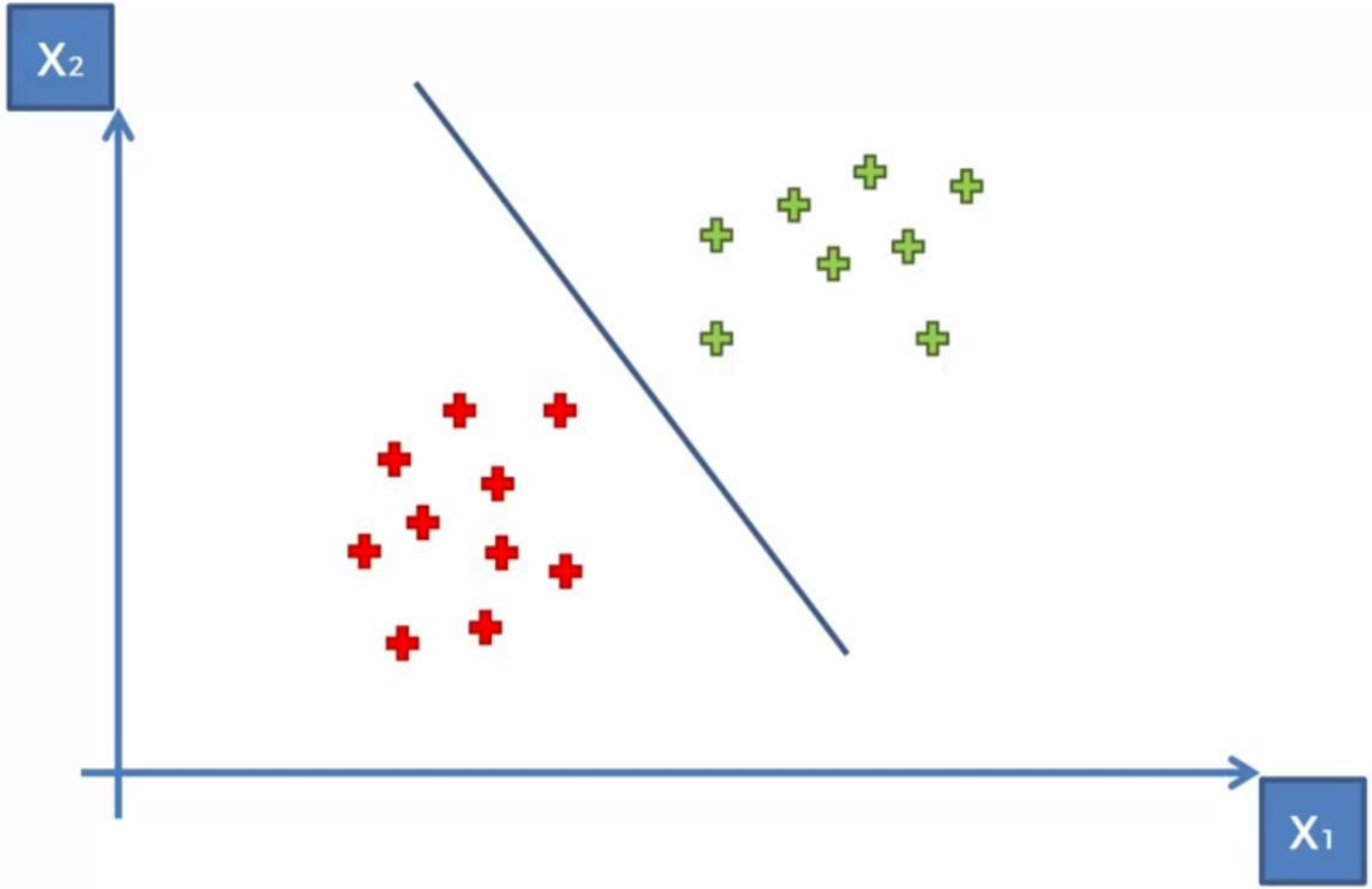
How to separate these points ?



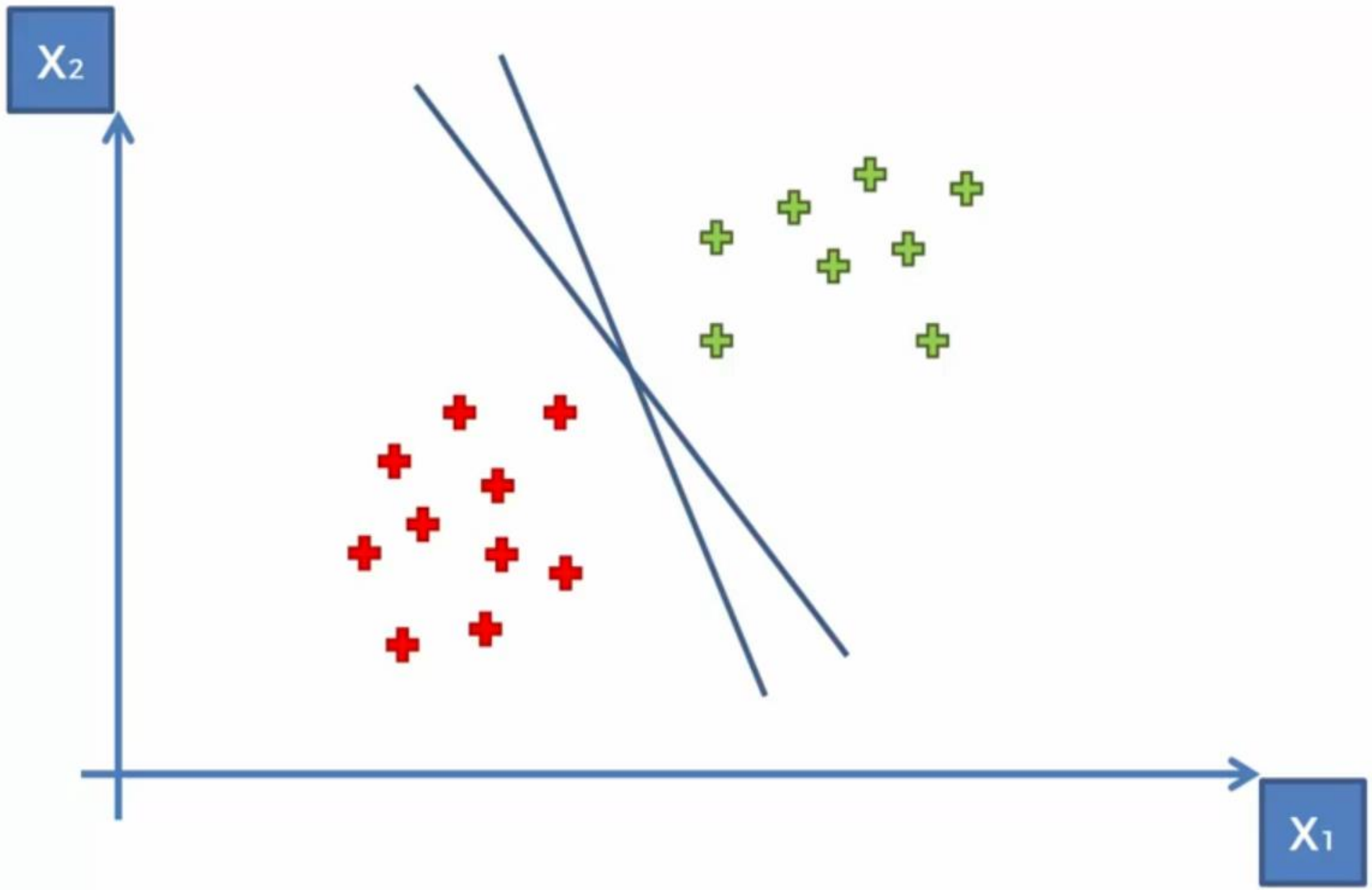
How to separate these points ?



