Machine Learning

Support Vector Machine

Session 5

Session 5

Content

- What is support vector machine
- Cost function
- SVM decision boundaries
- Large margin classifier in present of outlier
- Kernel

What is support vector machine

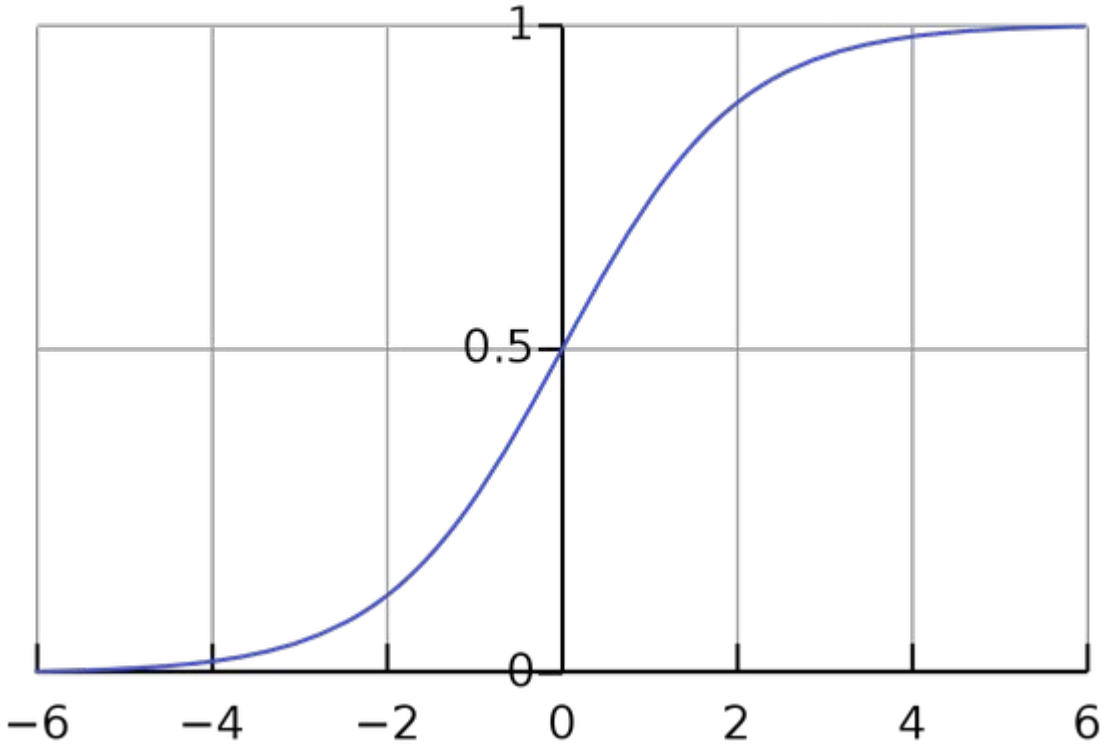
A supervised learning classification algorithm used to predict labels for categorical data.

What is support vector machine

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

if y = 1?

if y = 0?