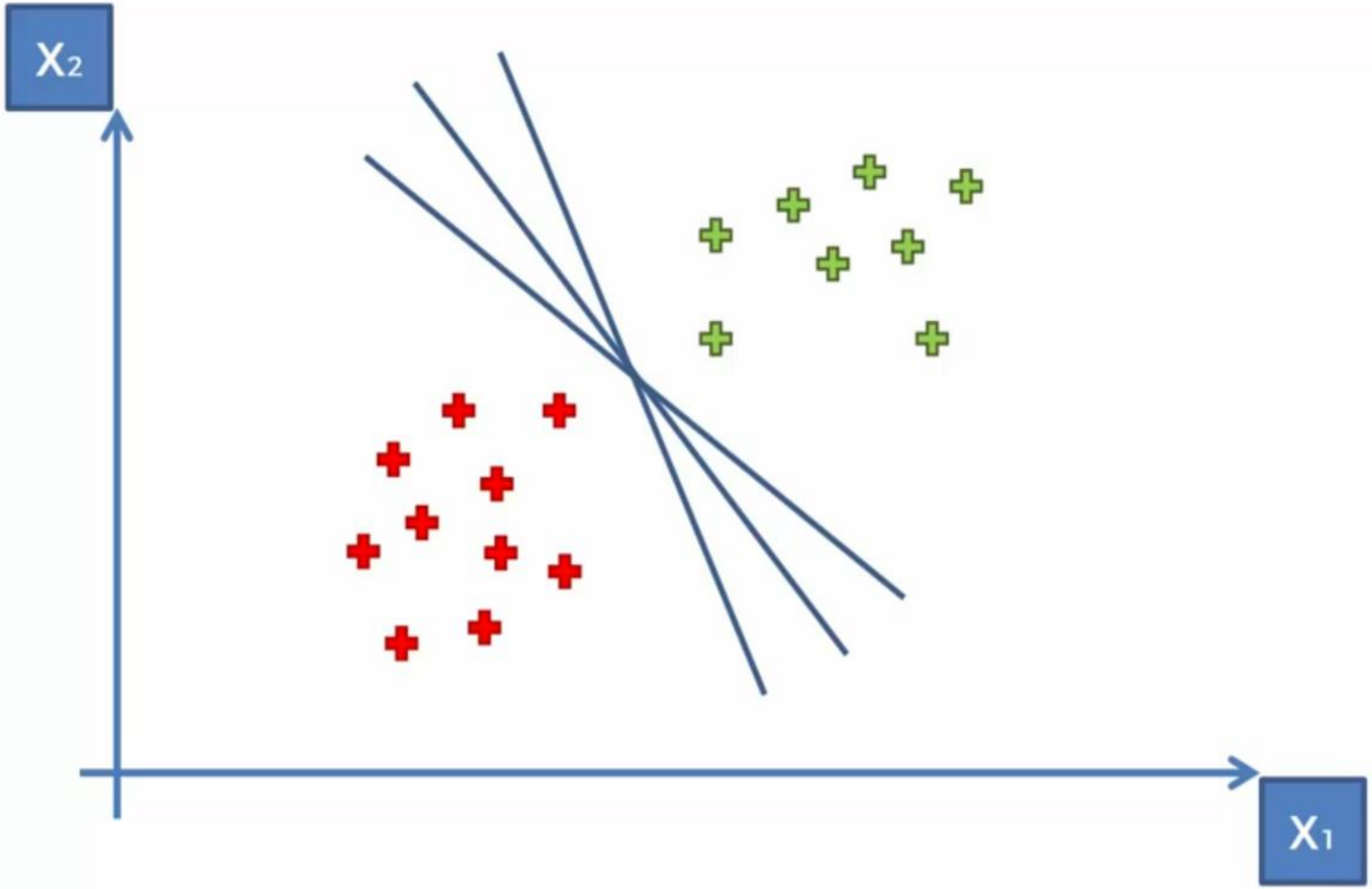
SVM



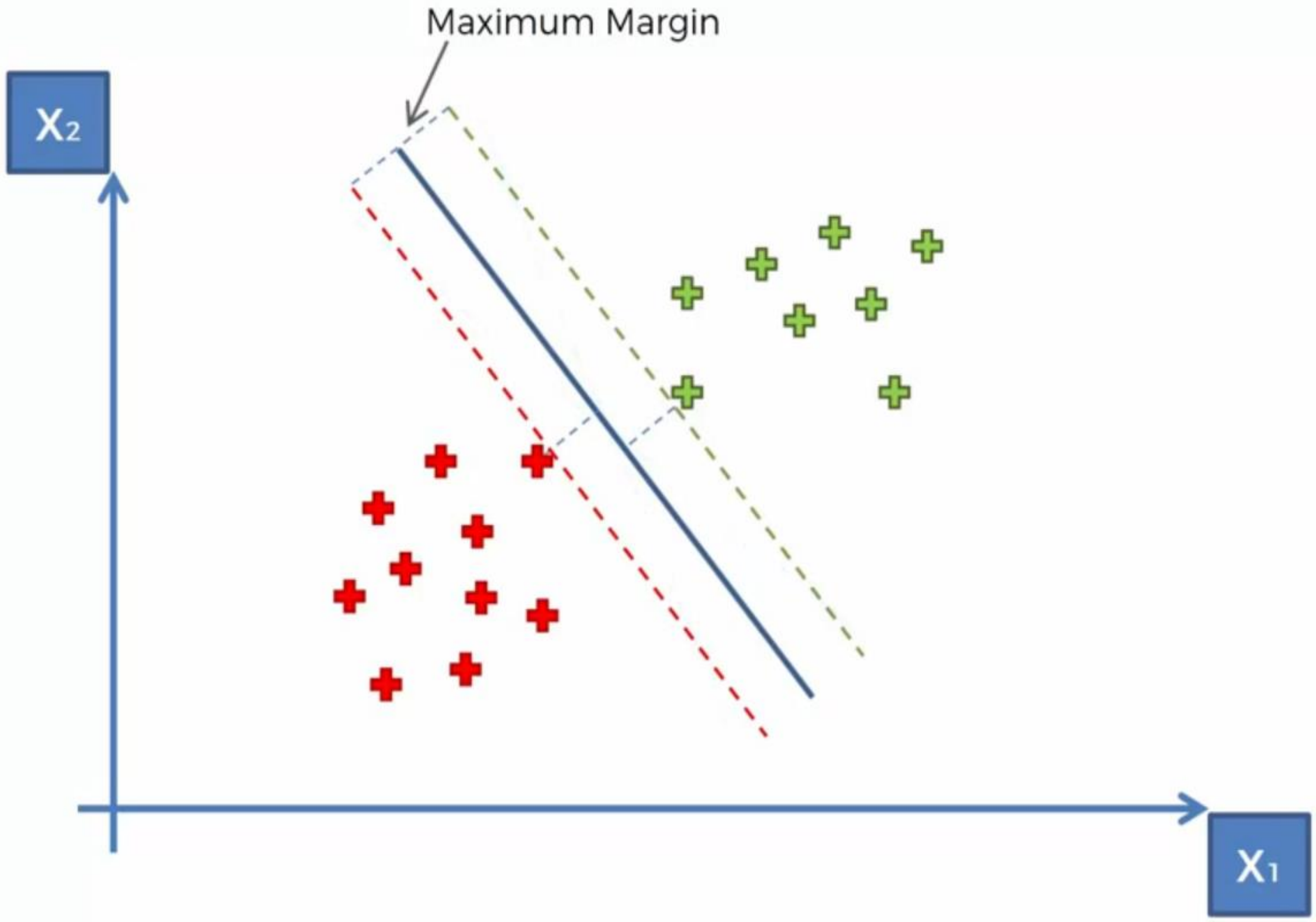
SVM



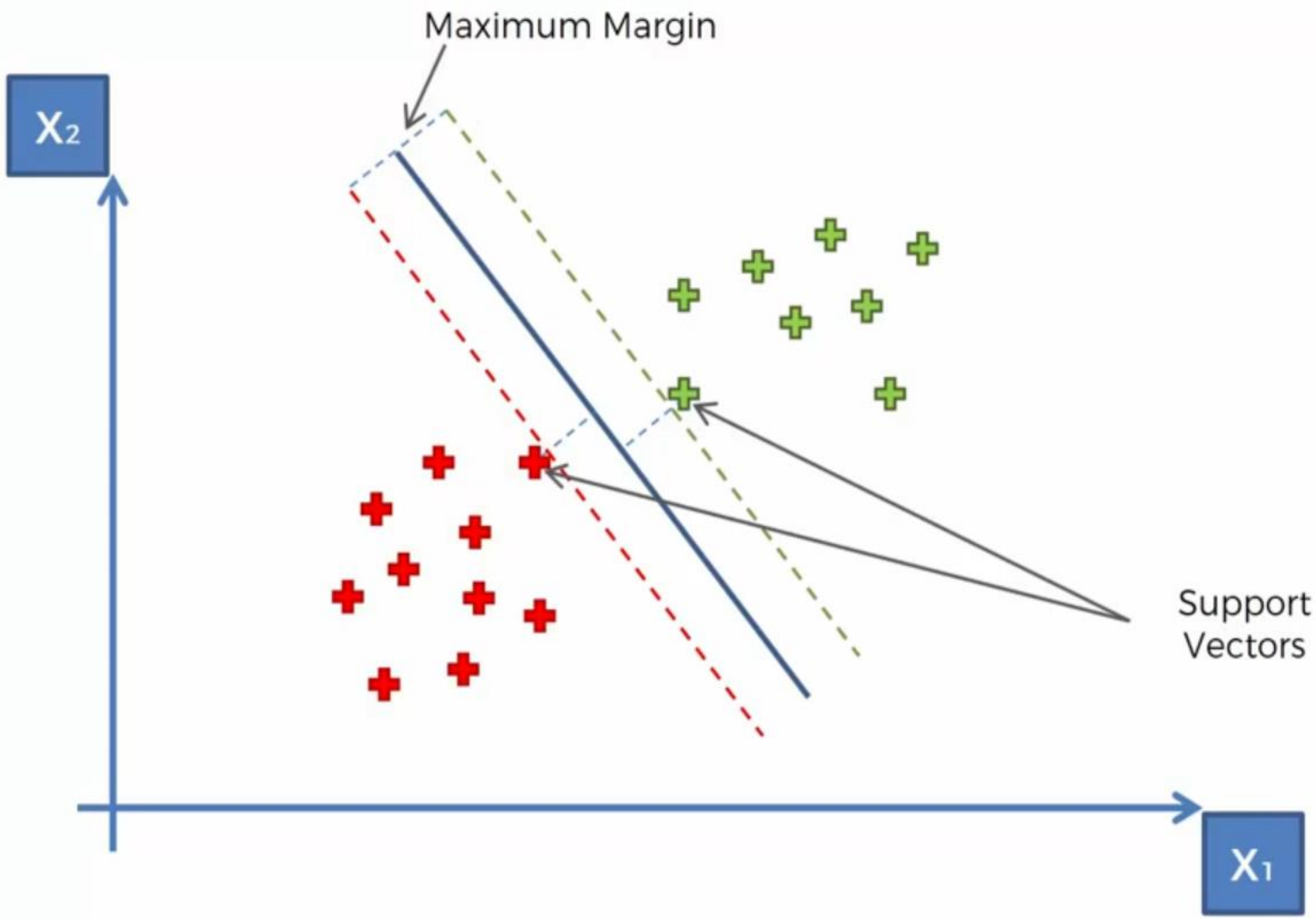
SVM



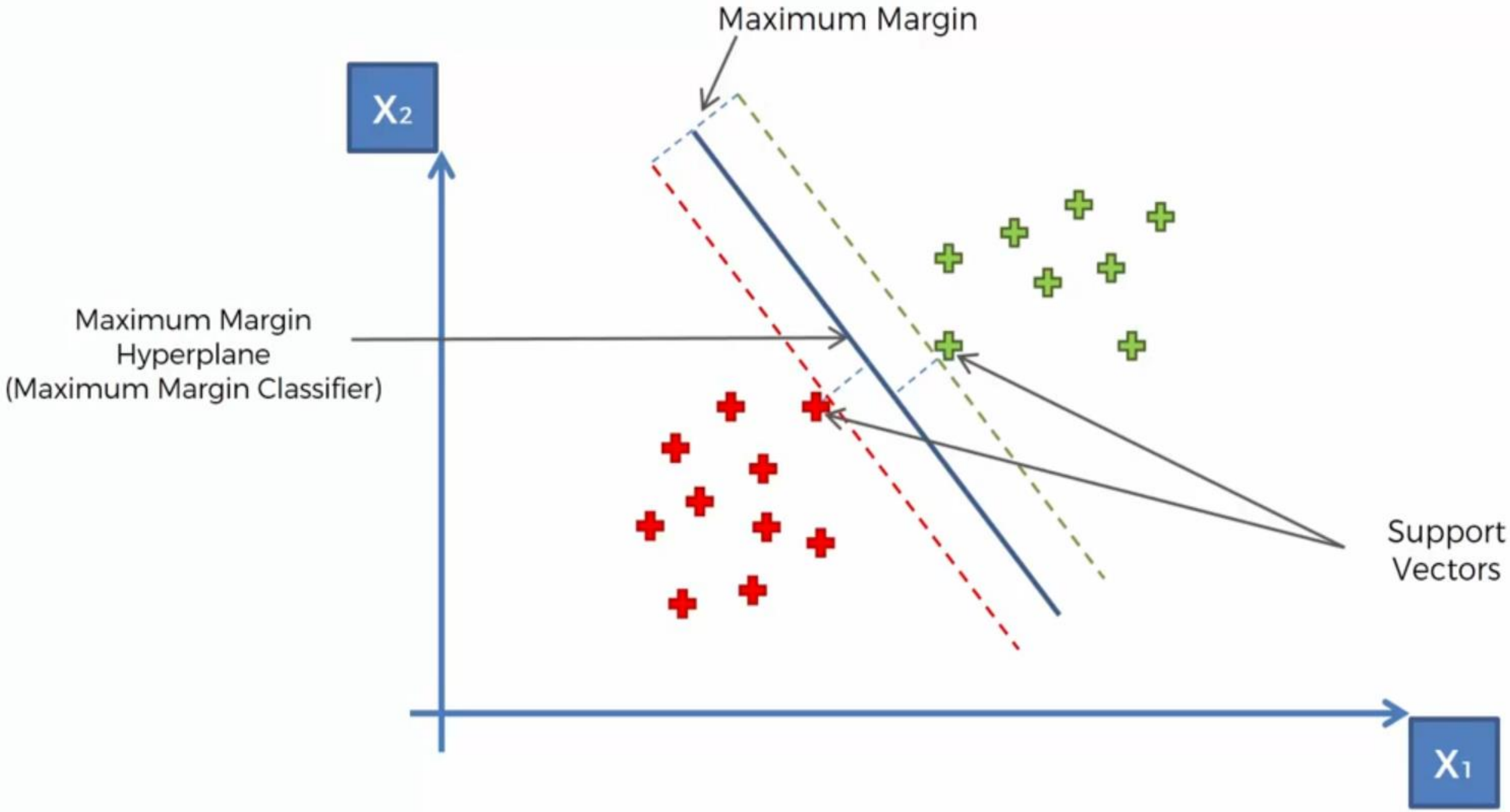
Maximum Margin



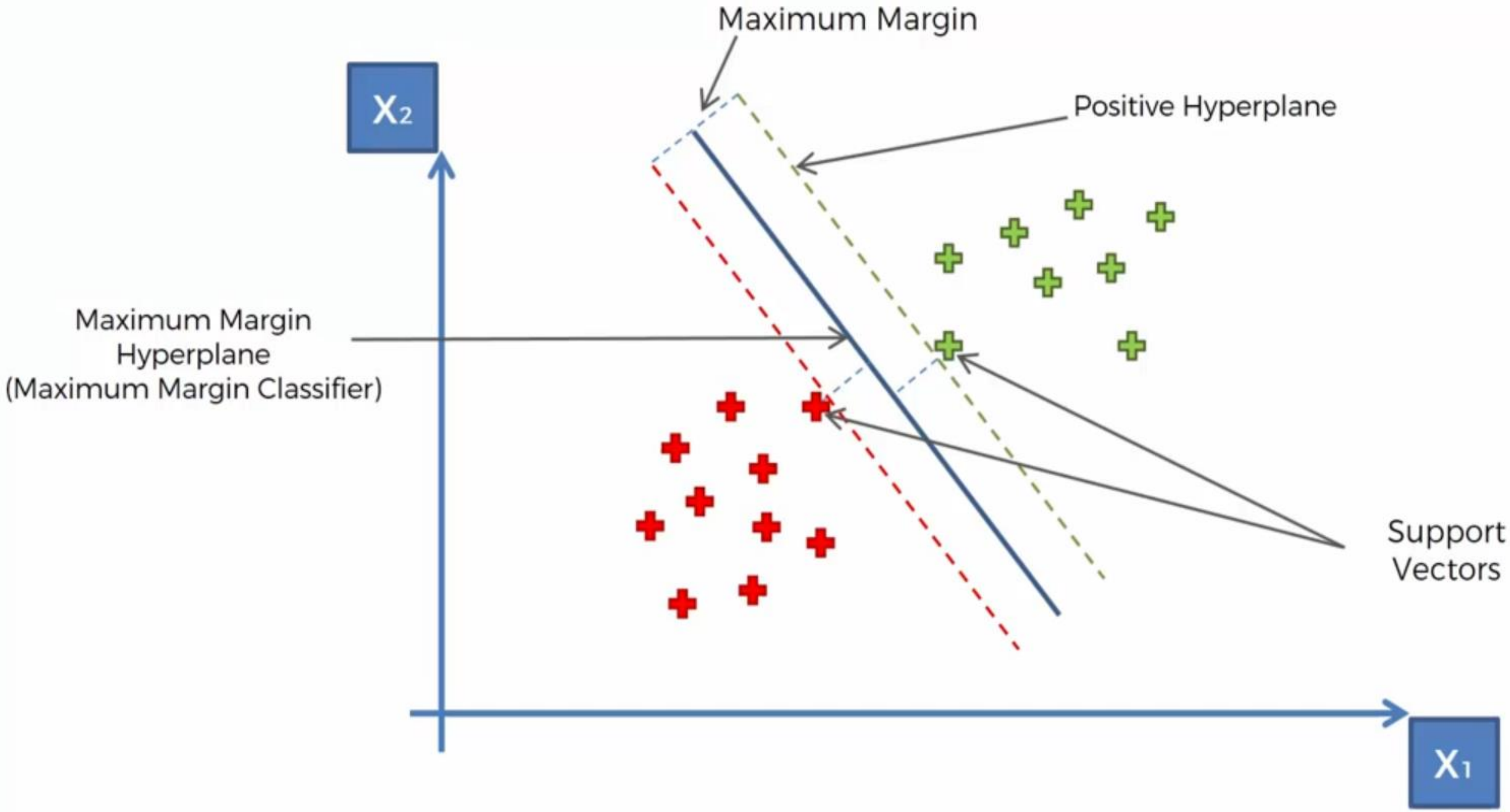
Support Vectors



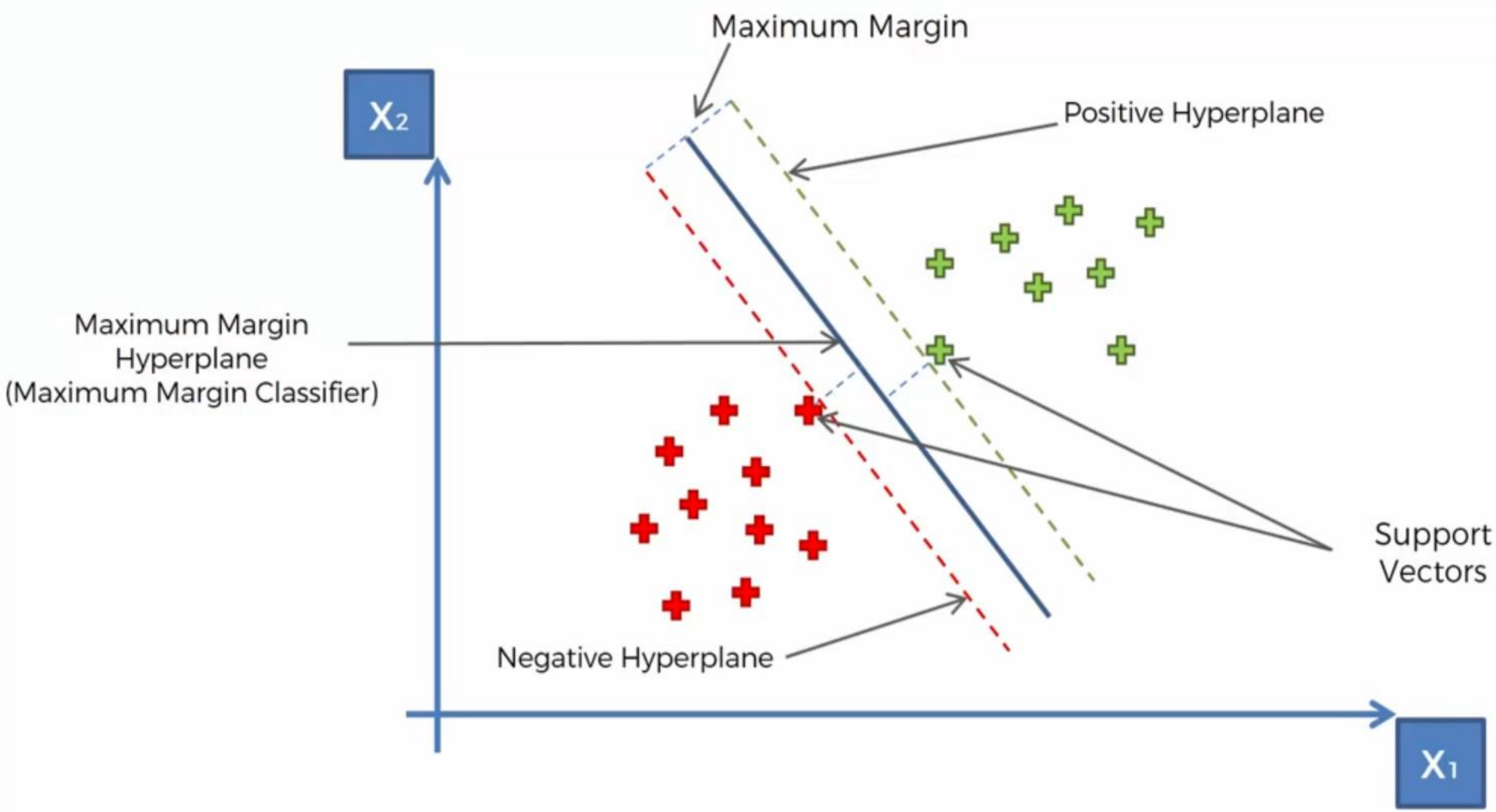
Hyperplanes



Hyperplanes

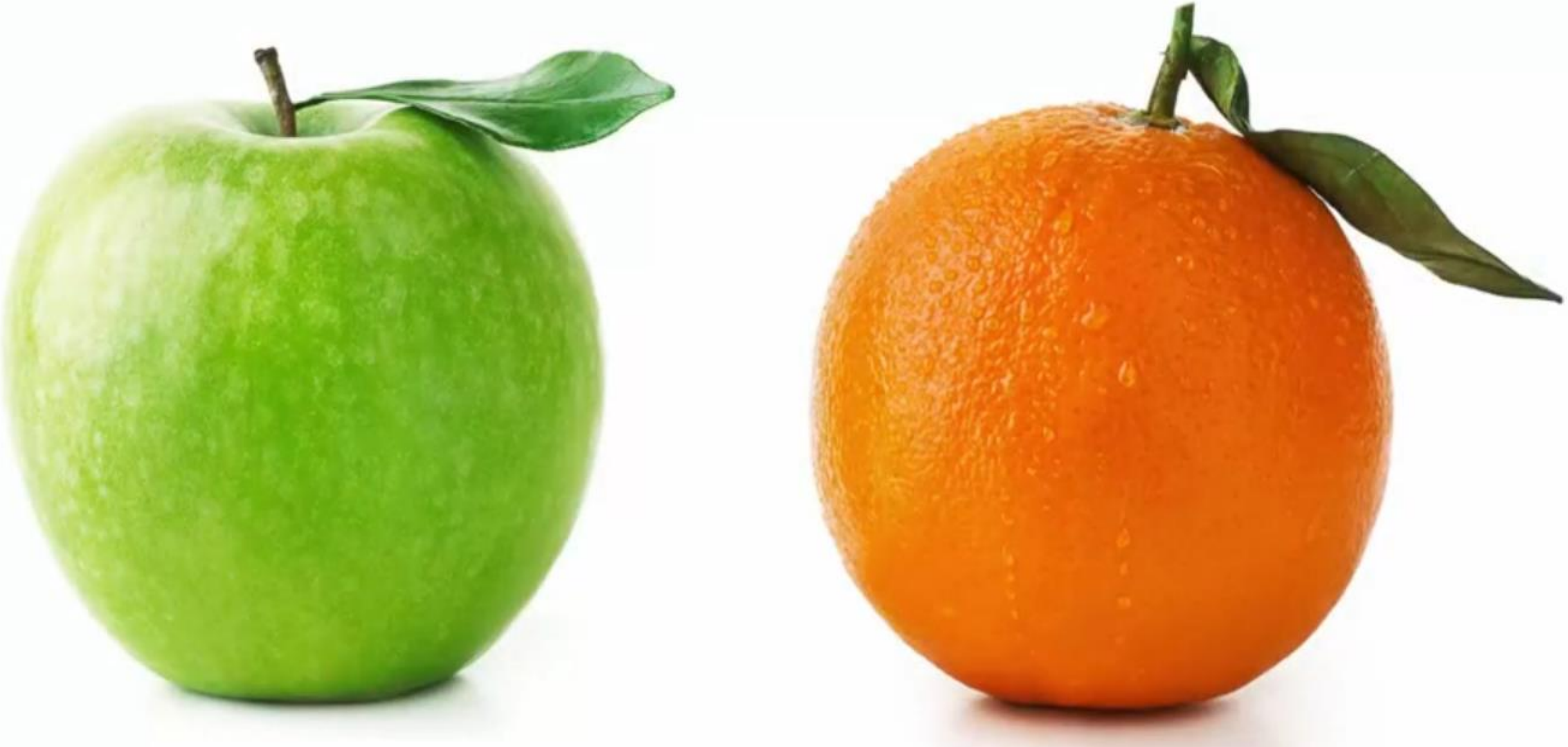


Hyperplanes

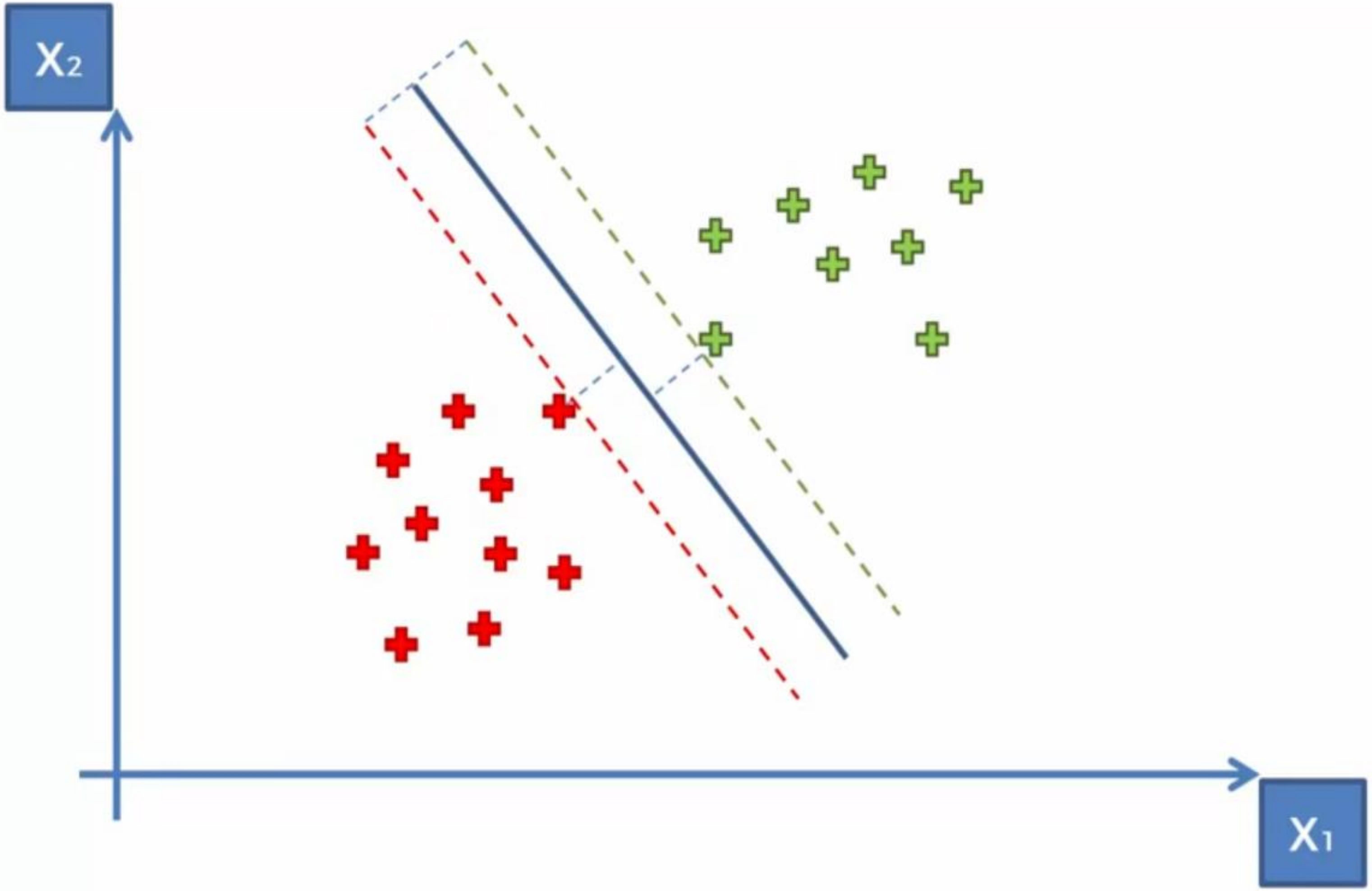


What’s So Special About SVMs?

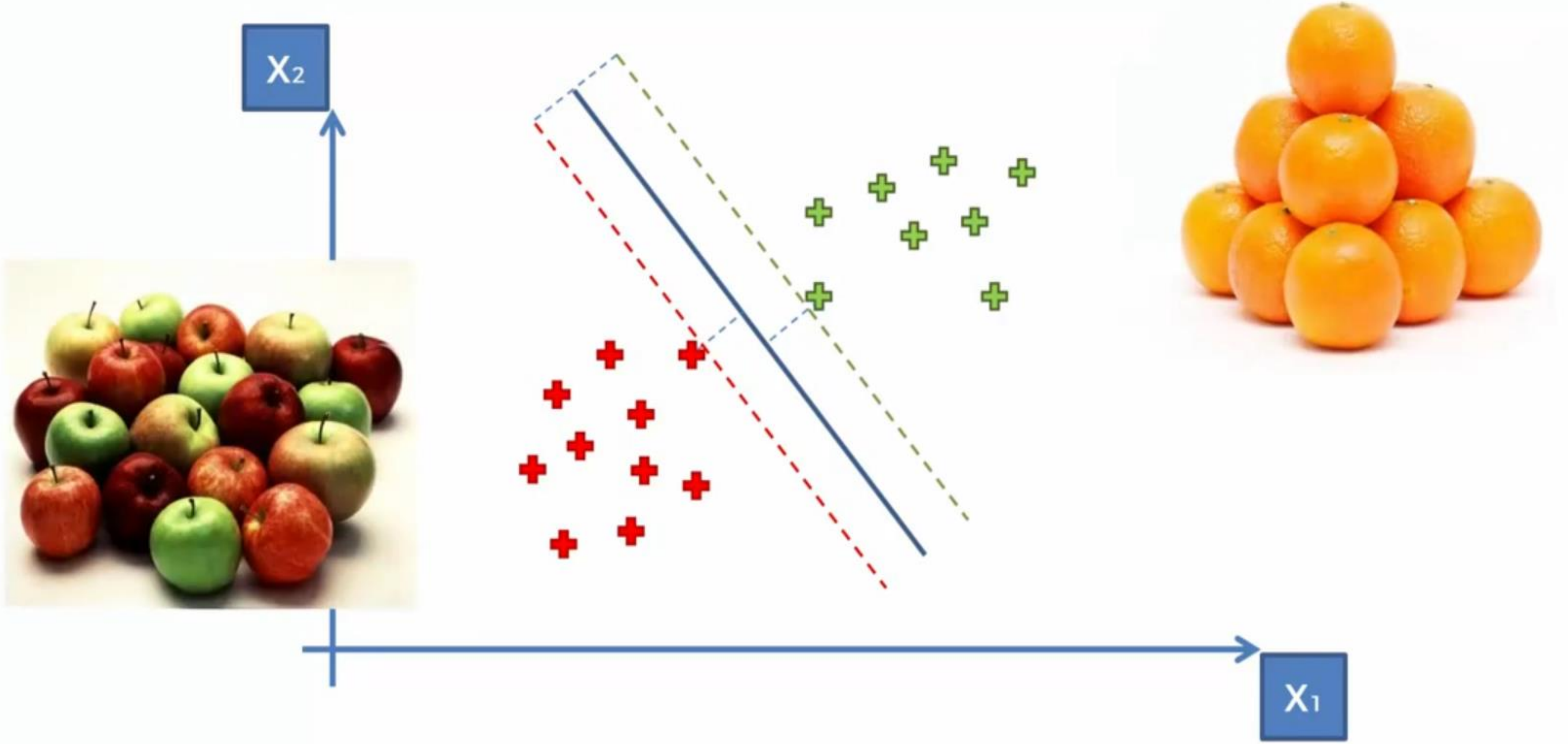
What's So Special About SVMs?



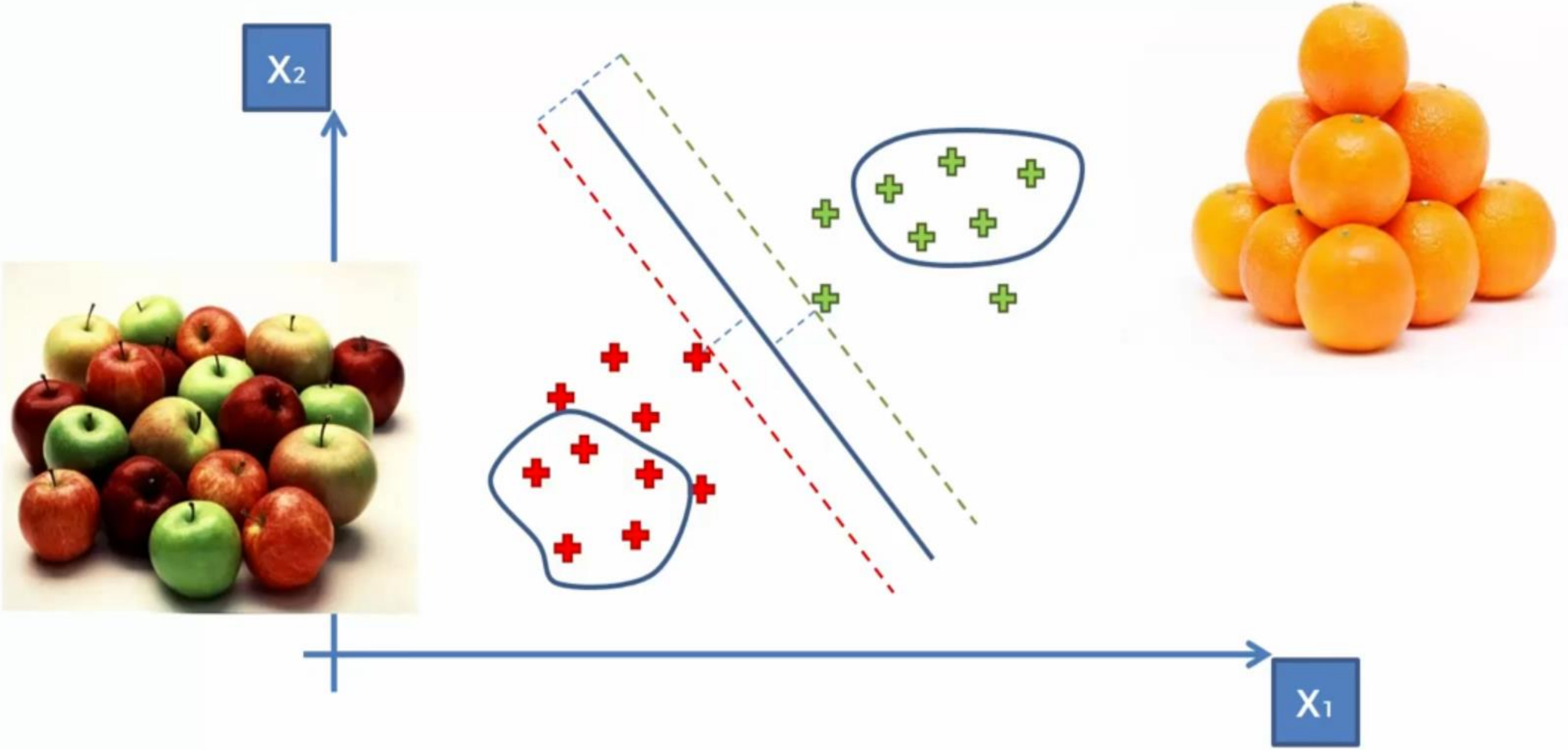
What's So Special About SVMs?



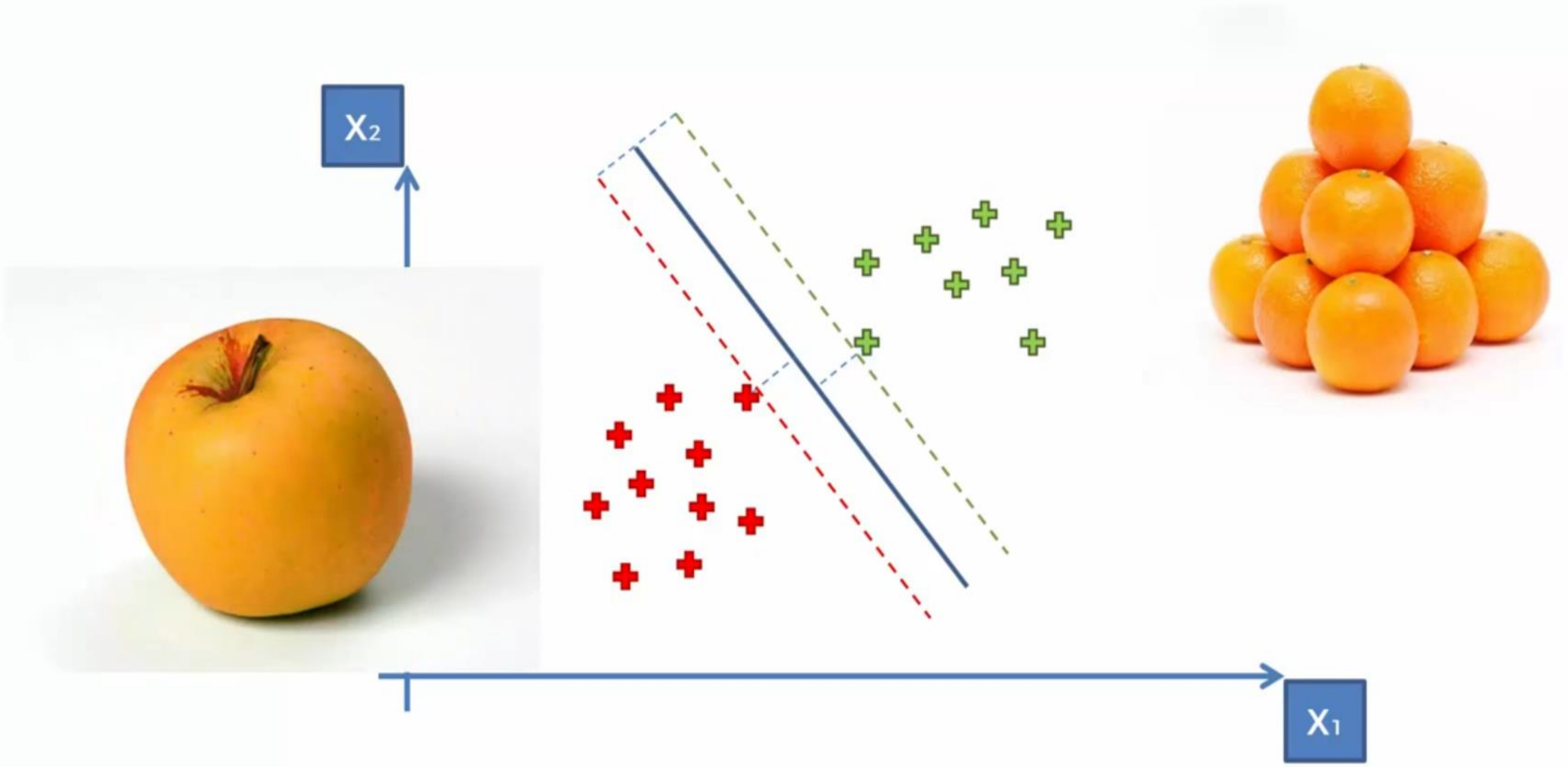
What's So Special About SVMs?



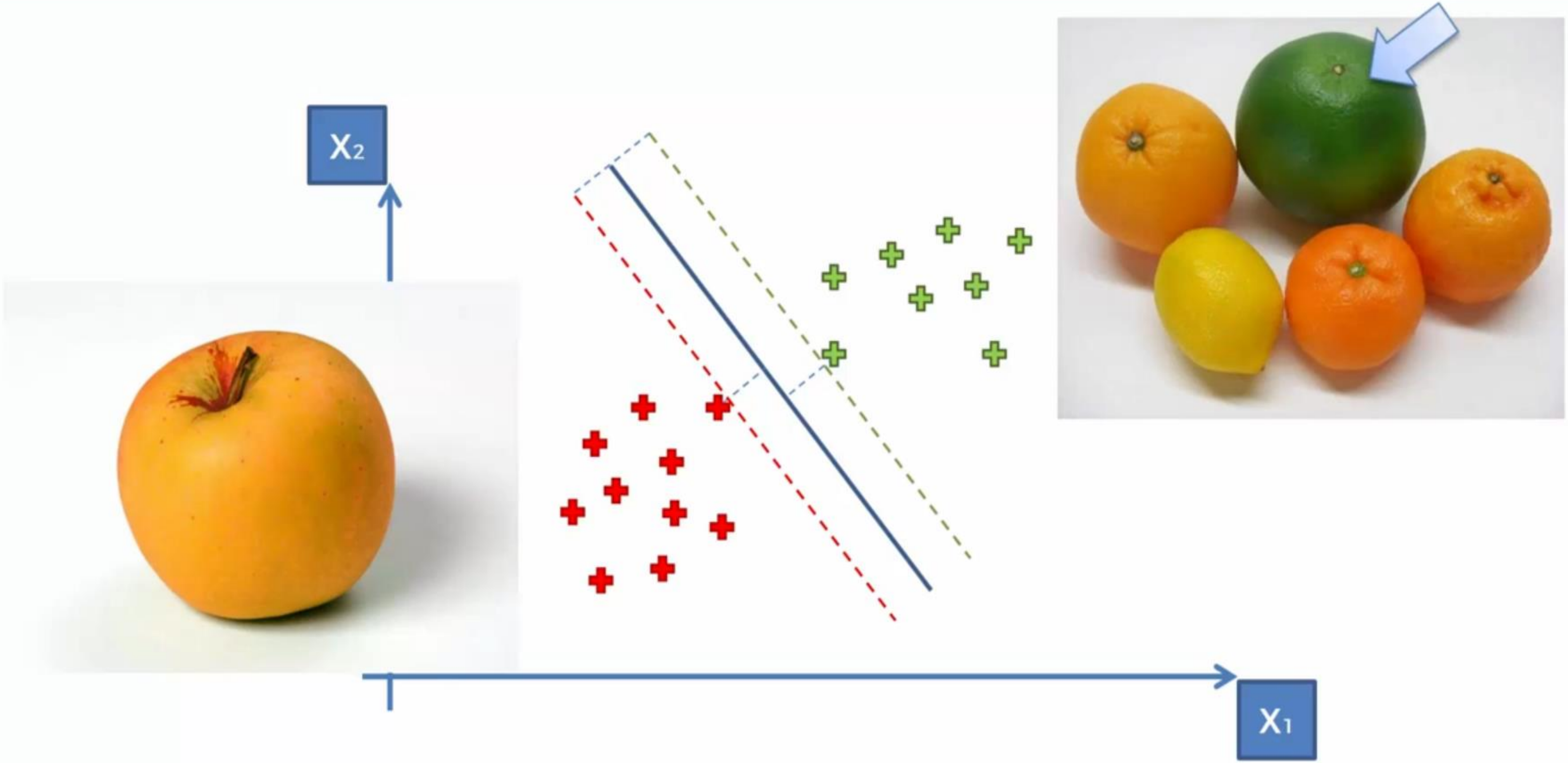
What's So Special About SVMs?



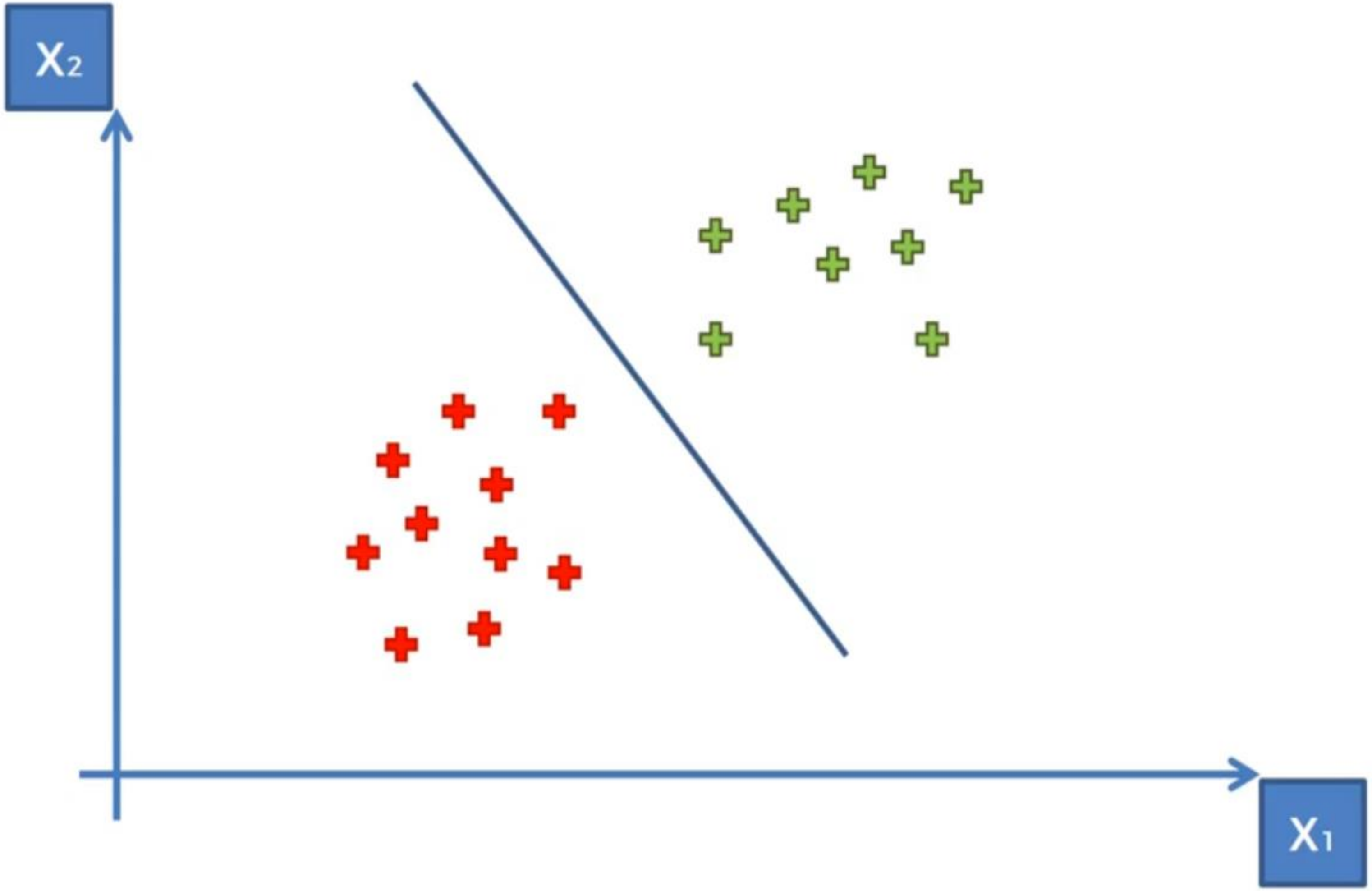
What's So Special About SVMs?



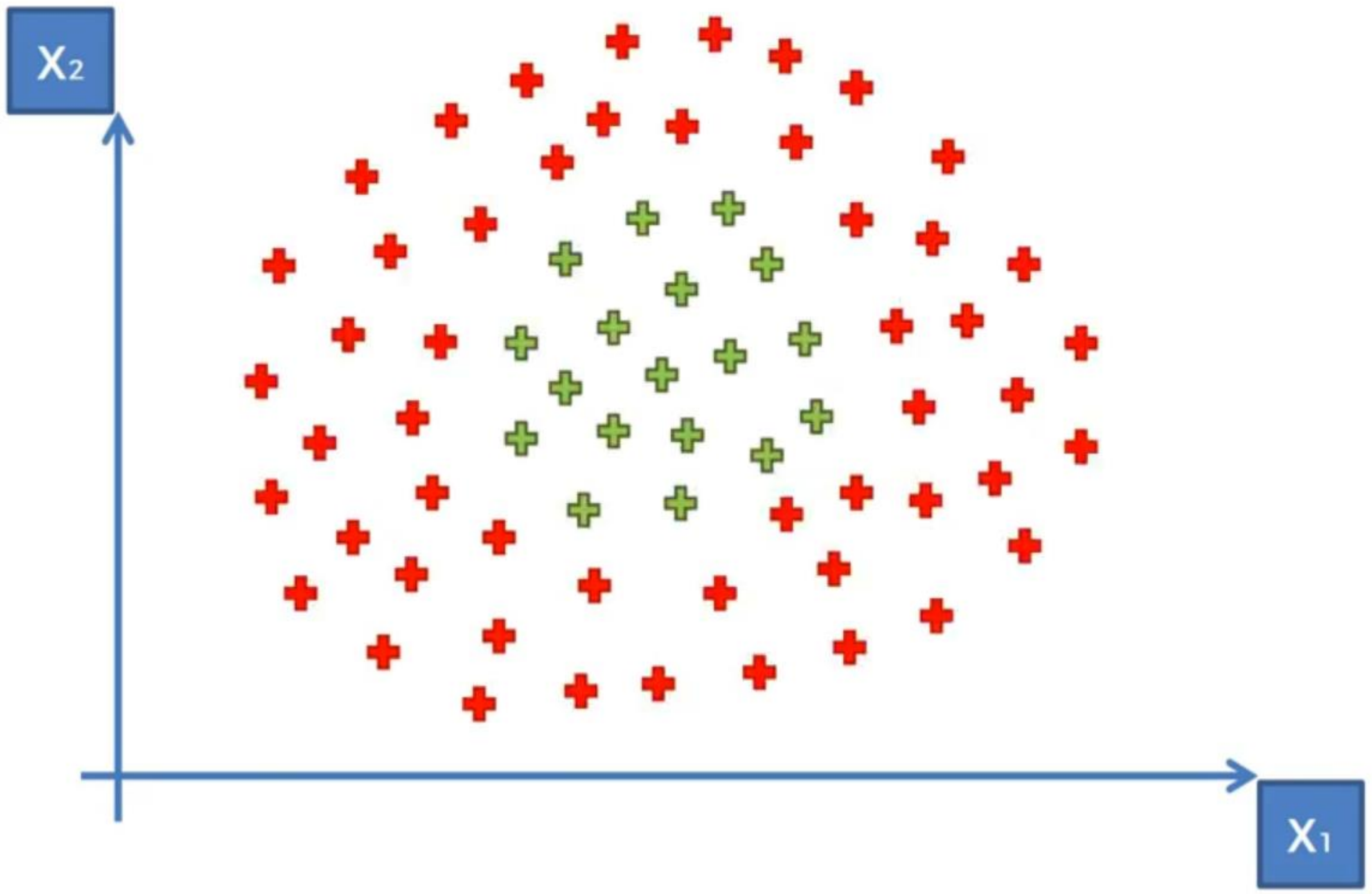
What's So Special About SVMs?



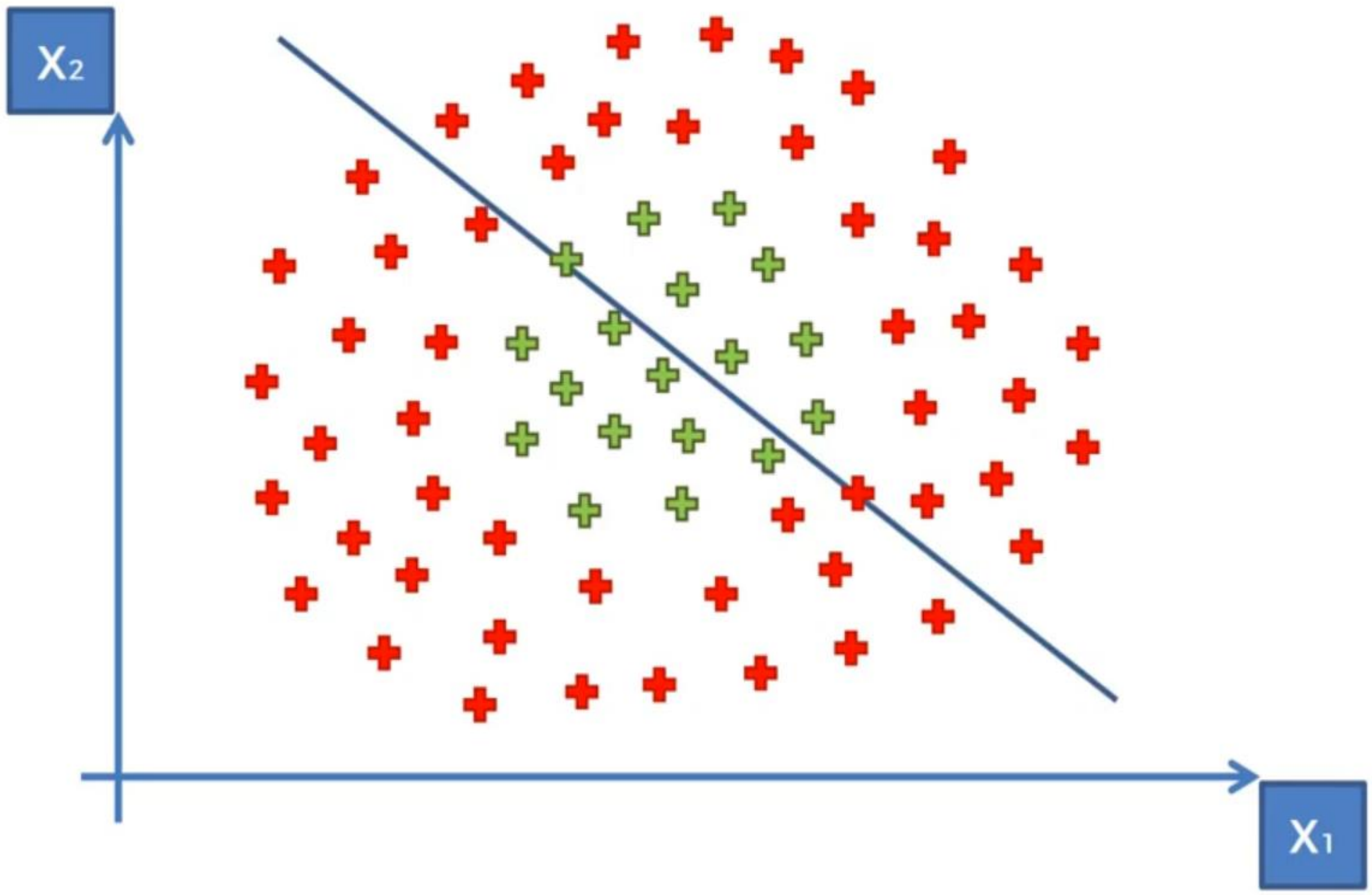
SVM separates well these points



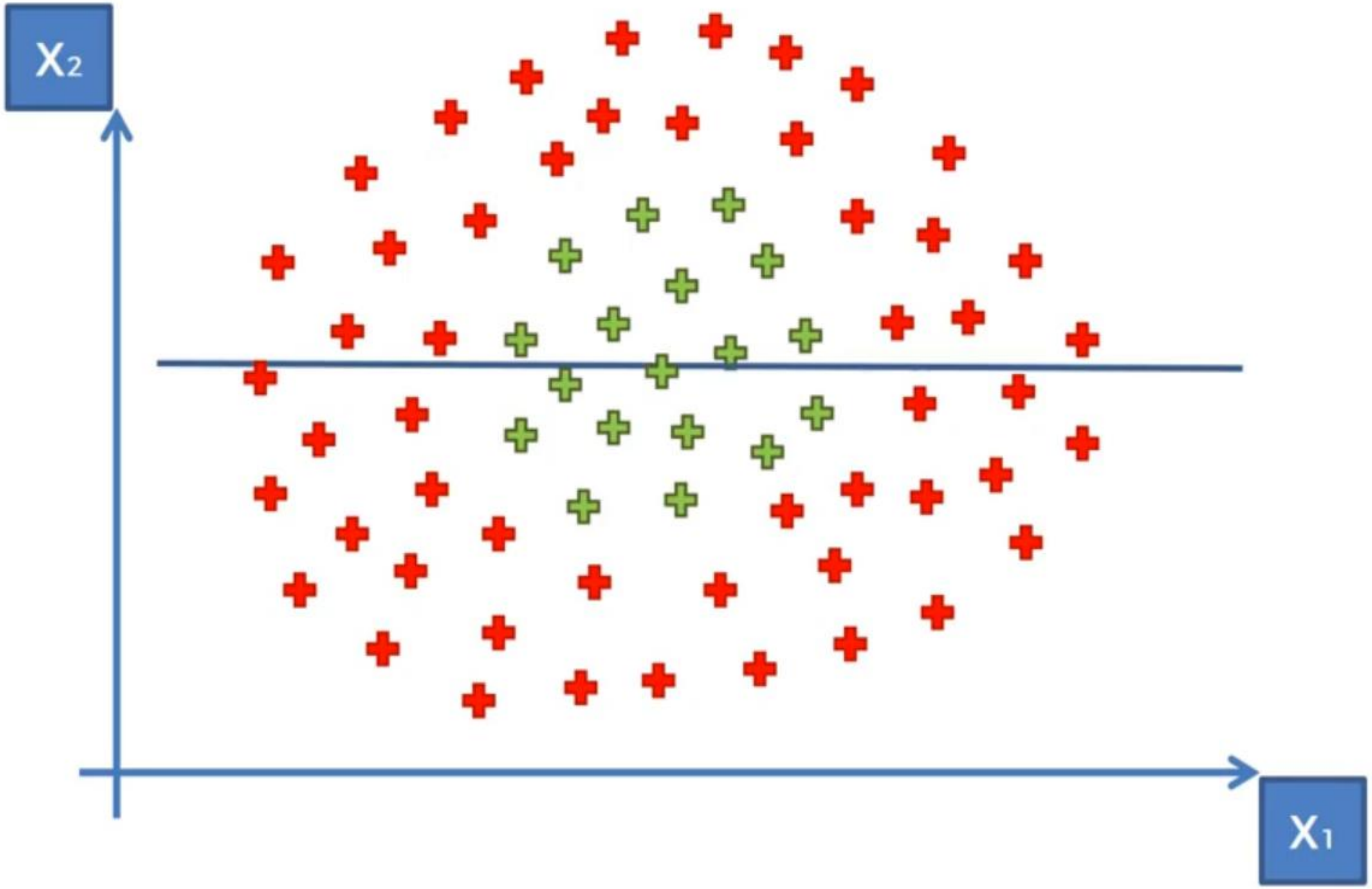
What about these points ?



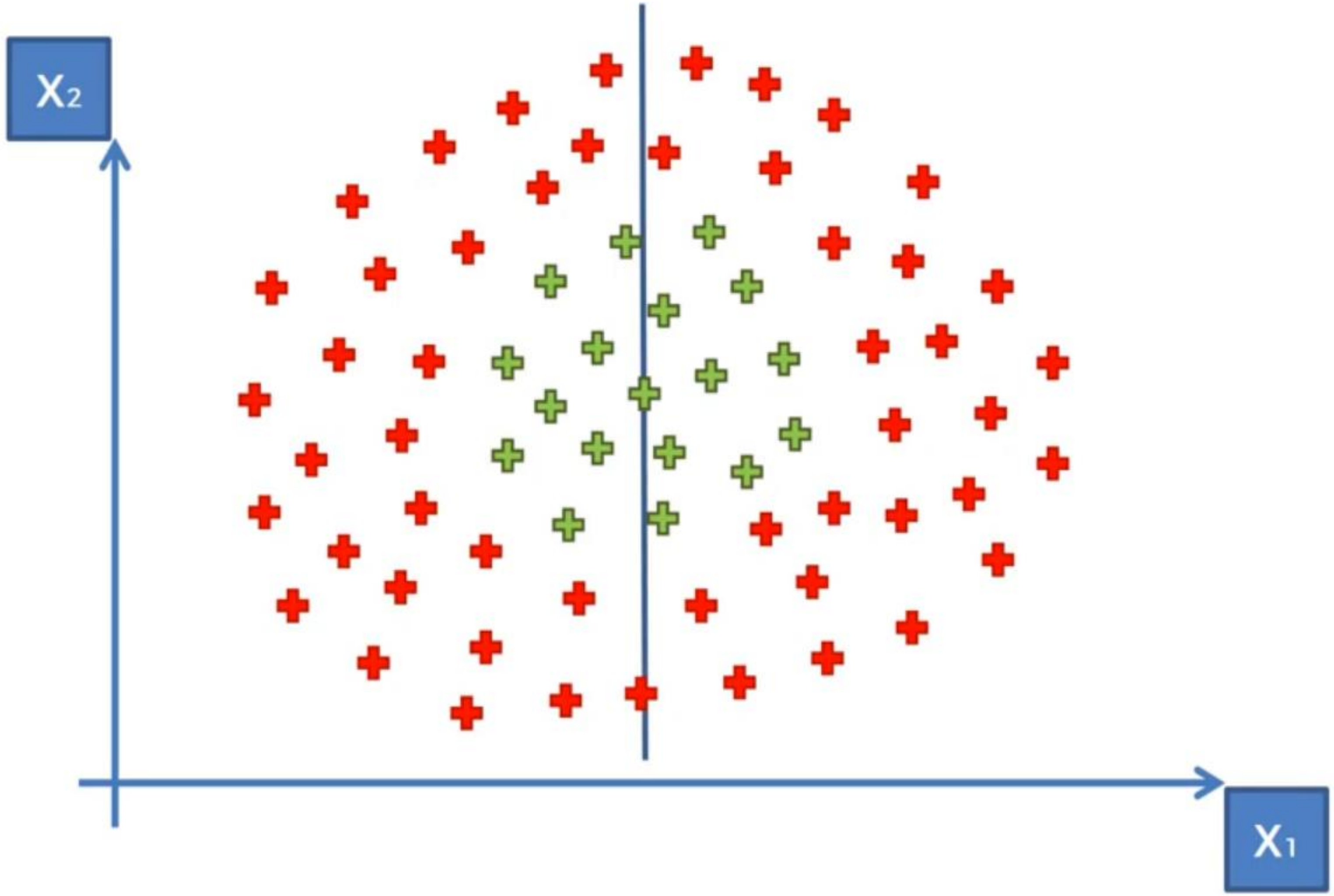
What about these points ?



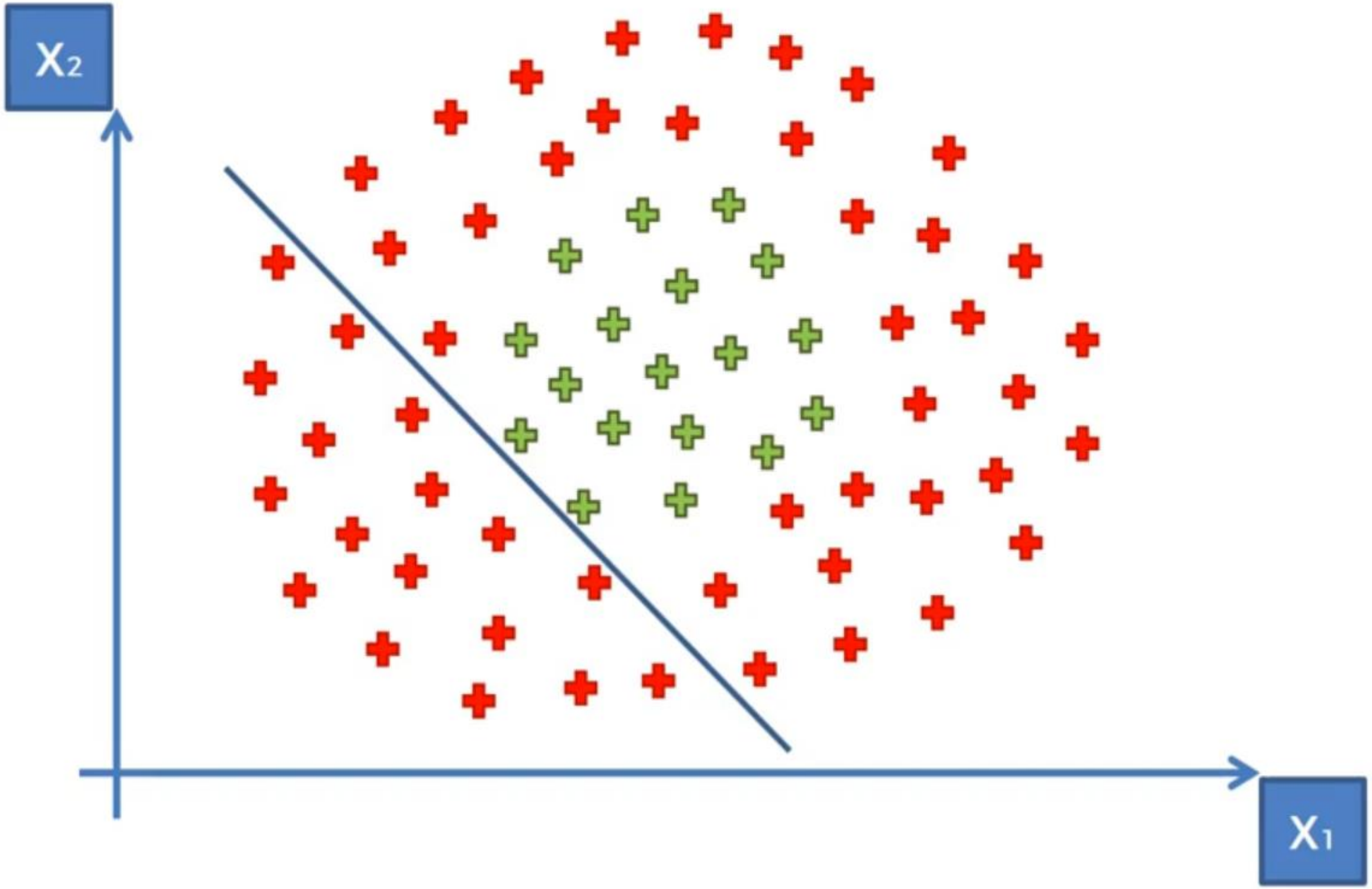
What about these points ?