Cost Function

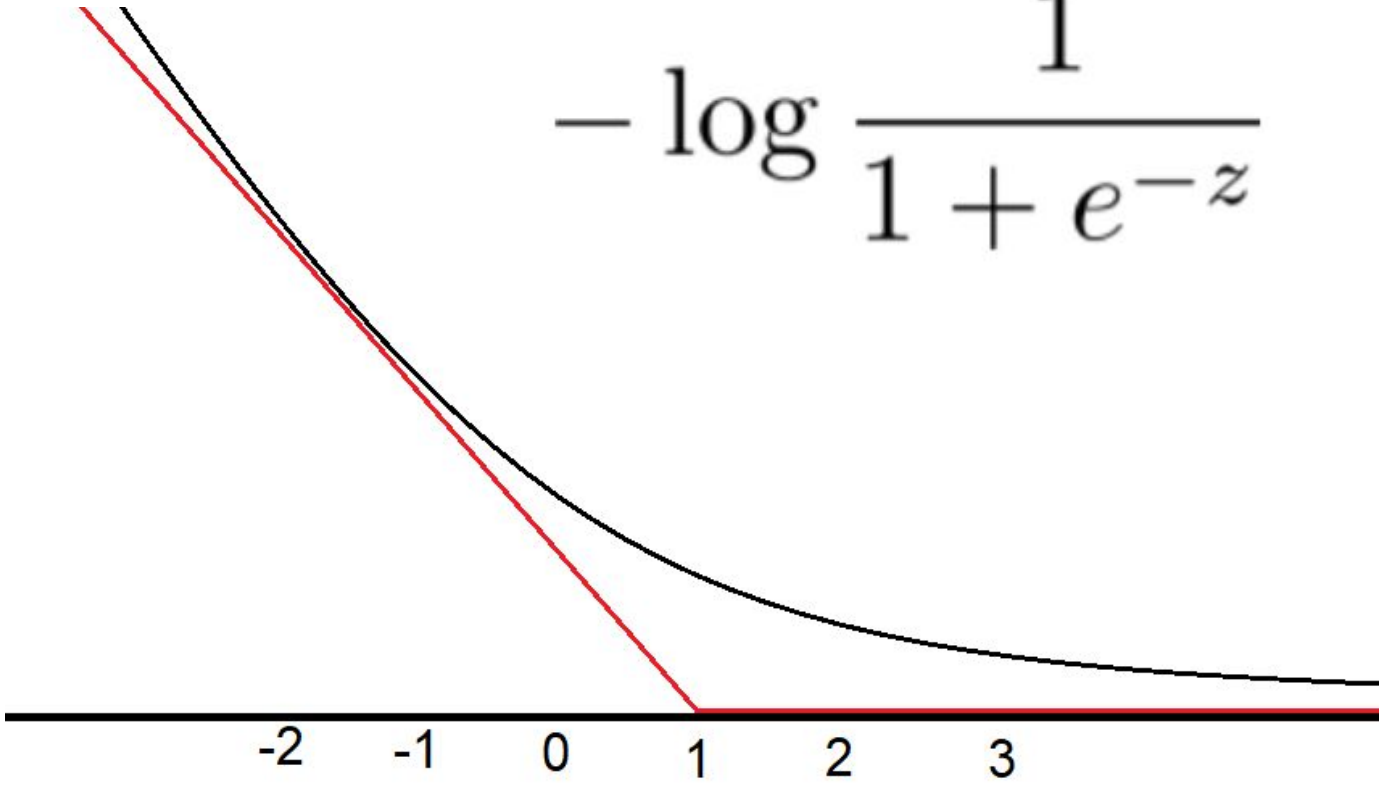
$$y\,log\big(h_\theta(x)\big)+(1-y)log\big(1-h_\theta(x)\big)$$

$$h_\theta(x)=\frac{1}{1+e^{-\theta^T x}}$$

Cost Function

If $y = 1$ (want $\theta^T x \gg 0$):