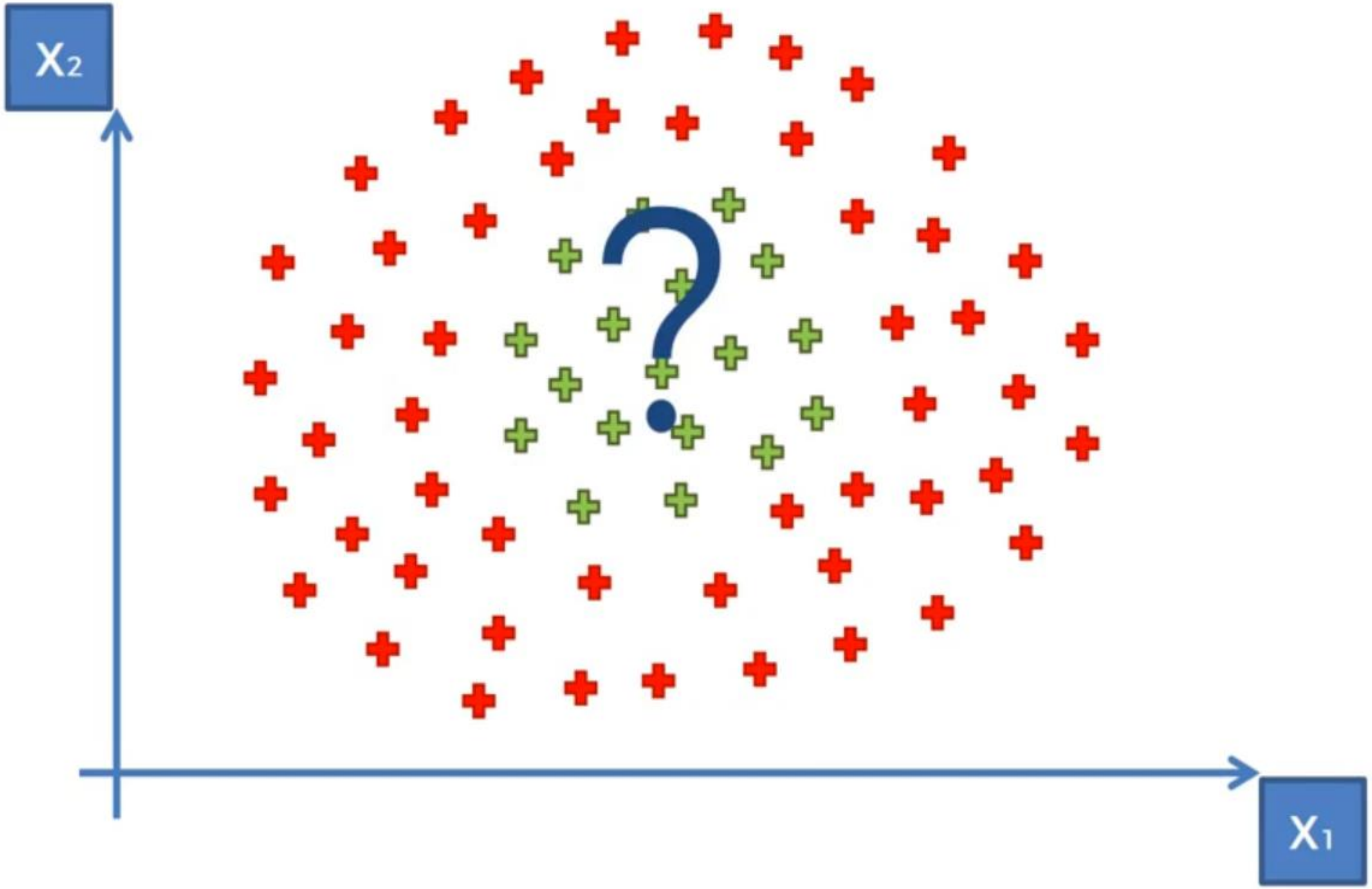
What about these points ?



What about these points ?



What about these points ?

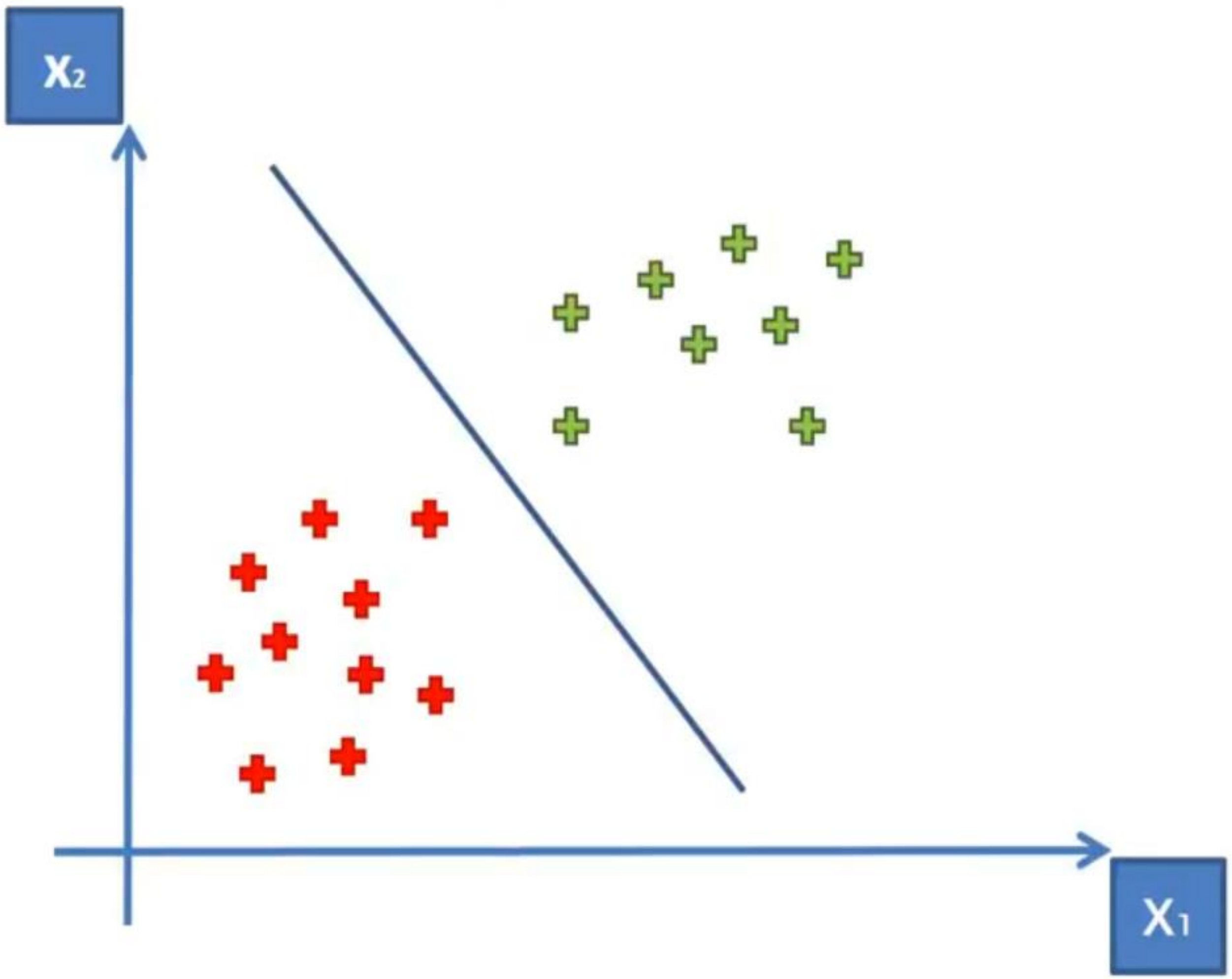


Why ?

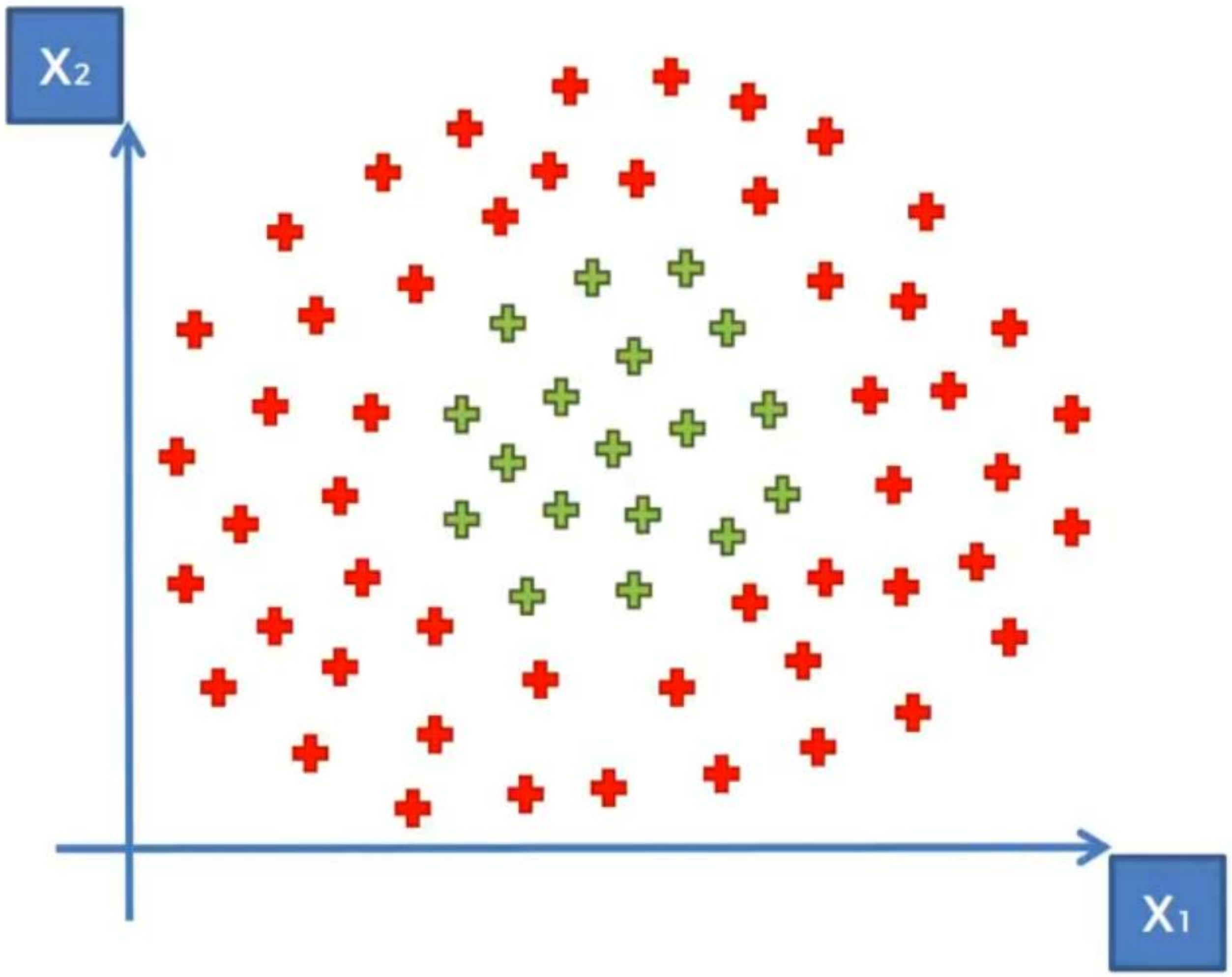
Because the data points are
not LINEARLY SEPARABLE

Linear Separability

Linearly Separable

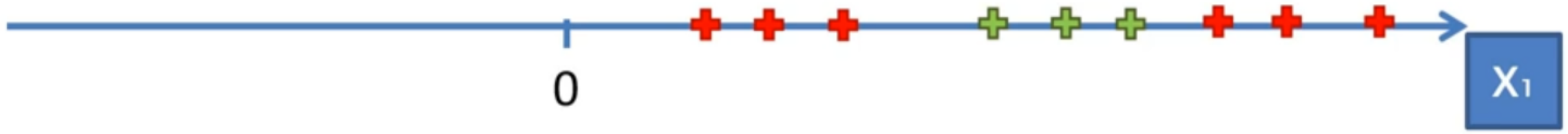


Not Linearly Separable



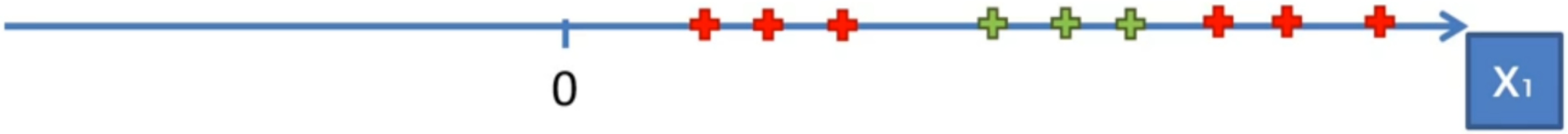
A Higher-Dimensional Space

Mapping to a Higher Dimension



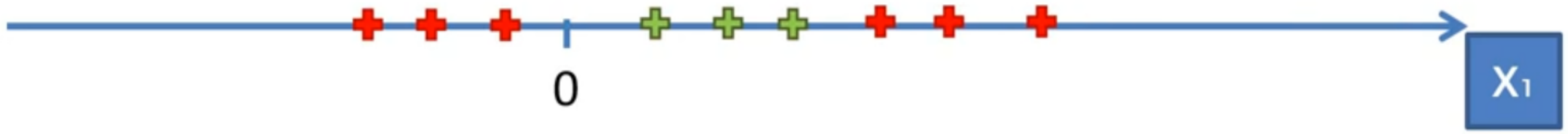
Mapping to a Higher Dimension

$f = x - 5$



Mapping to a Higher Dimension

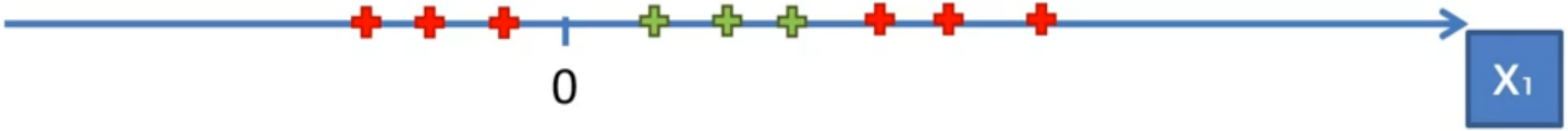
$f = x - 5$



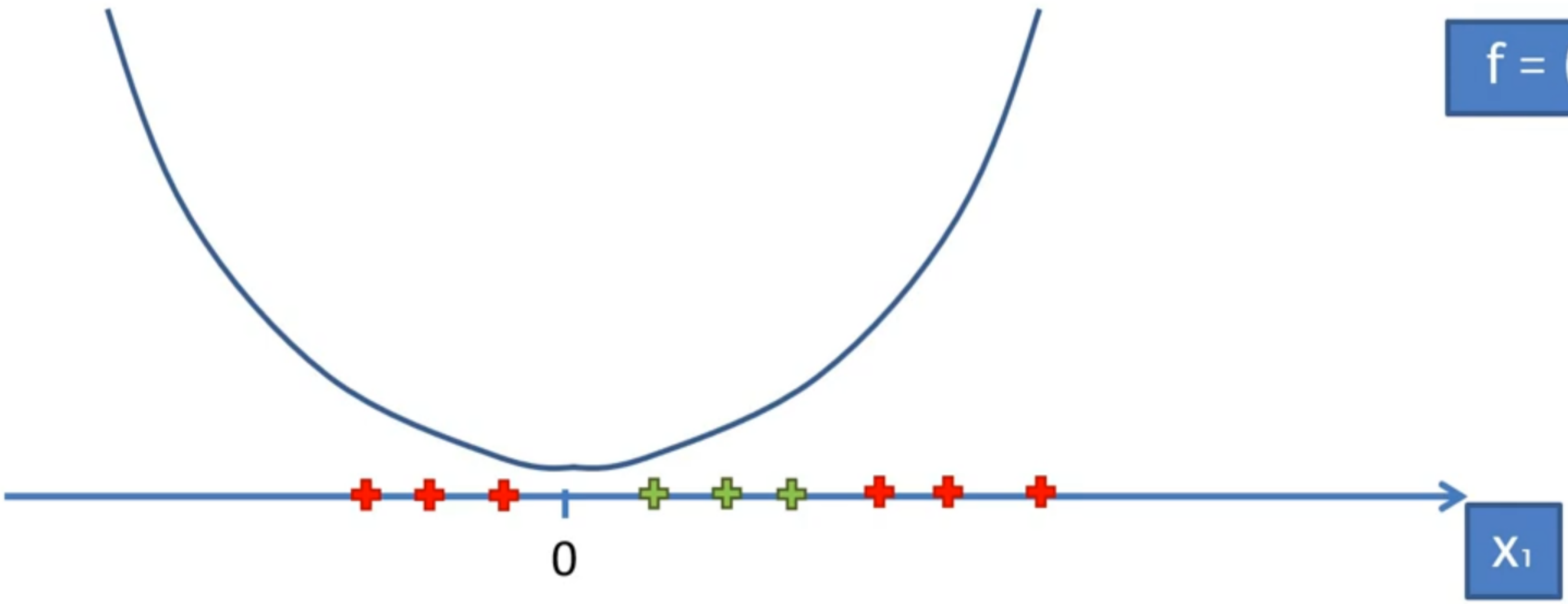
Mapping to a Higher Dimension

$f = x - 5$

$f = (x - 5)^2$



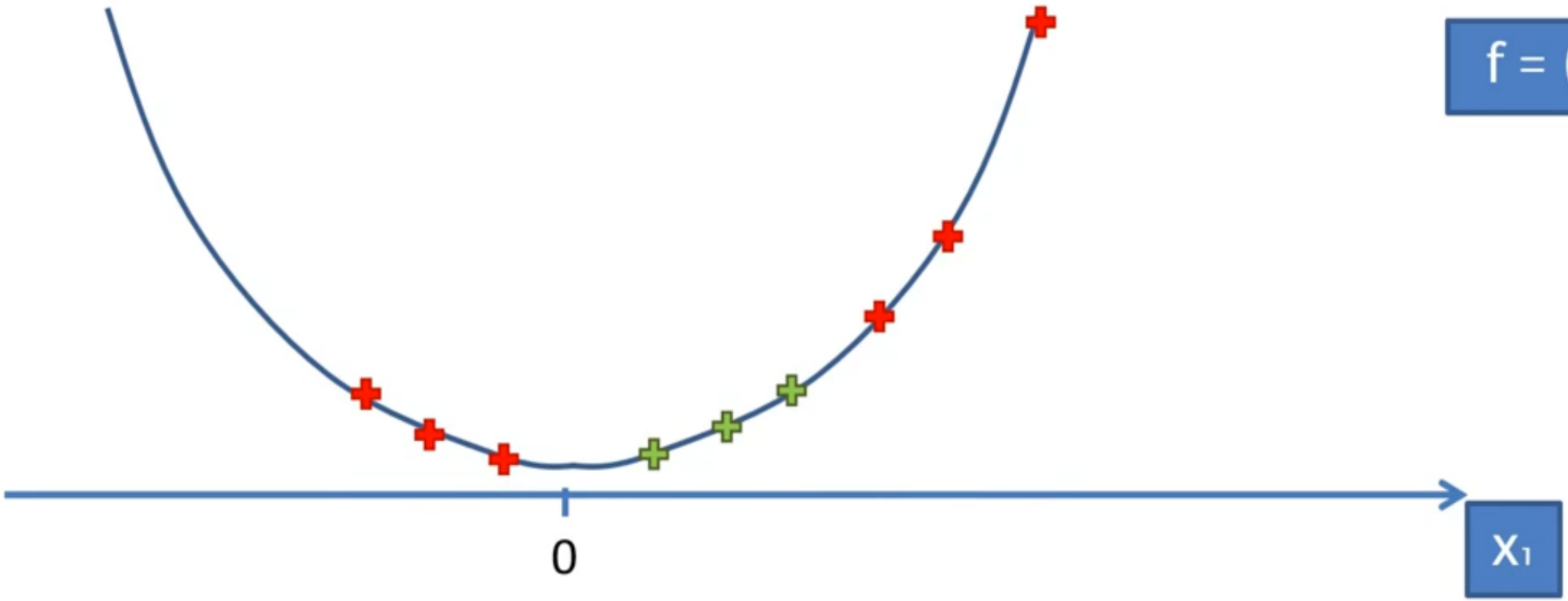
Mapping to a Higher Dimension



$f = x - 5$

$f = (x - 5)^2$

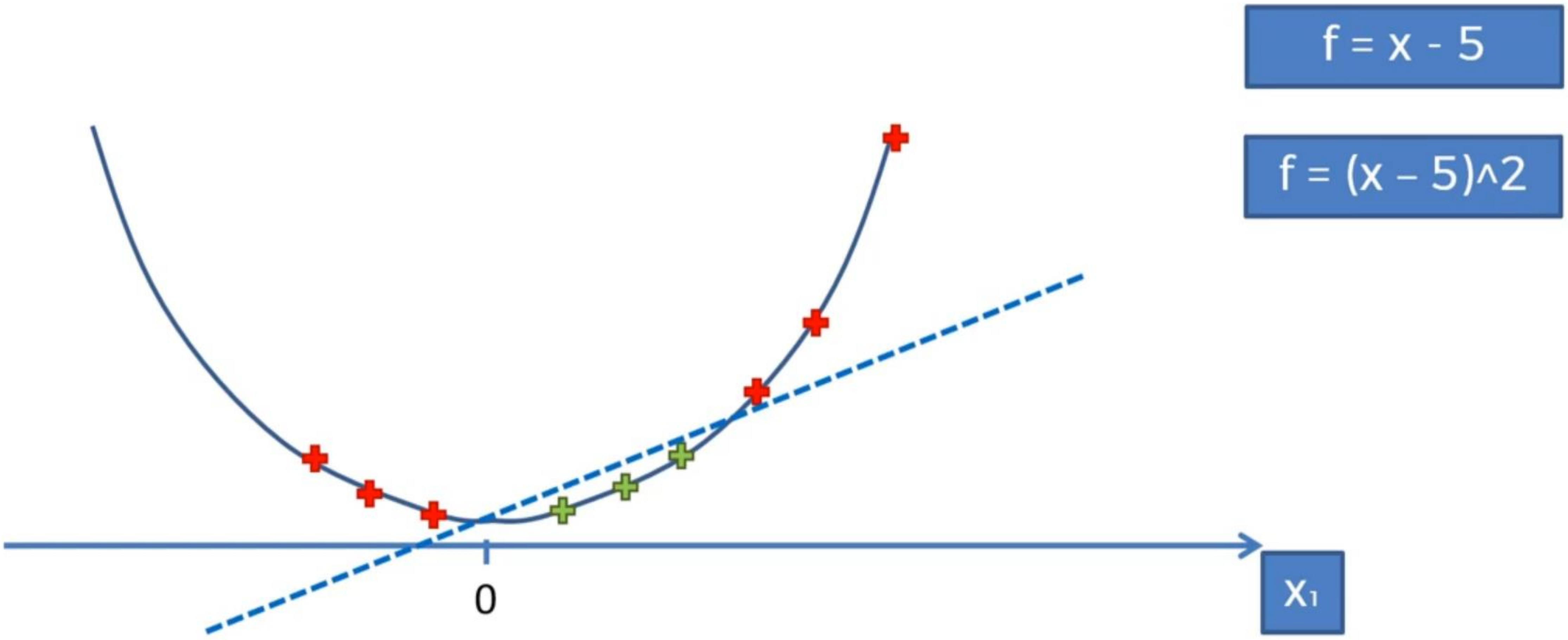
Mapping to a Higher Dimension



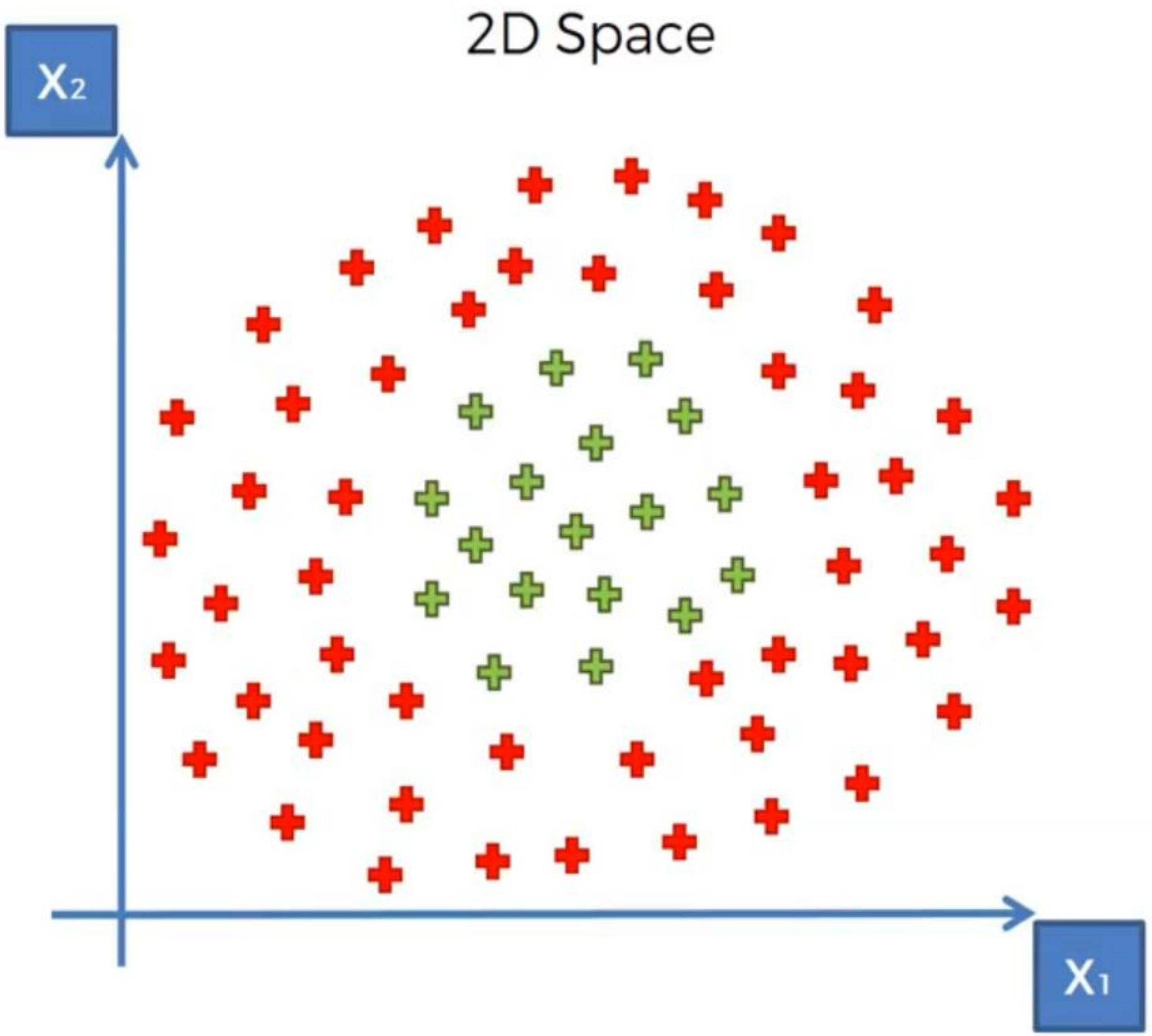
f = x - 5

f = (x – 5)^2

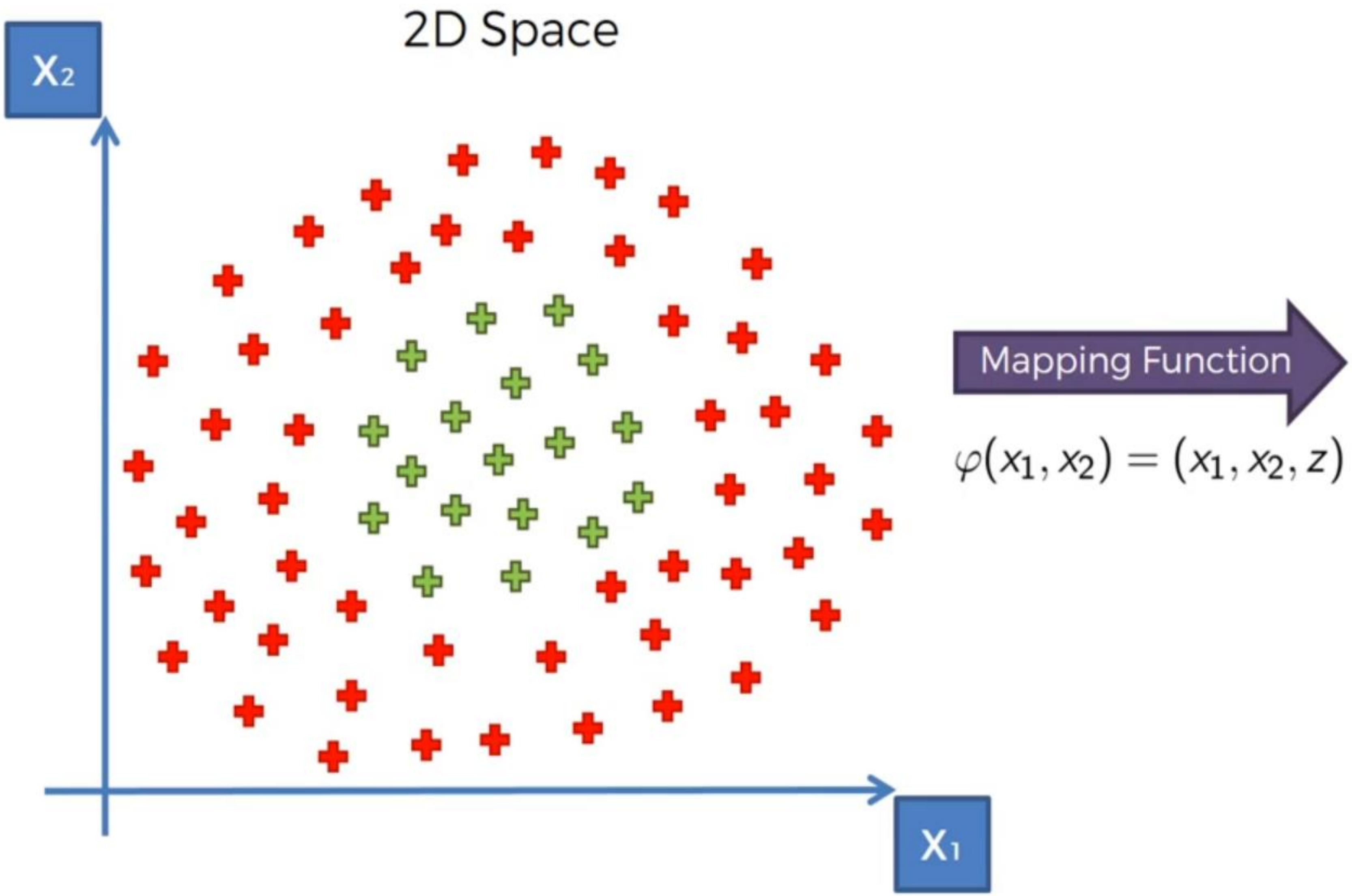
Mapping to a Higher Dimension



Mapping to a Higher Dimension

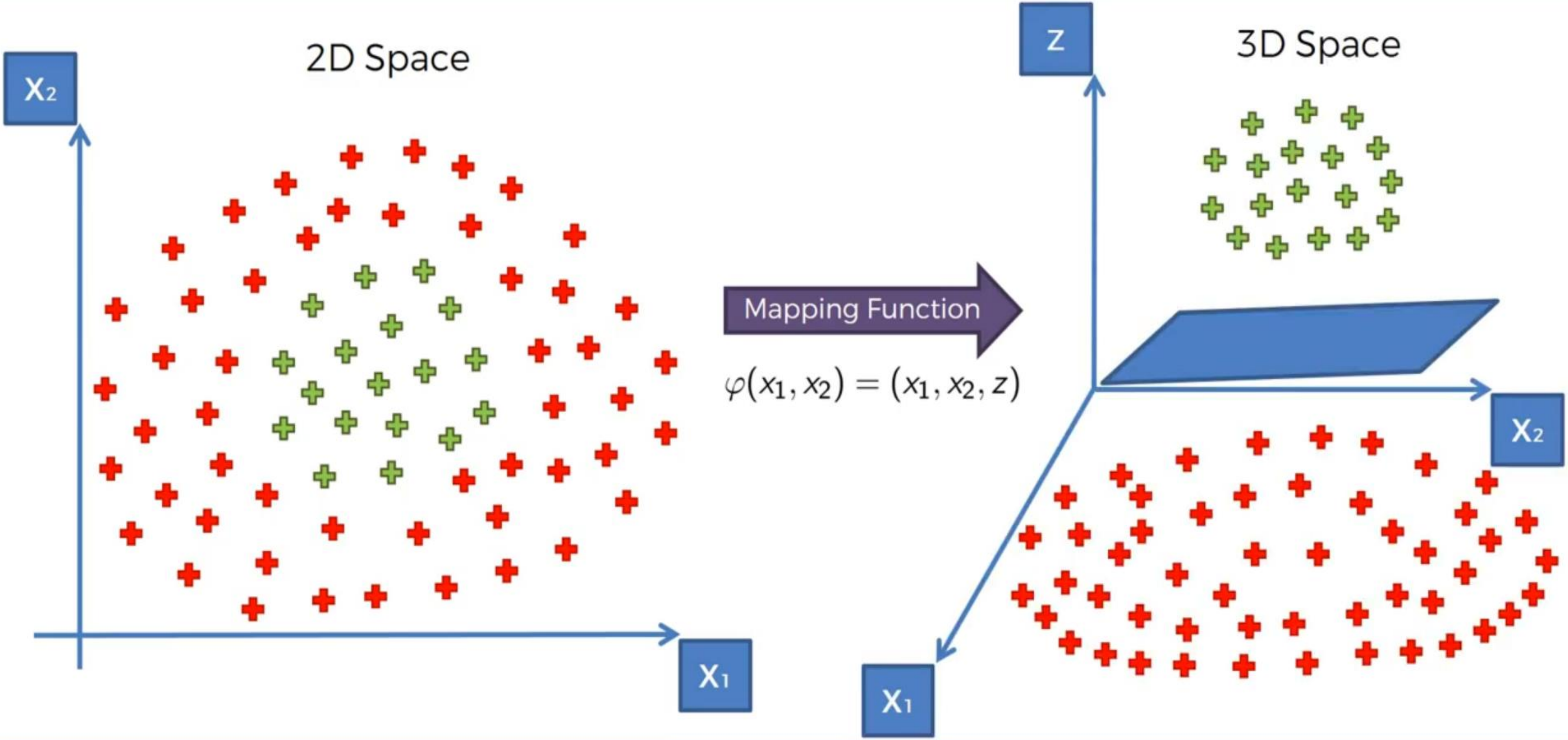


Mapping to a Higher Dimension

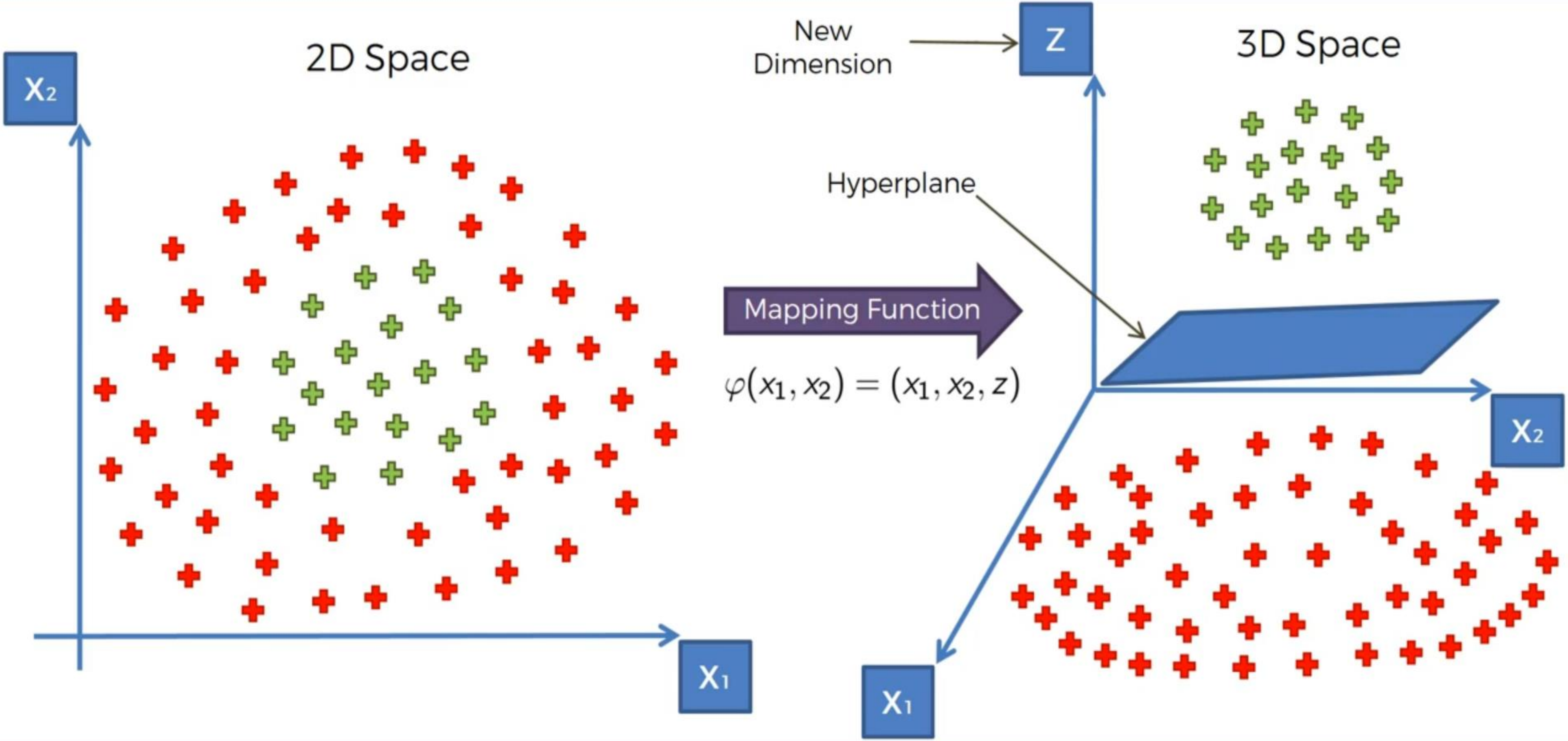


$$\varphi(x_1, x_2) = (x_1, x_2, z)$$

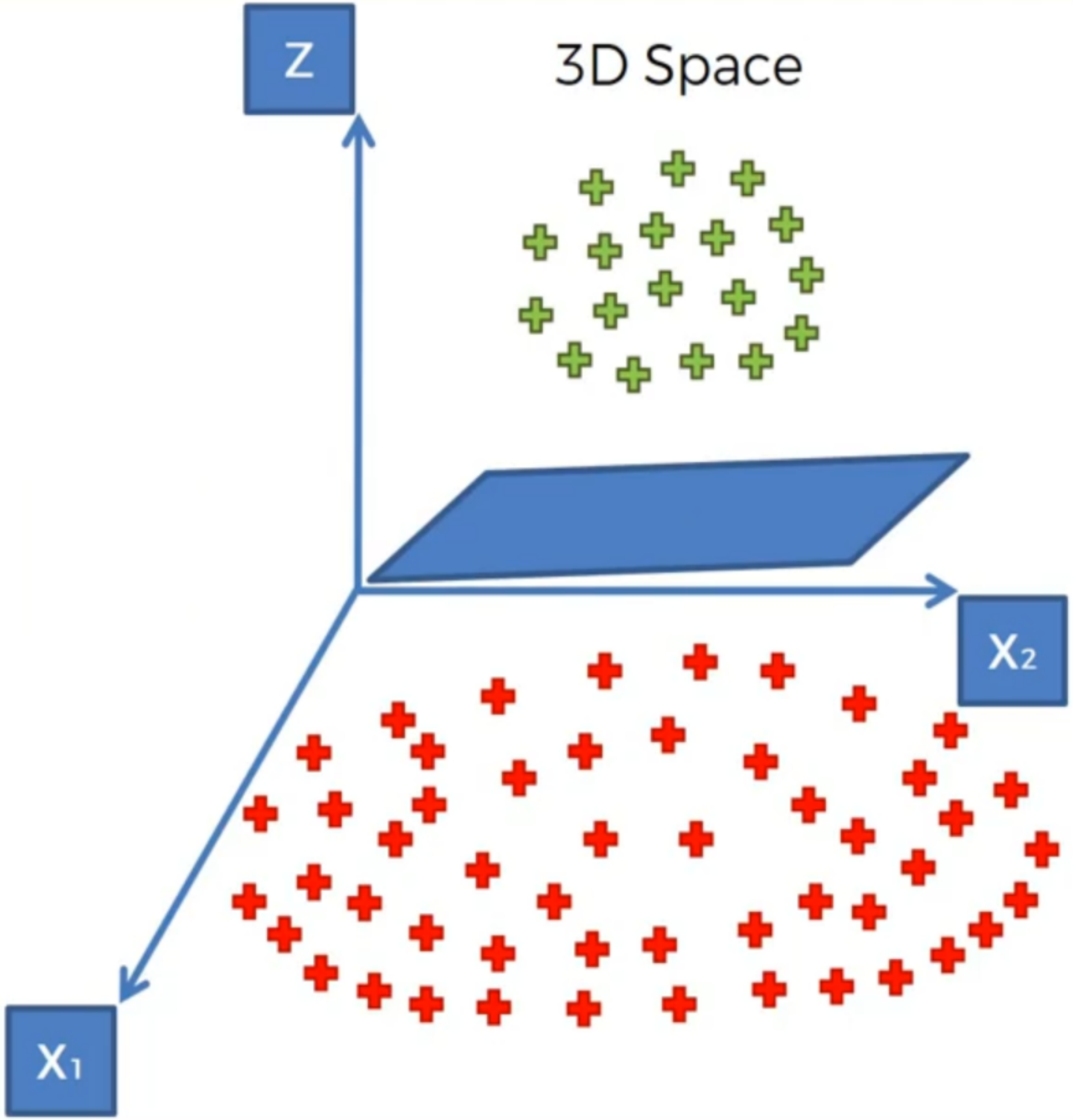
Mapping to a Higher Dimension



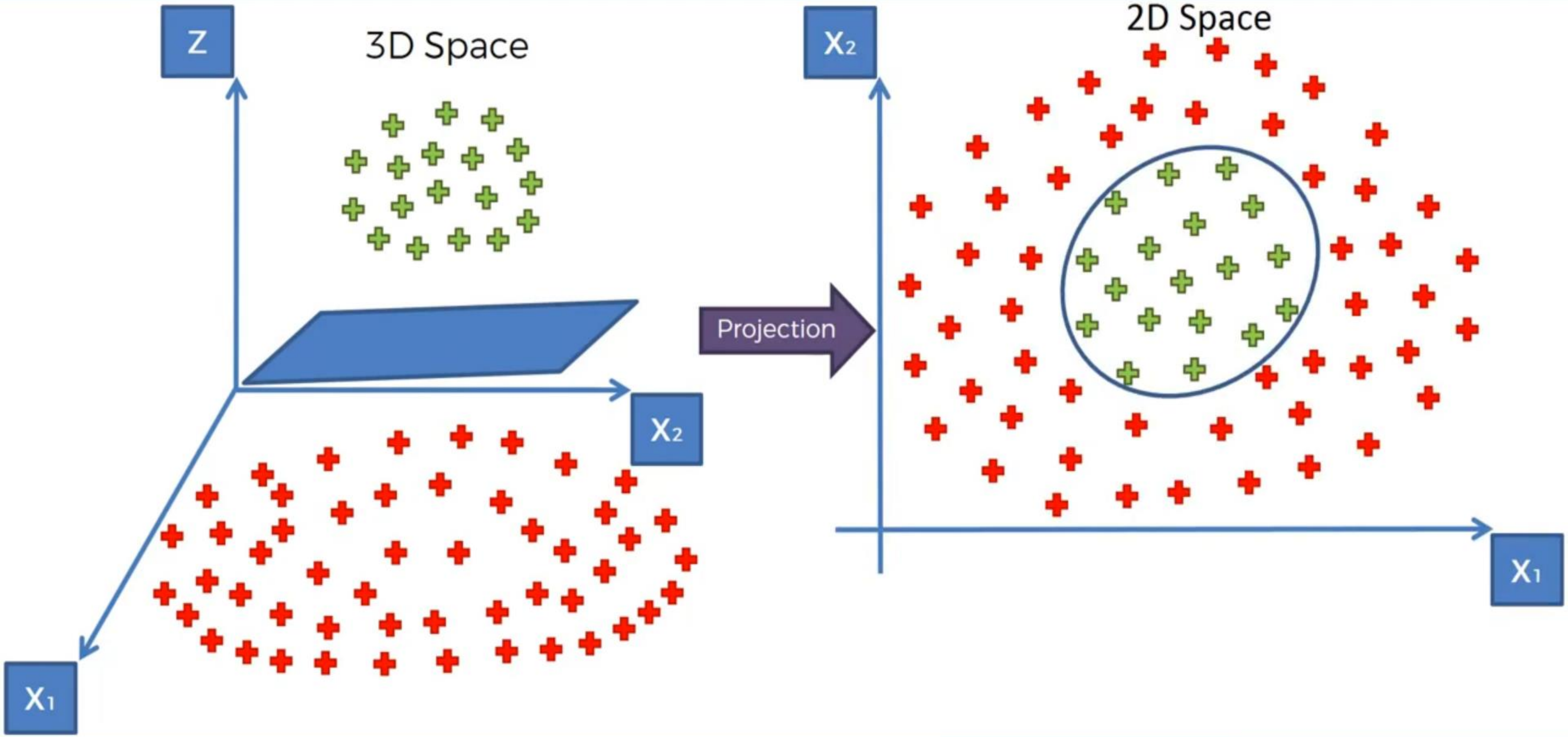
Mapping to a Higher Dimension



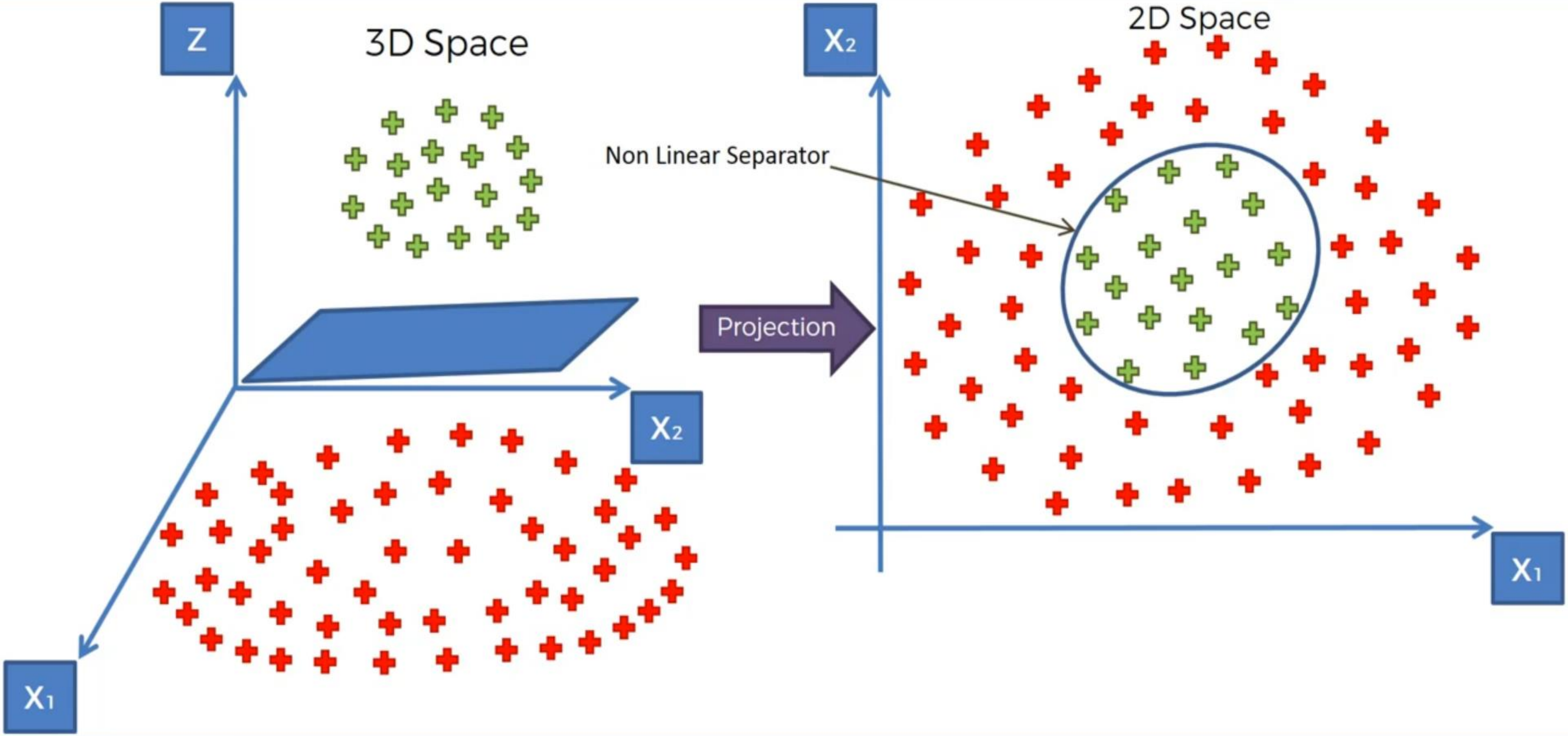
Projecting back to 2D Space



Projecting back to 2D Space



Projecting back to 2D Space



But there is a catch...