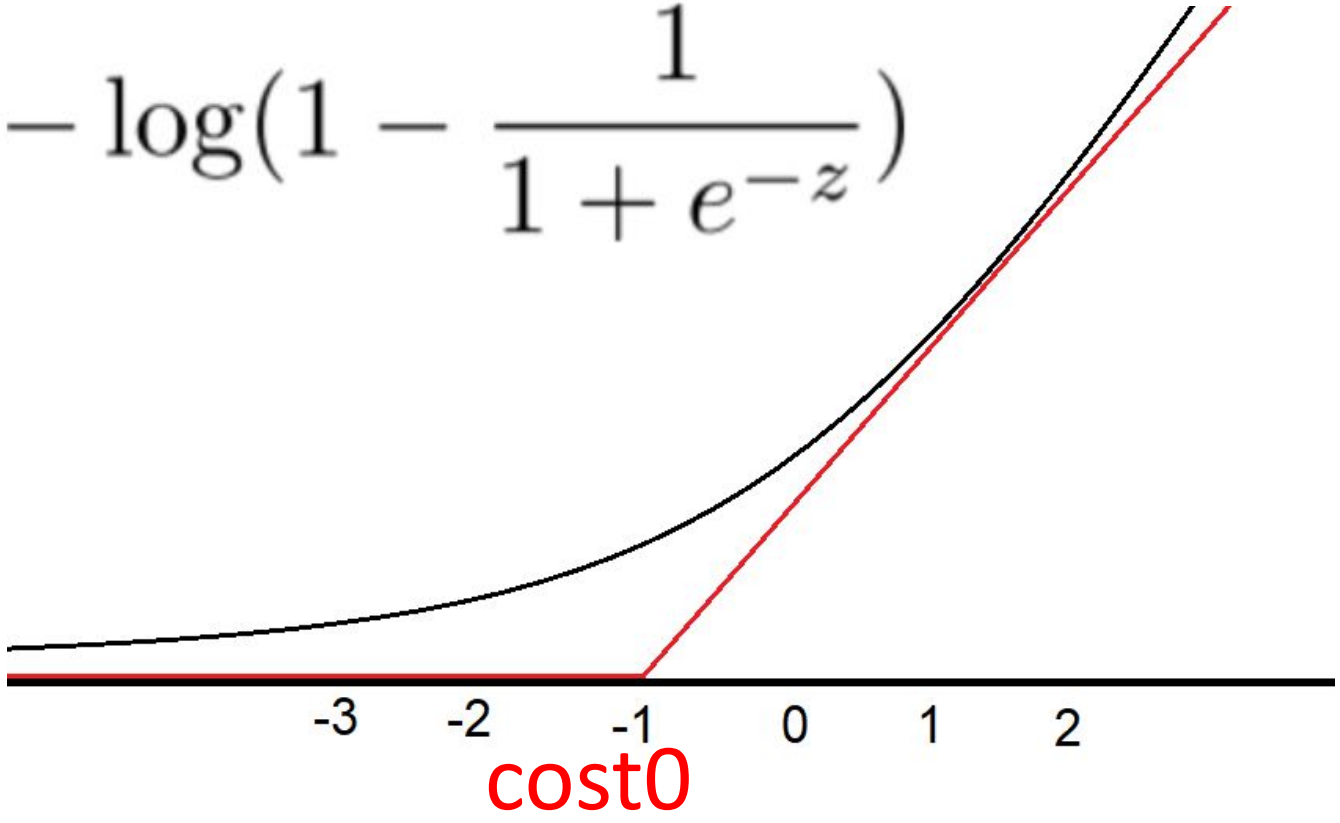
$$-\log \frac{1}{1+e^{-z}}$$



cost1

Cost Function

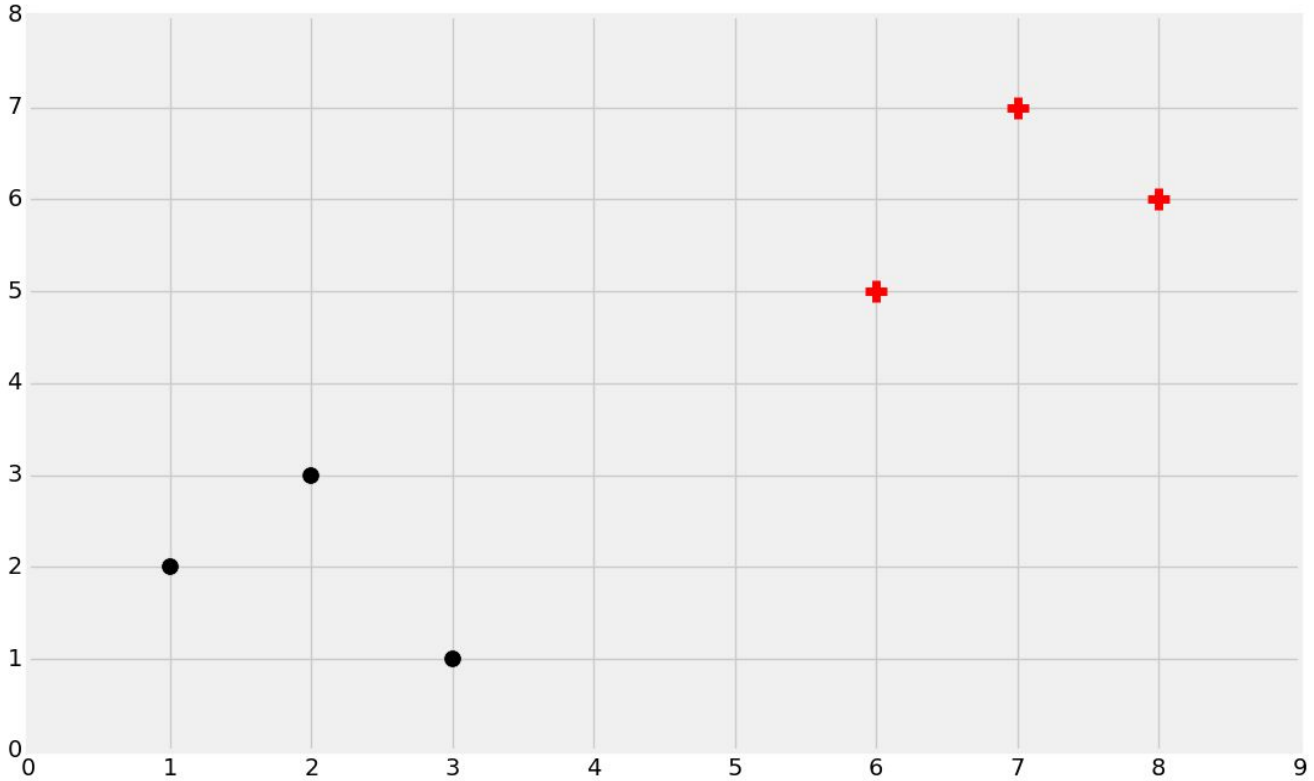
If $y = 0$ (want $\theta^T x \ll 0$):