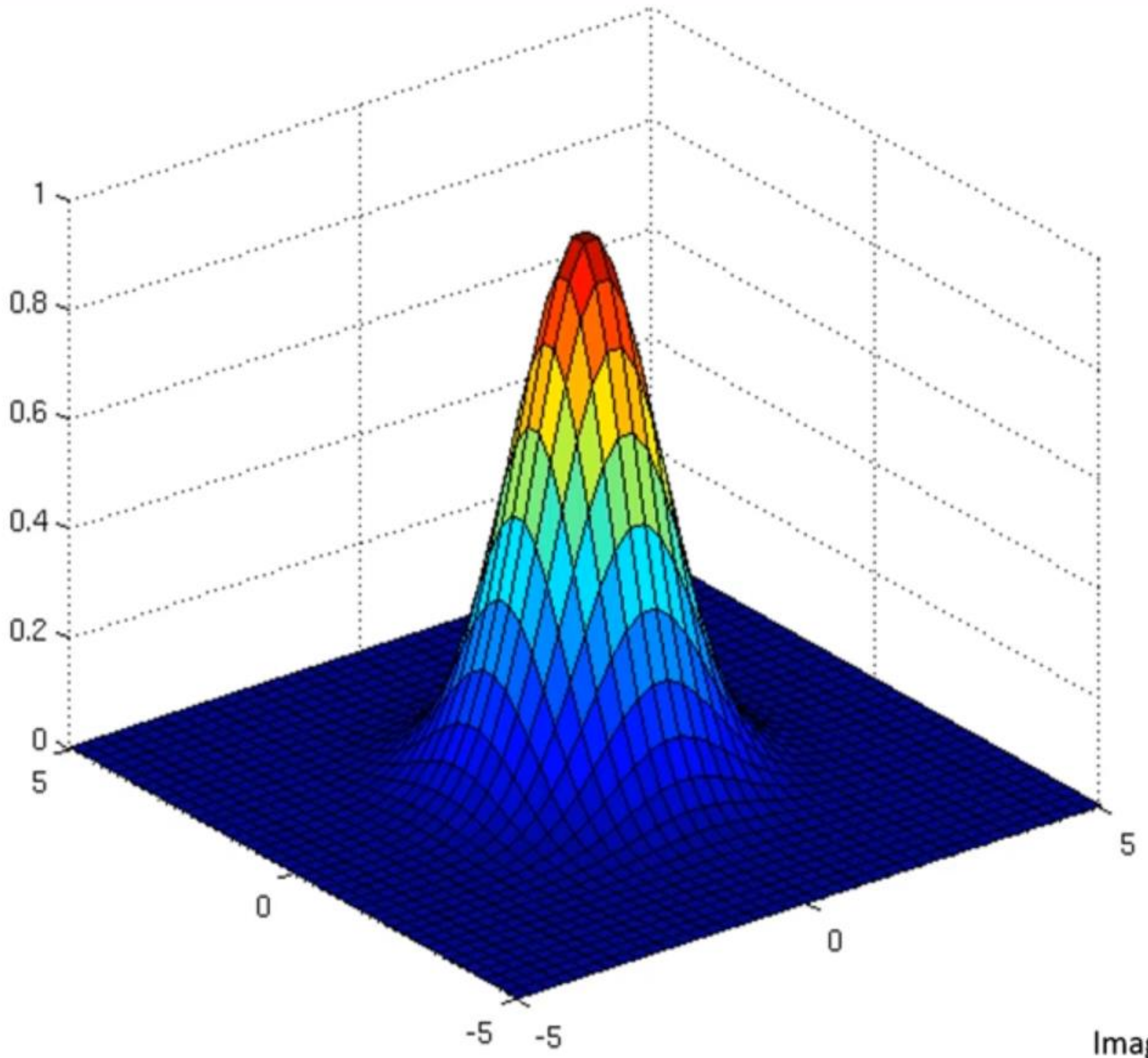
Mapping to a Higher Dimensional Space
can be highly compute-intensive

The Kernel Trick

The Gaussian RBF Kernel

$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x} - \vec{l}^i\|^2}{2\sigma^2}}$$

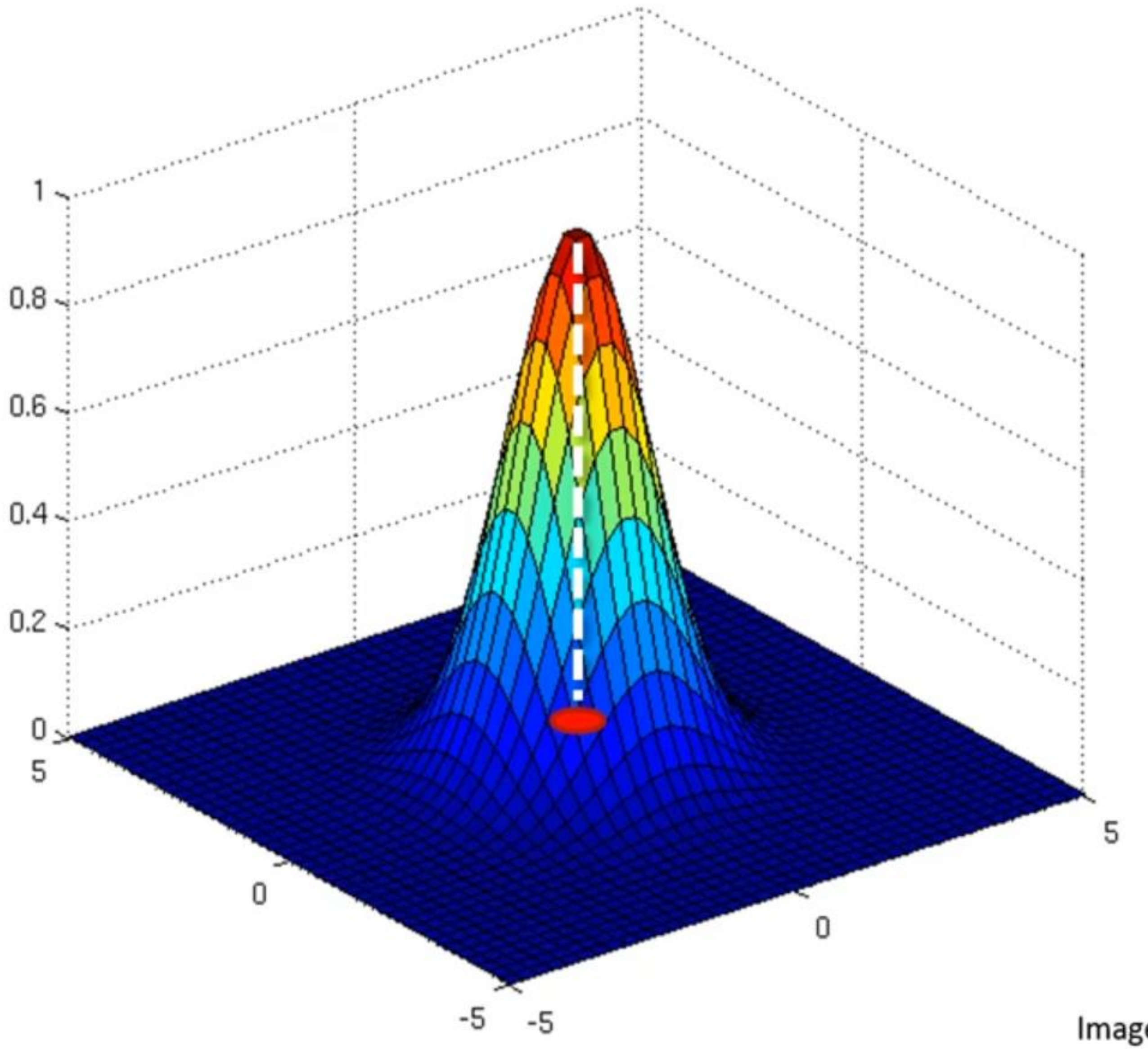
The Gaussian RBF Kernel



$$K\left(\vec{x}, \vec{l}^{i}\right) = e^{-\frac{\left\|\vec{x}-\vec{l}^{i}\right\|^{2}}{2\sigma^{2}}}$$

Image source: http://www.cs.toronto.edu/~duvenaud/cookbook/index.html

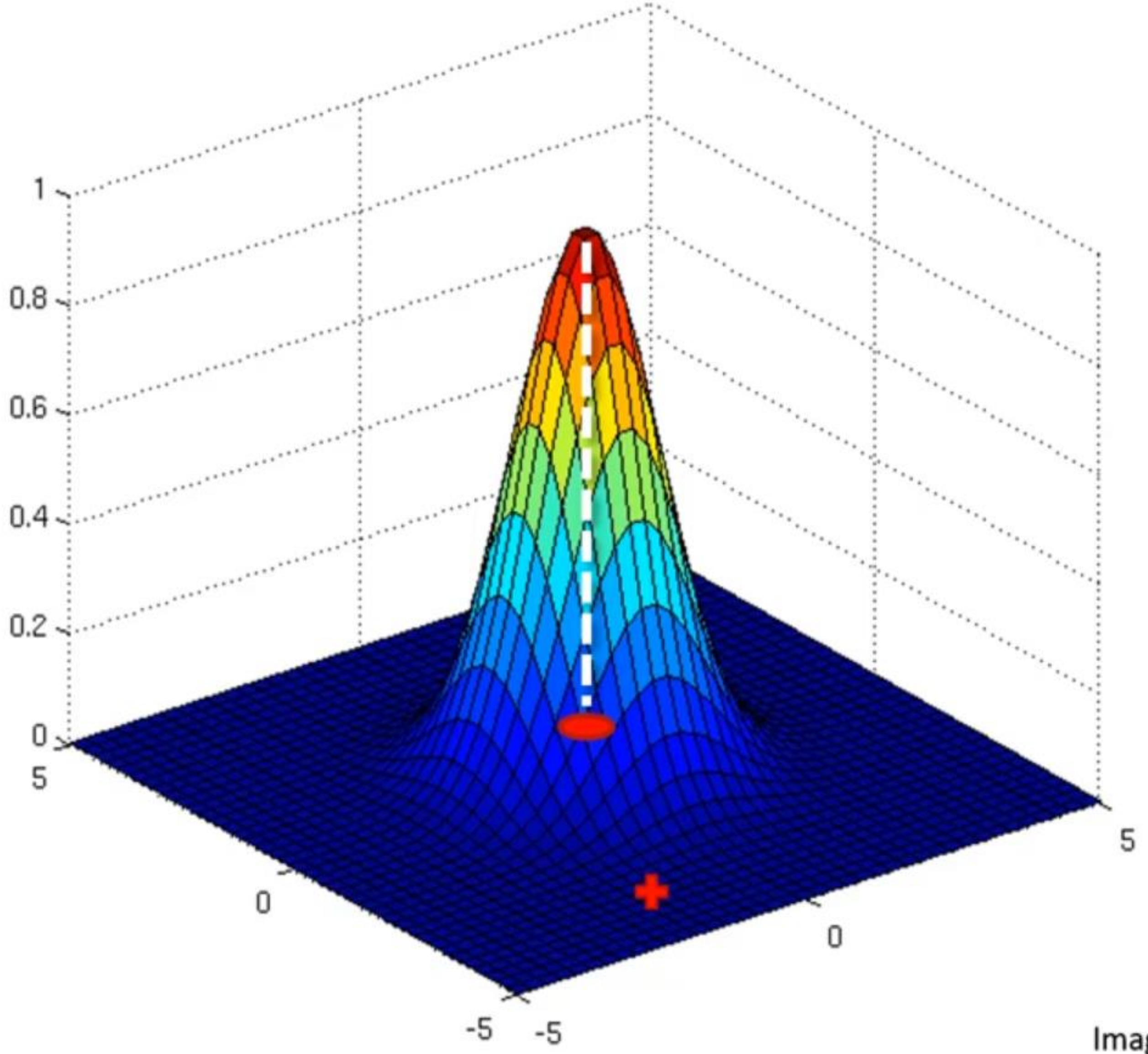
The Gaussian RBF Kernel



$$K\left(\vec{x}, \vec{l}^{i}\right) = e^{-\frac{\left\|\vec{x}-\vec{l}^{i}\right\|^{2}}{2\sigma^{2}}}$$

Image source: http://www.cs.toronto.edu/~duvenaud/cookbook/index.html

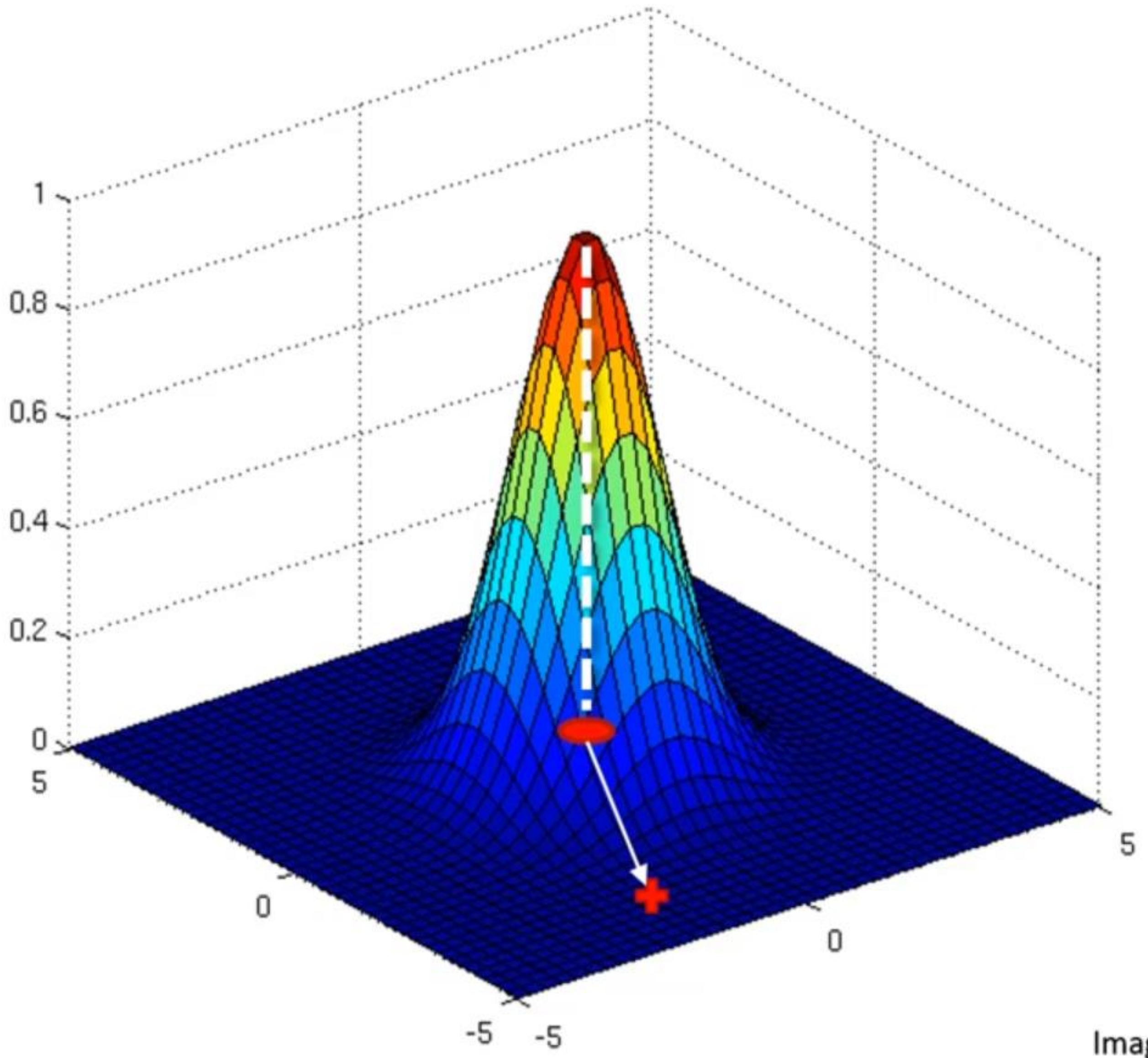
The Gaussian RBF Kernel



$$K\left(\vec{x}, \vec{l}^{i}\right) = e^{-\frac{\left\|\vec{x}-\vec{l}^{i}\right\|^{2}}{2\sigma^{2}}}$$

Image source: http://www.cs.toronto.edu/~duvenaud/cookbook/index.html

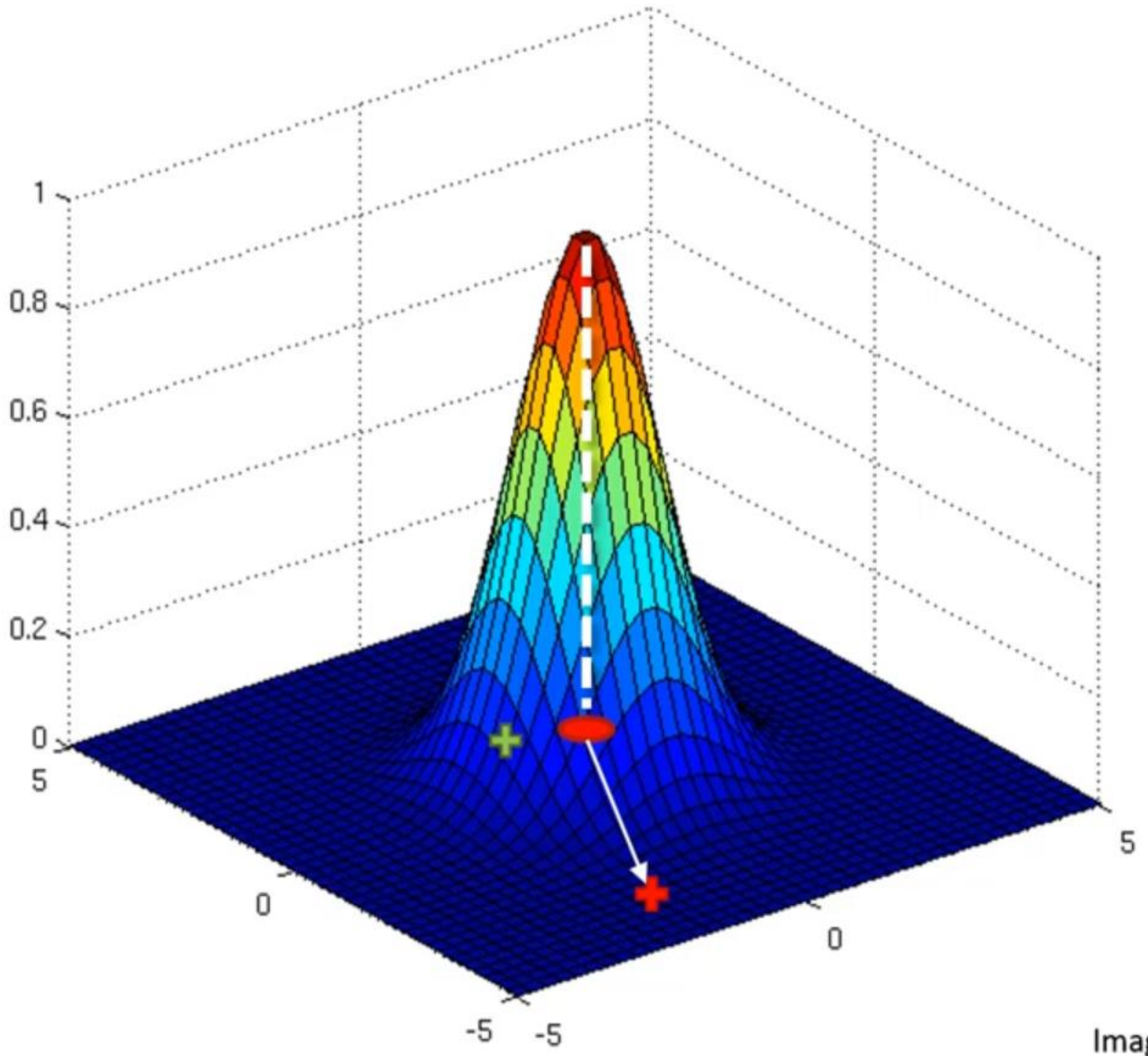
The Gaussian RBF Kernel



$$K\left(\vec{x}, \vec{l}^{i}\right) = e^{-\frac{\left\|\vec{x}-\vec{l}^{i}\right\|^{2}}{2\sigma^{2}}}$$

Image source: http://www.cs.toronto.edu/~duvenaud/cookbook/index.html

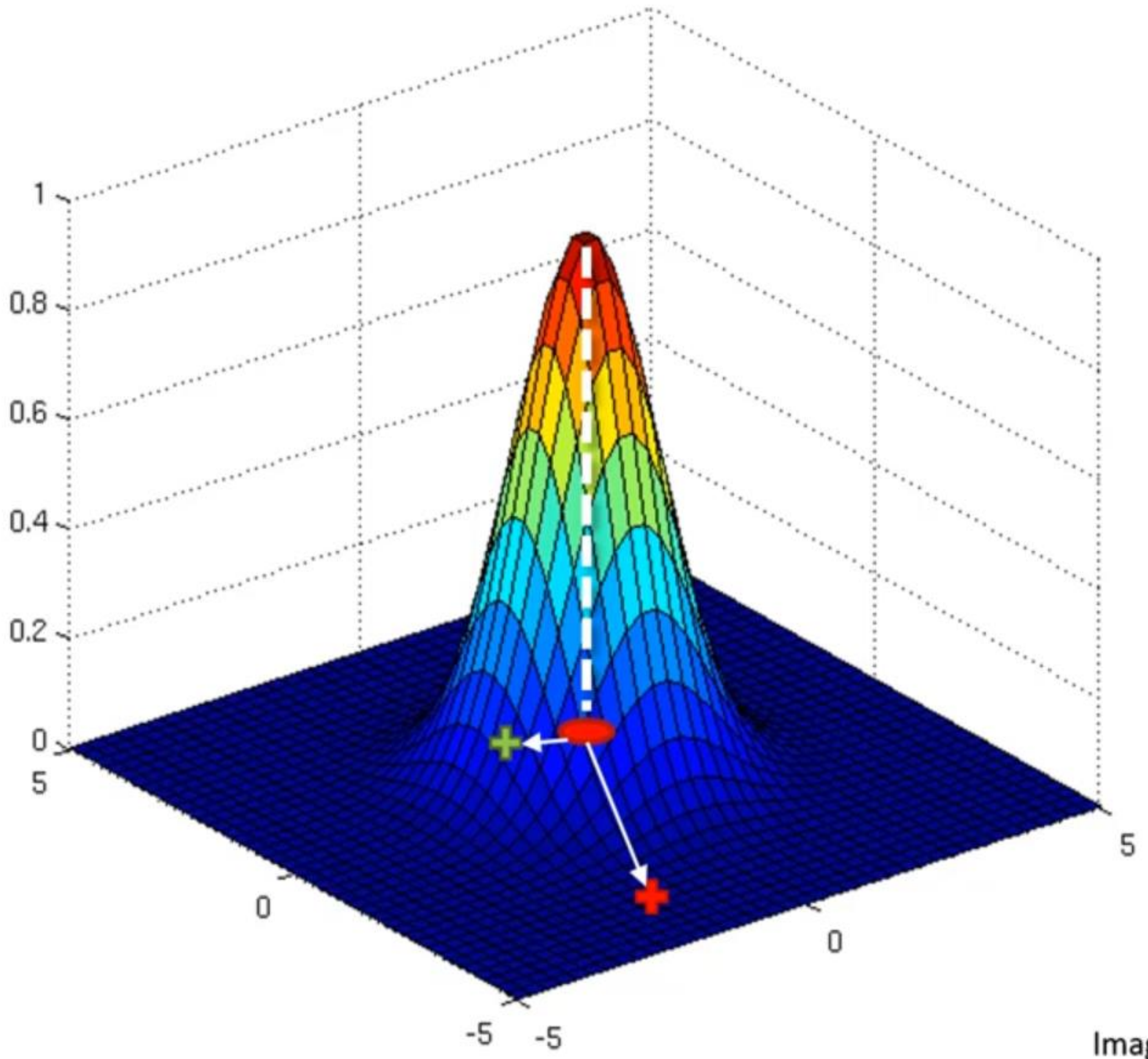
The Gaussian RBF Kernel



$$K(\vec{x},\vec{l}^i) = e^{-\frac{\left\|\vec{x}-\vec{l}^i\right\|^2}{2\sigma^2}}$$

Image source: http://www.cs.toronto.edu/~duvenaud/cookbook/index.html

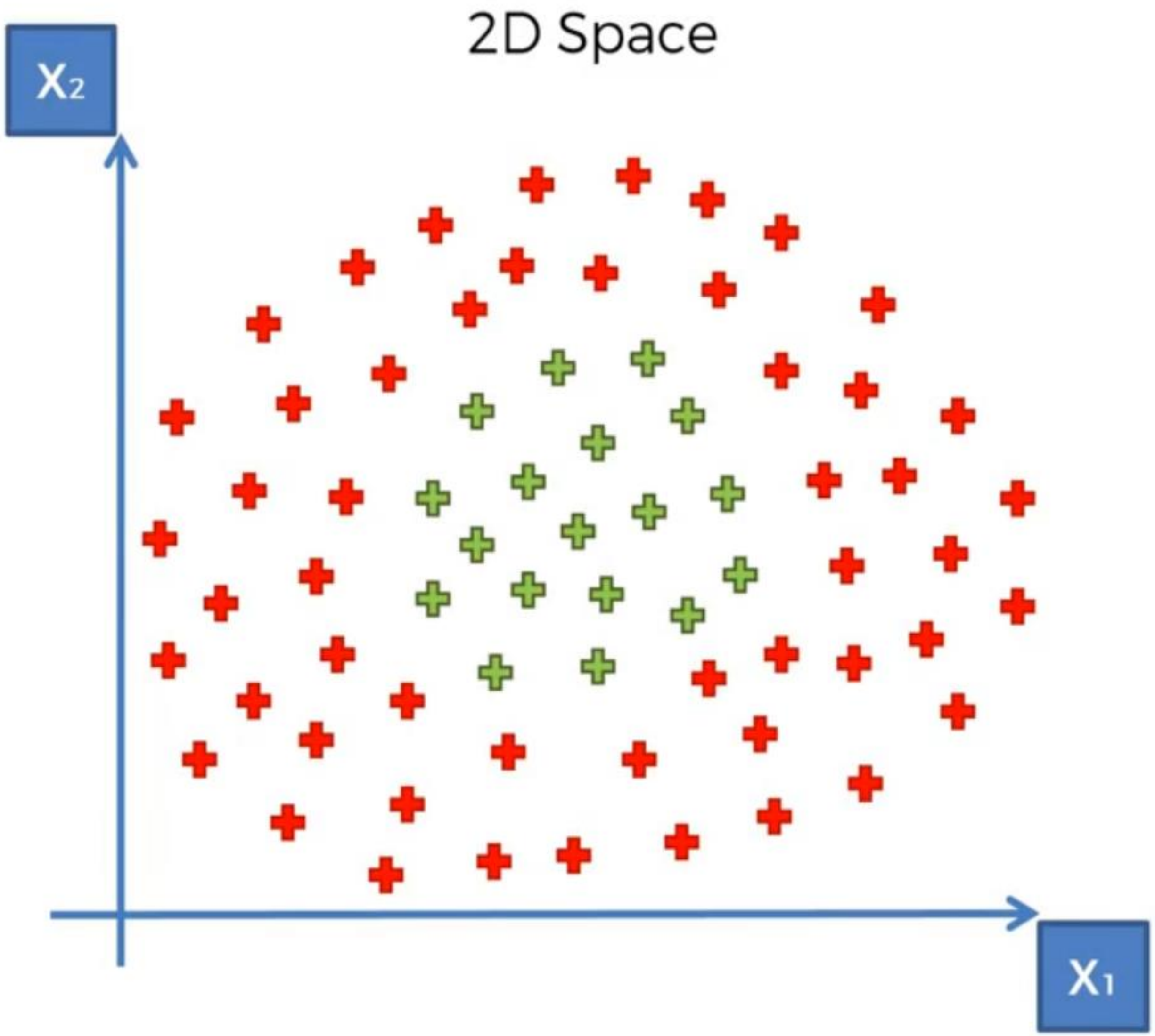
The Gaussian RBF Kernel

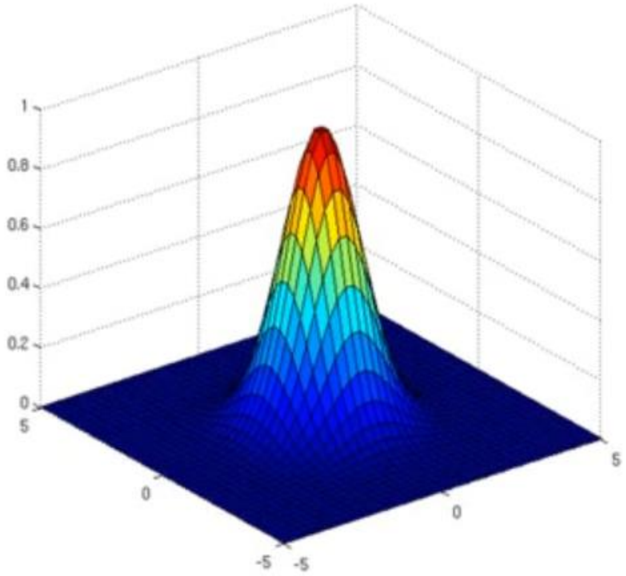


$$K\left(\vec{x}, \vec{l}^i\right) = e^{-\frac{\left\|\vec{x}-\vec{l}^i\right\|^2}{2\sigma^2}}$$

Image source: http://www.cs.toronto.edu/~duvenaud/cookbook/index.html

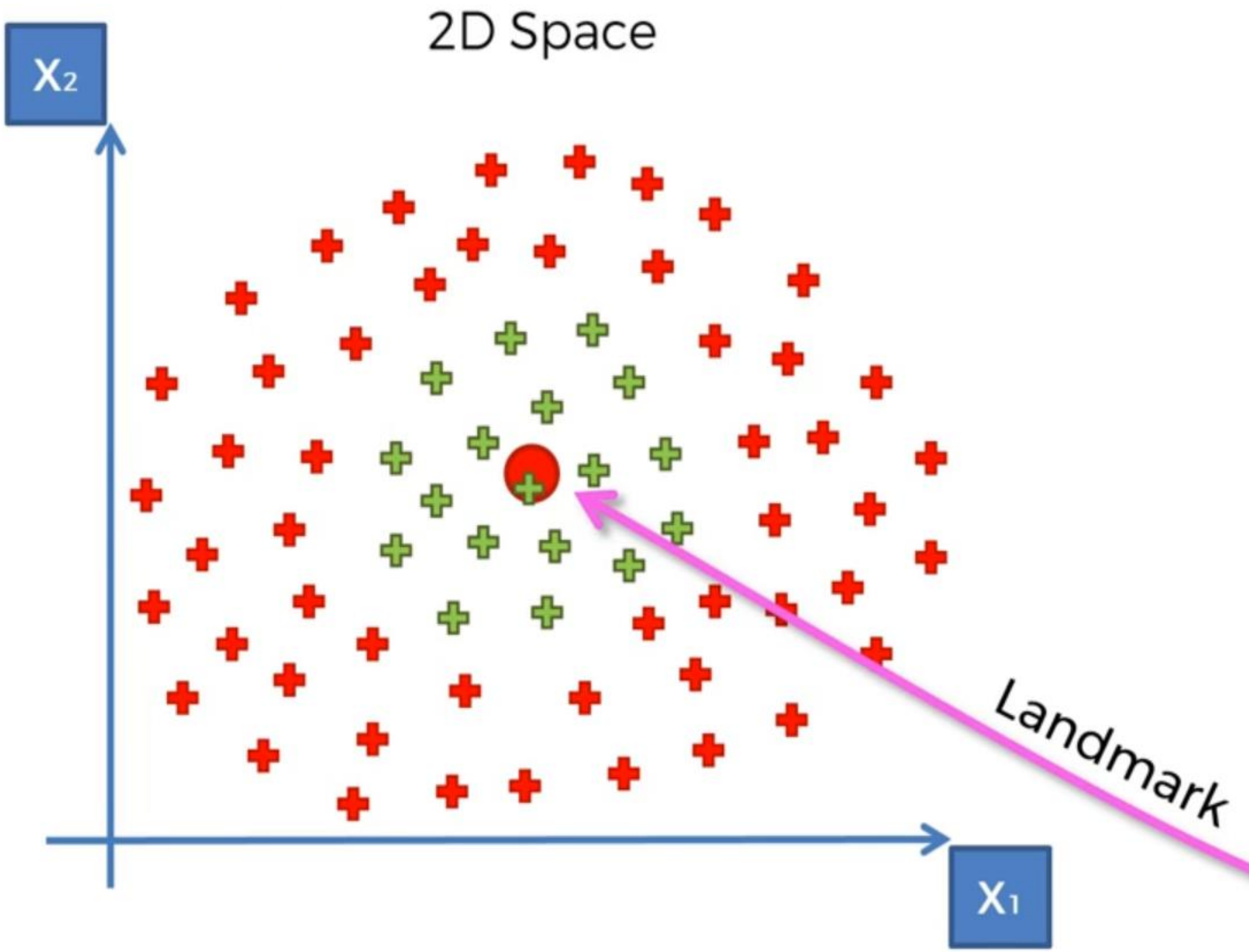
The Gaussian RBF Kernel

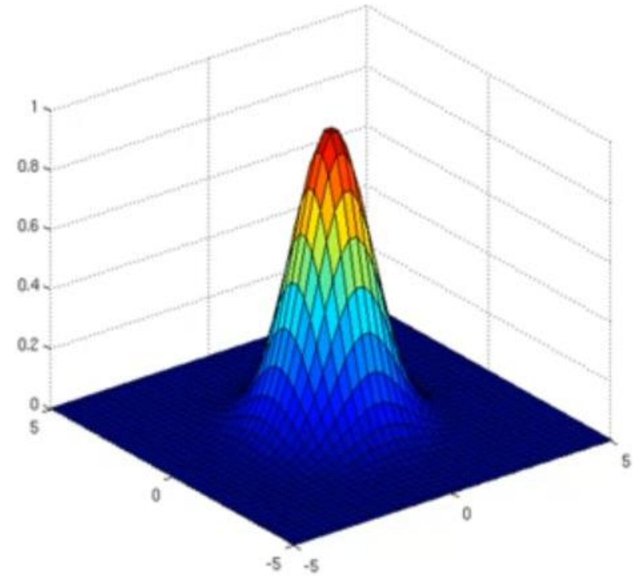




$$K\left(\vec{x},\vec{l}^{i}\right) = e^{-\frac{\left\|\vec{x}-\vec{l}^{i}\right\|^{2}}{2\sigma^{2}}}$$

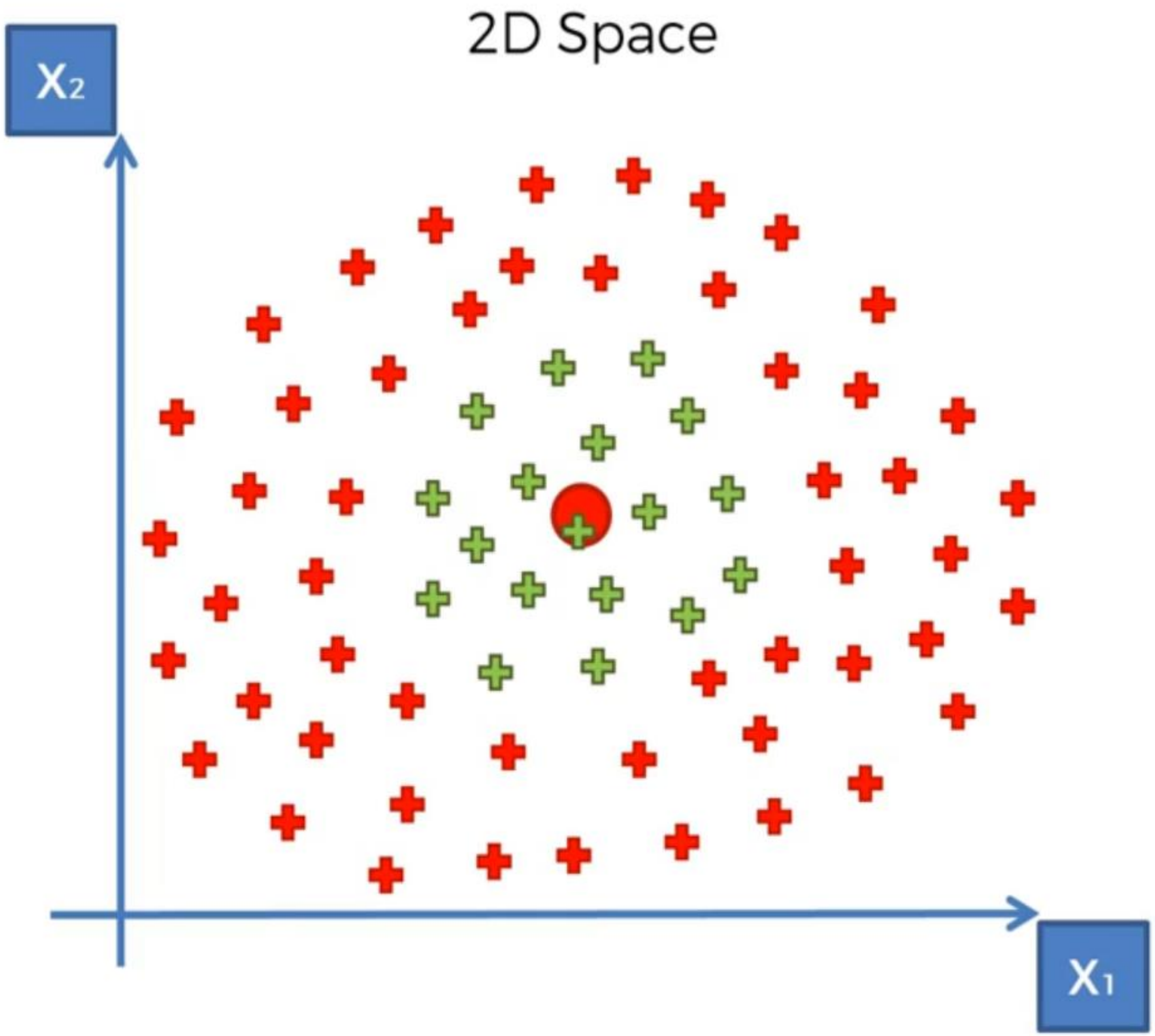
The Gaussian RBF Kernel

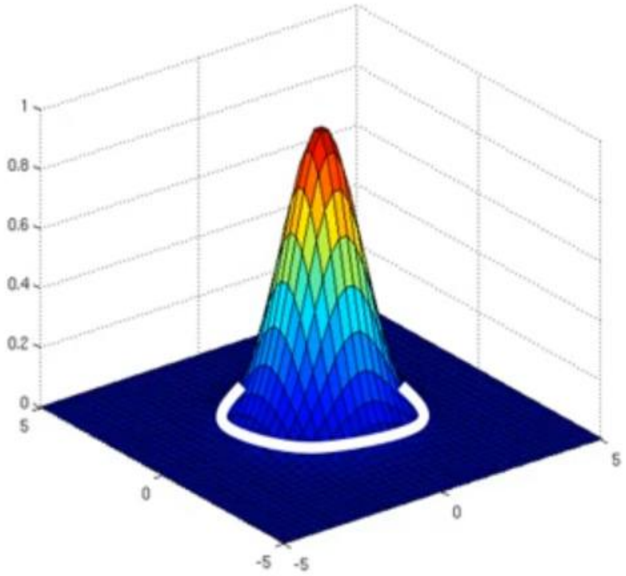




$$K(\vec{x}, \vec{l}^{i}) = e^{-\frac{\|\vec{x}-\vec{l}^{i}\|^{2}}{2\sigma^{2}}}$$

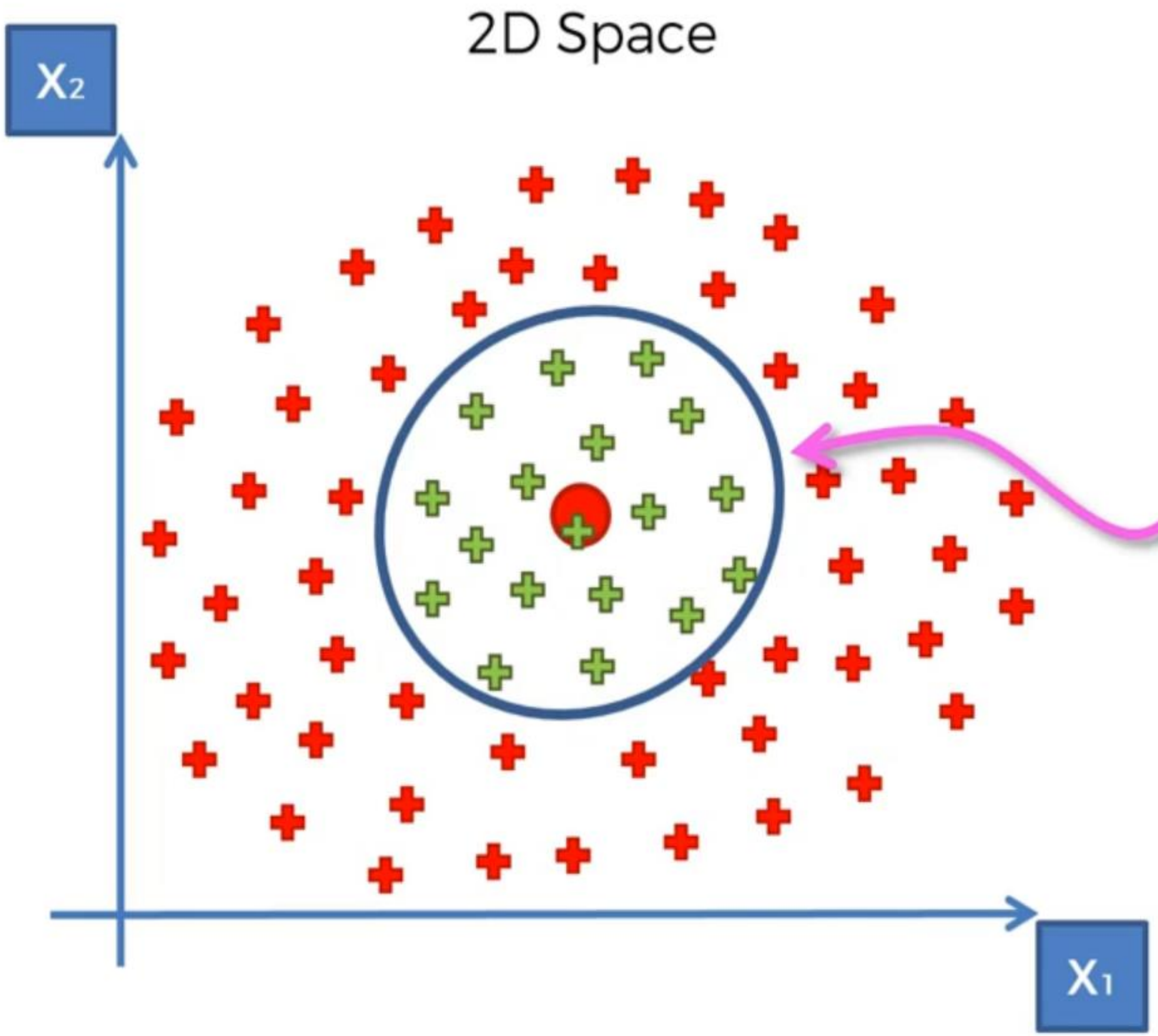
The Gaussian RBF Kernel

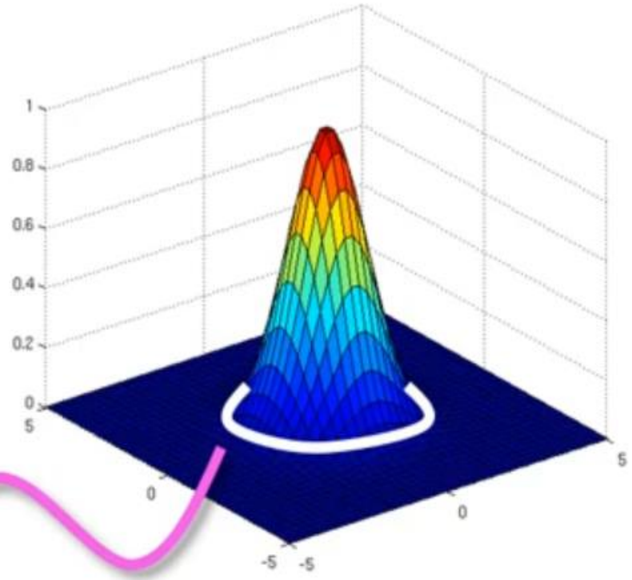




$$K\left(\vec{x}, \vec{l}^{i}\right) = e^{-\frac{\left\|\vec{x}-\vec{l}^{i}\right\|^{2}}{2\sigma^{2}}}$$

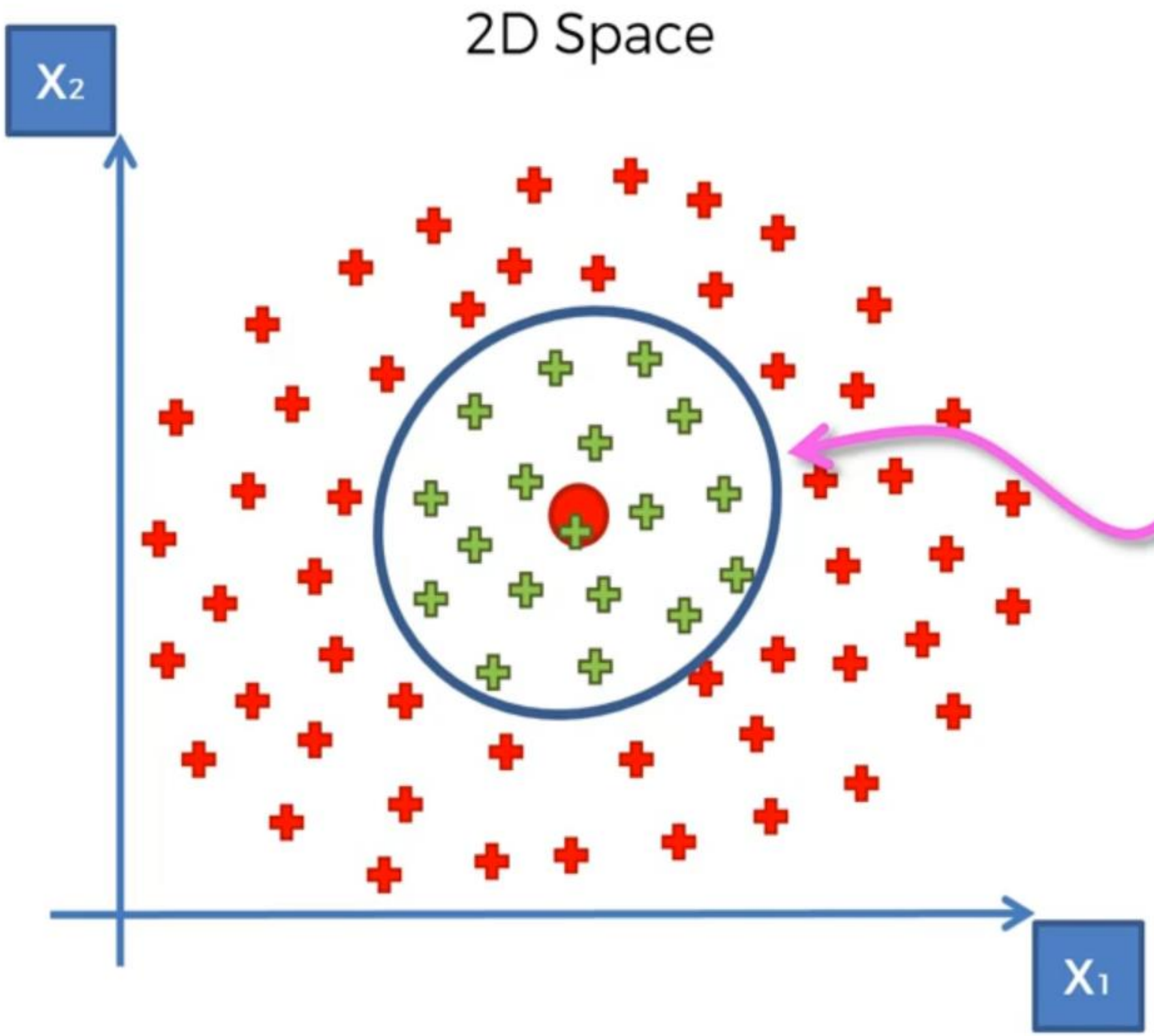
The Gaussian RBF Kernel

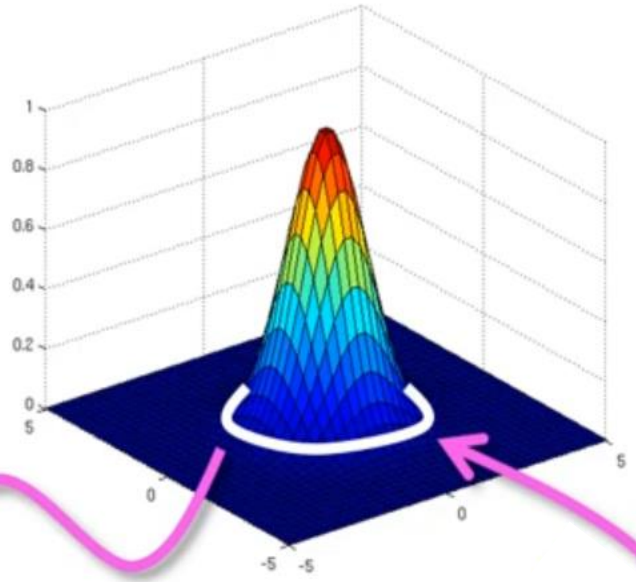




$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\left\|\vec{x}-\vec{l}^i\right\|^2}{2\sigma^2}}$$

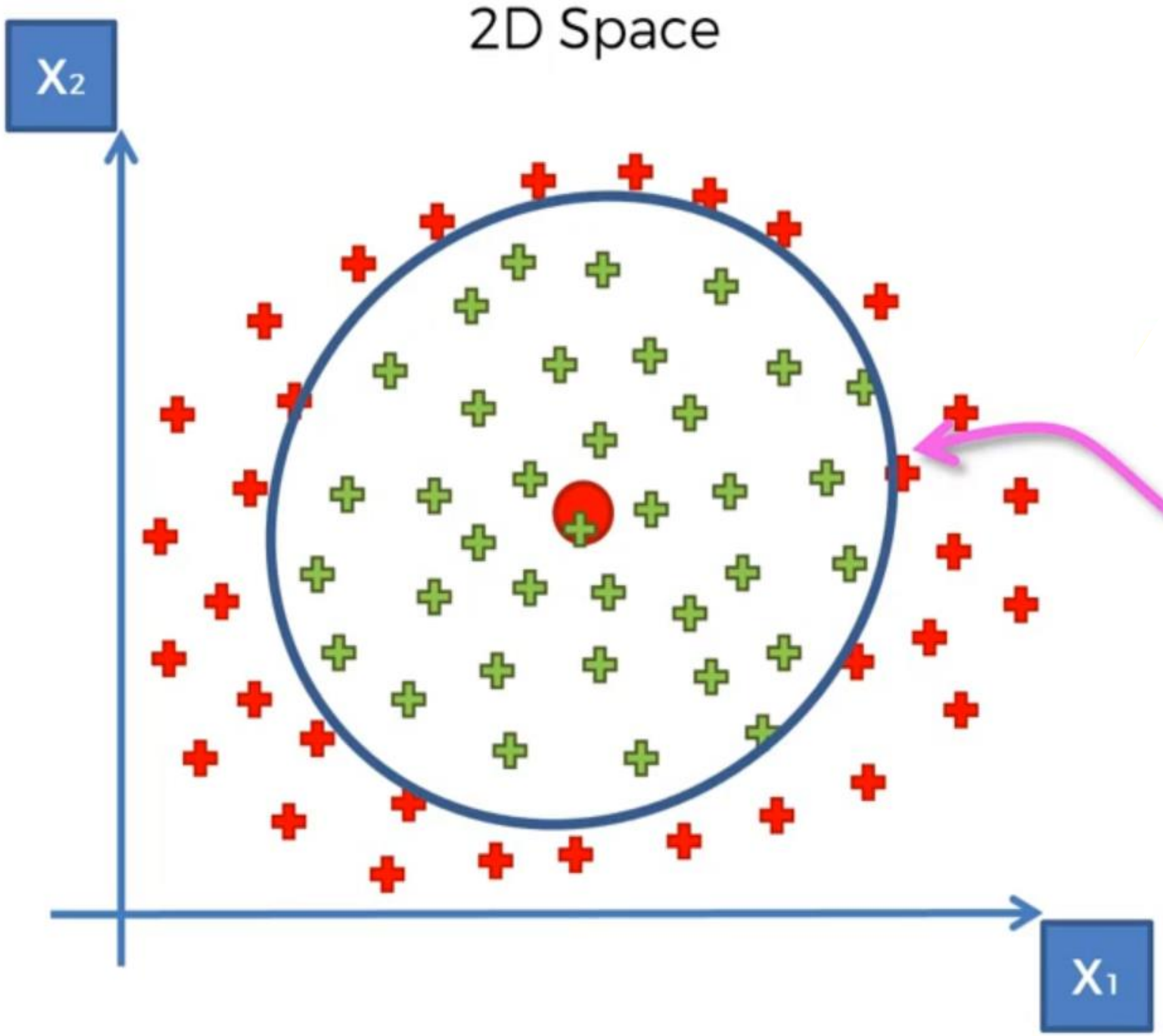
The Gaussian RBF Kernel



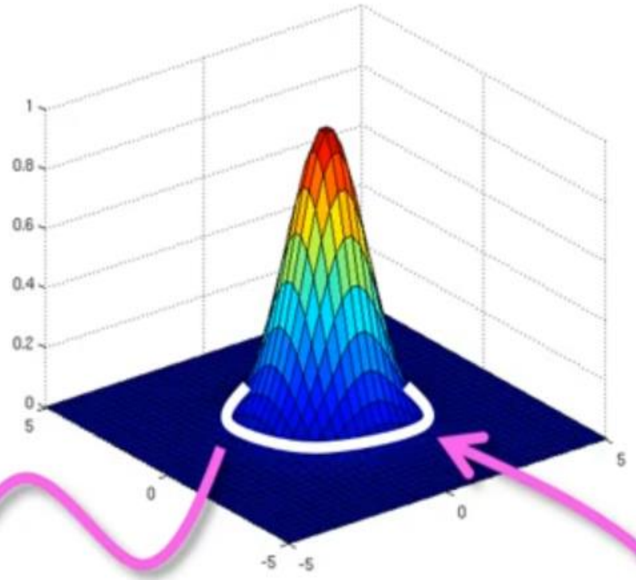


$$K\left(\vec{x},\vec{l}^{i}\right)=e^{-\frac{\left\|\vec{x}-\vec{l}^{i}\right\|^{2}}{2\sigma^{2}}}$$

The Gaussian RBF Kernel

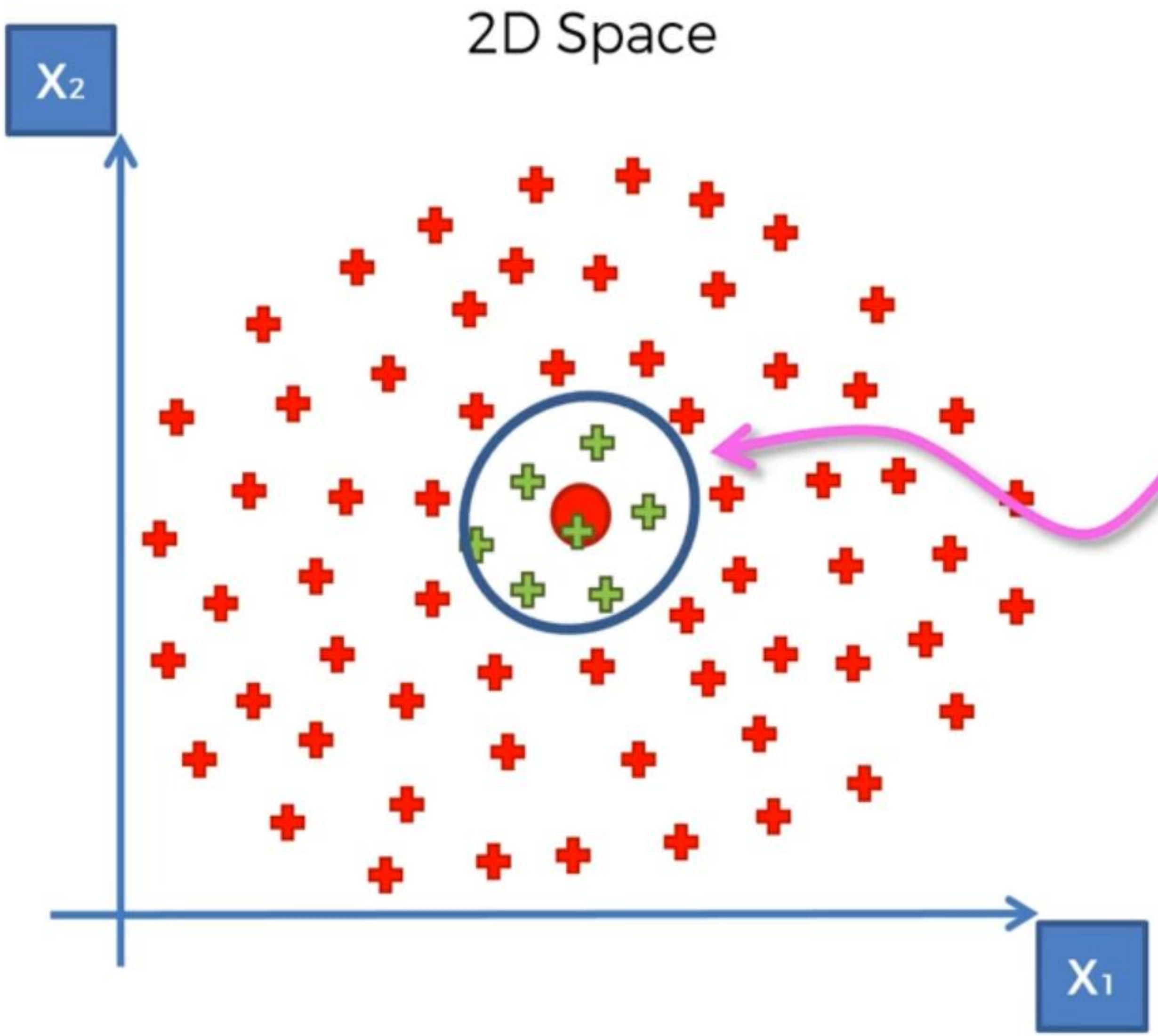


σ

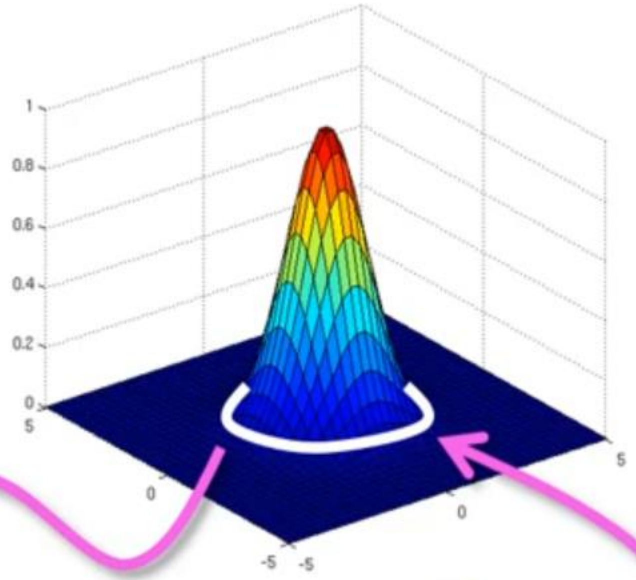


$$K\left(\vec{x}, \vec{l}^{i}\right) = e^{-\frac{\left\|\vec{x}-\vec{l}^{i}\right\|^{2}}{2\sigma^{2}}}$$

The Gaussian RBF Kernel

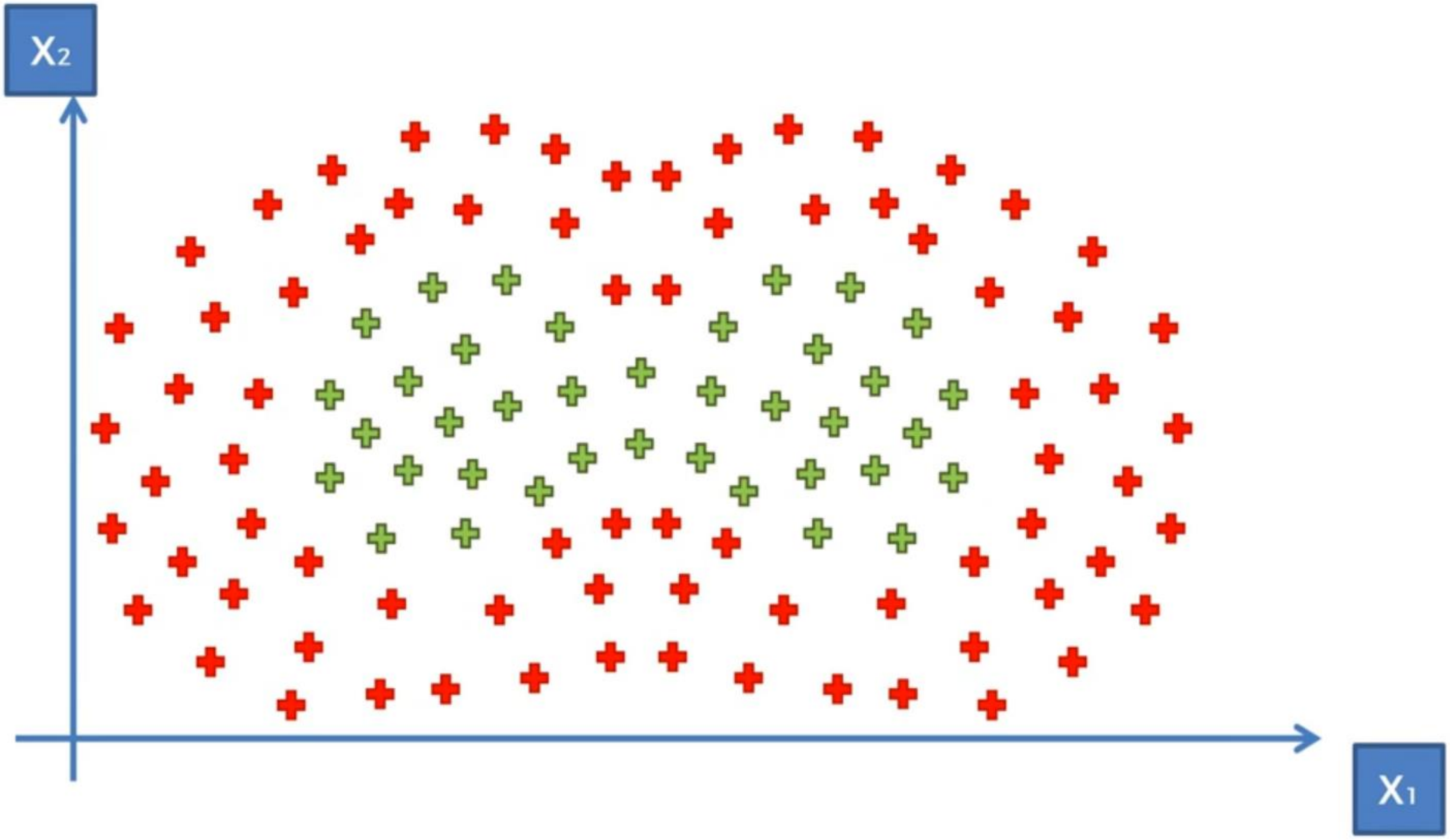


$\sigma \downarrow$

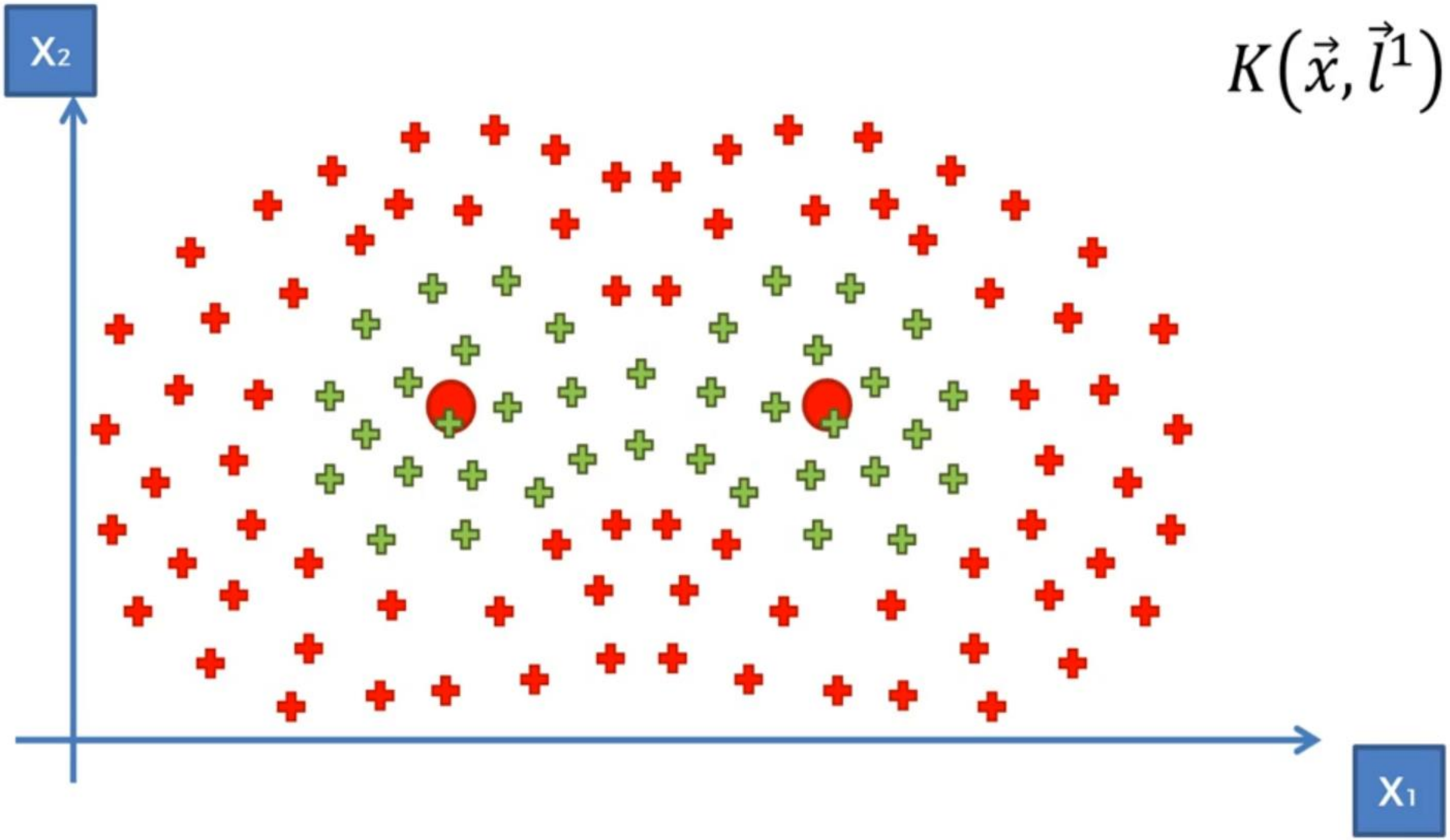


$$K(\vec{x},\vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}}$$

The Gaussian RBF Kernel



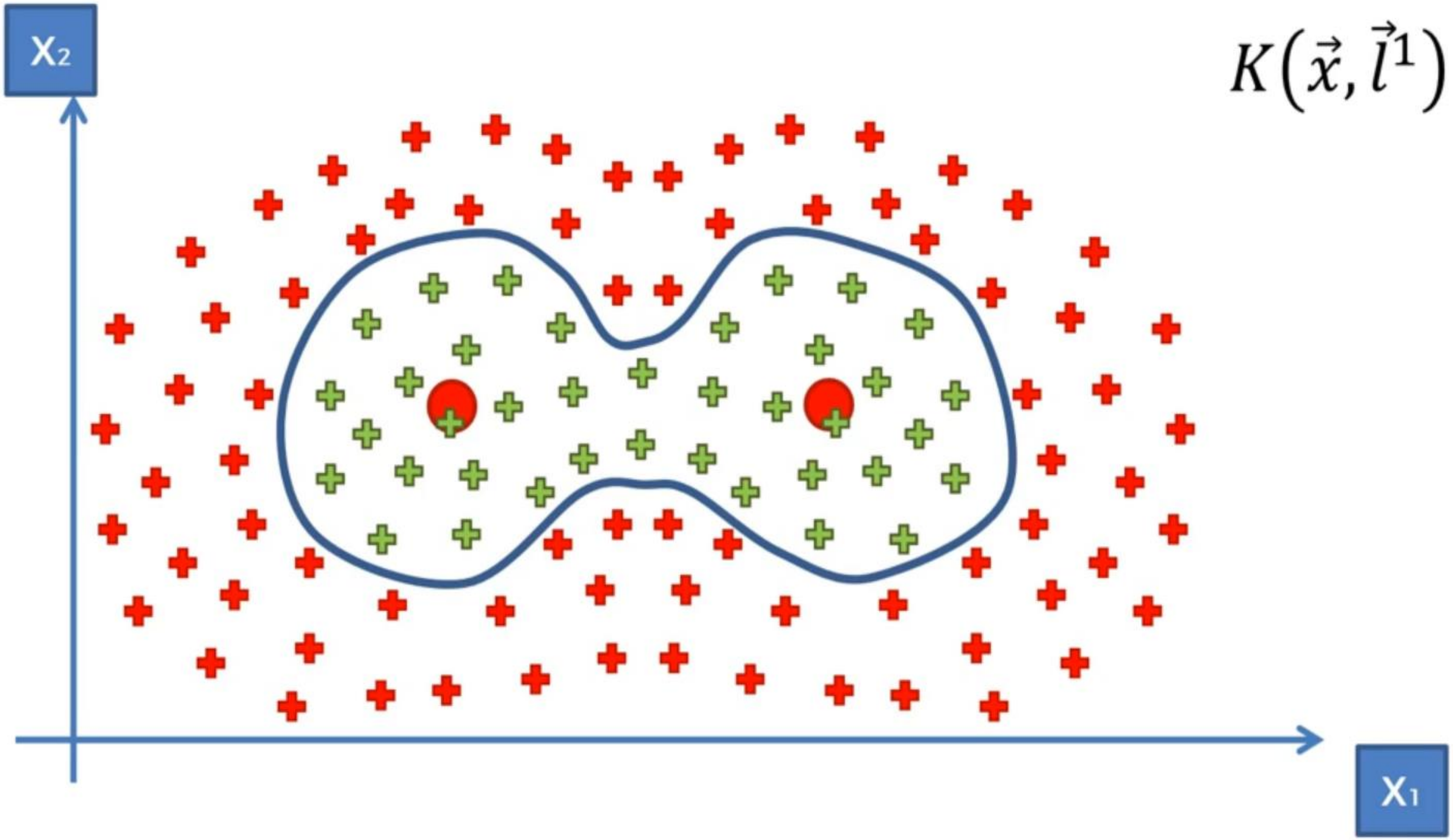
The Gaussian RBF Kernel



$$K(\vec{x}, \vec{l}^1) + K(\vec{x}, \vec{l}^2)$$

(Simplified Formula)

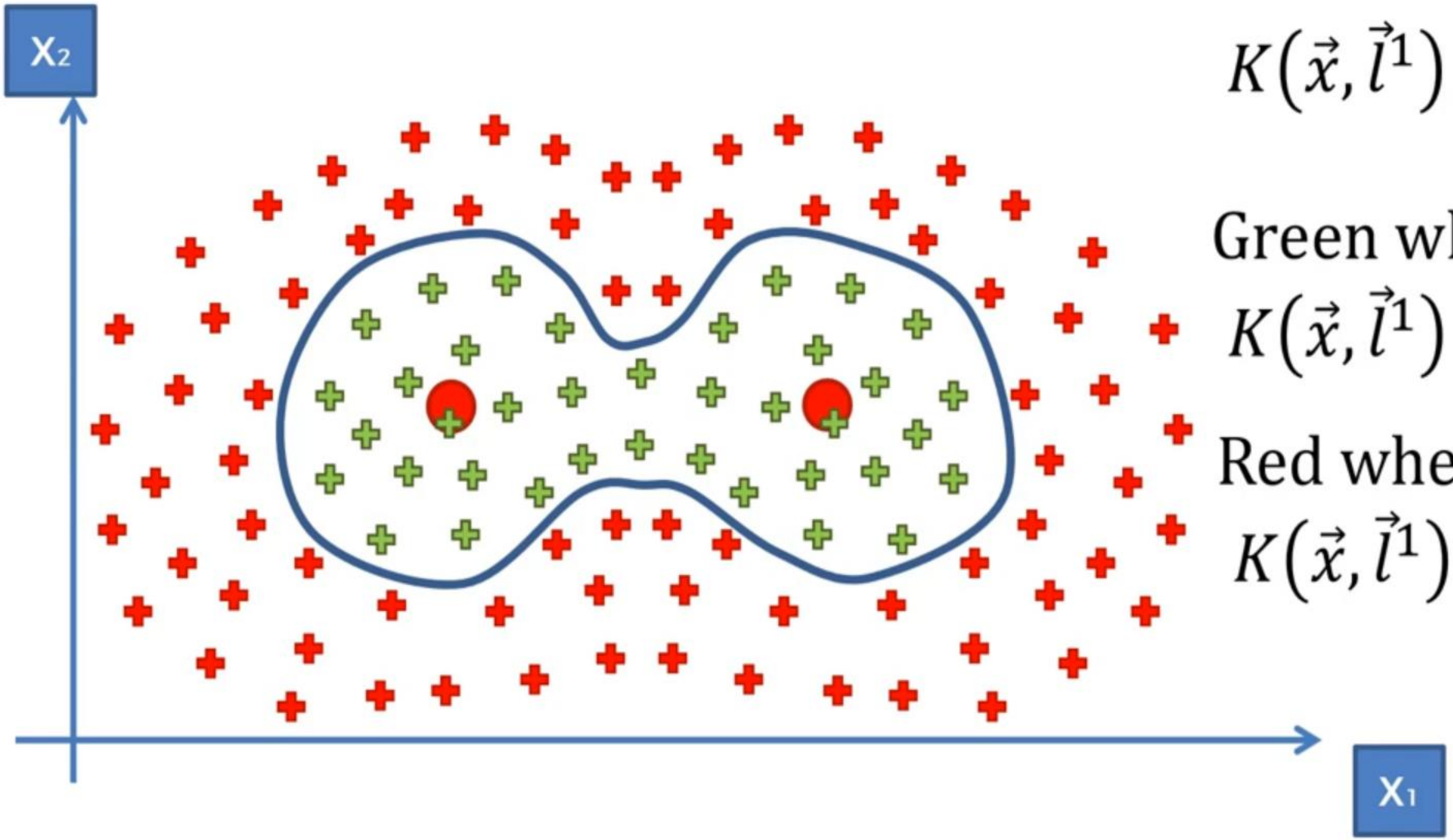
The Gaussian RBF Kernel



$$K\left(\vec{x}, \vec{l}^{1}\right) + K\left(\vec{x}, \vec{l}^{2}\right)$$

(Simplified Formula)

The Gaussian RBF Kernel



$$K\left(\vec{x},\vec{l}^{1}\right)+K\left(\vec{x},\vec{l}^{2}\right)$$

(Simplified Formula)

Green when:

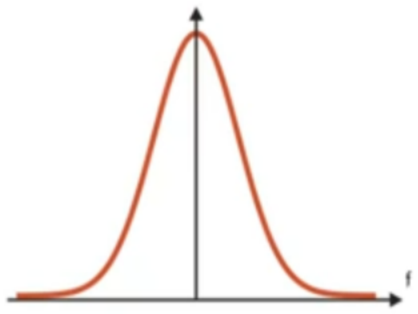
$$K\left(\vec{x},\vec{l}^{1}\right)+K\left(\vec{x},\vec{l}^{2}\right)>0$$

Red when:

$$K\left(\vec{x},\vec{l}^{1}\right)+K\left(\vec{x},\vec{l}^{2}\right)=0$$

Types of Kernel Functions

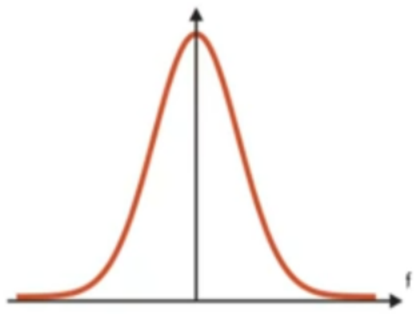
Types of Kernel Functions



Gaussian RBF Kernel

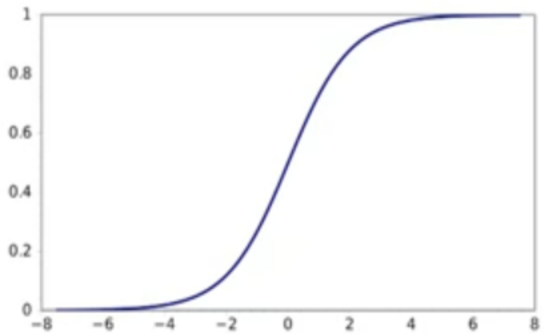
$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\left\|\vec{x}-\vec{l}^i\right\|^2}{2\sigma^2}}$$

Types of Kernel Functions



Gaussian RBF Kernel

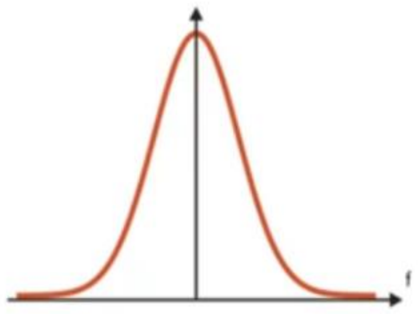
$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}}$$



Sigmoid Kernel

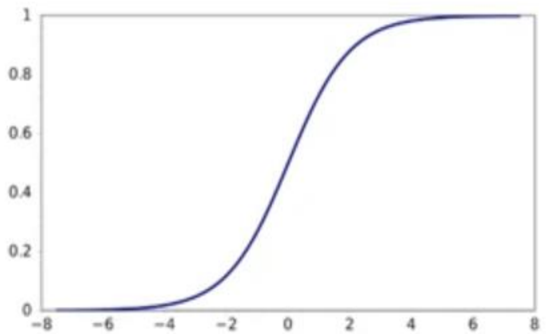
$$K(X, Y) = \tanh(\gamma \cdot X^T Y + r)$$

Types of Kernel Functions



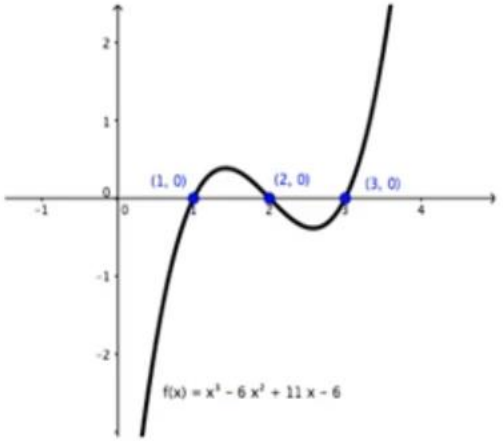
Gaussian RBF Kernel

$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}}$$



Sigmoid Kernel

$$K(X,Y) = \tanh(\gamma \cdot X^T Y + r)$$



Polynomial Kernel

$$K(X,Y) = (\gamma \cdot X^T Y + r)^d, \gamma > 0$$