SVM Decision Boundary:

Linearly separable case

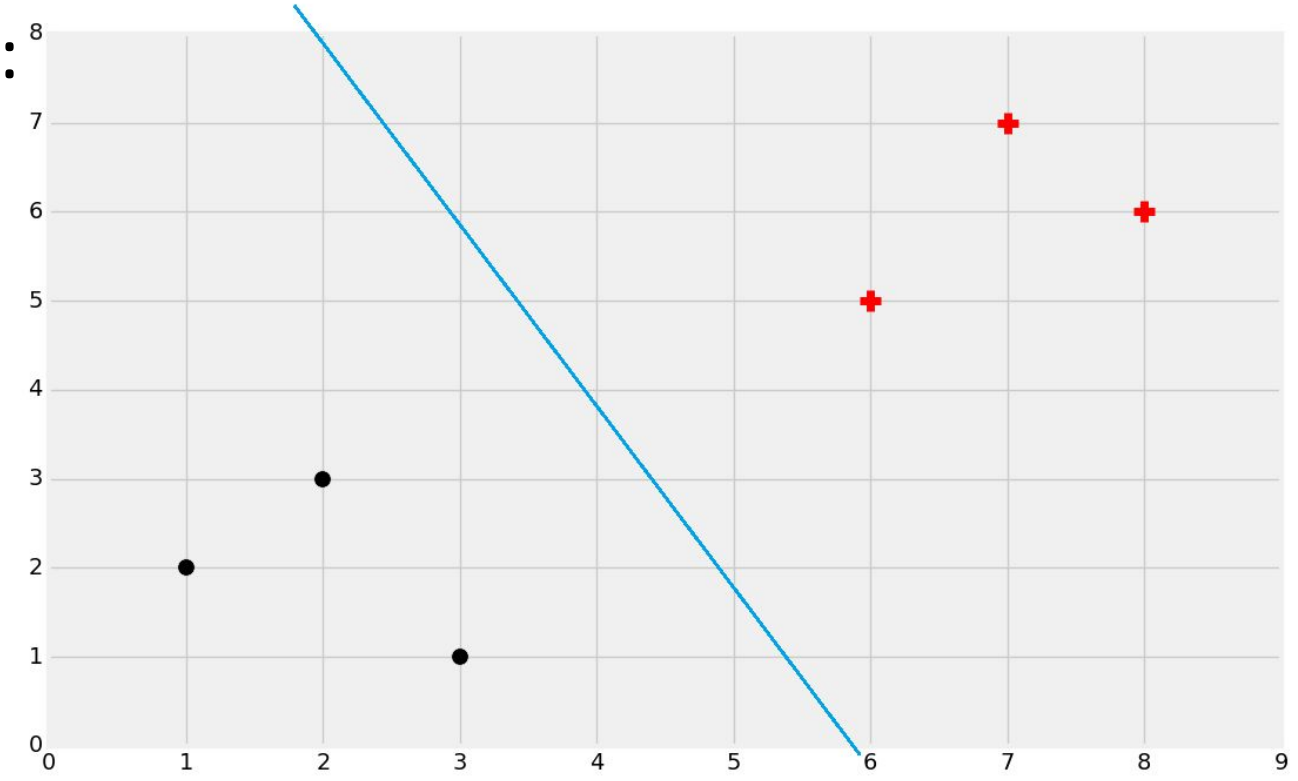
- There will be two classes named as :
  1. positive class
  2. negative class



SVM Decision Boundary:

The objective of the Support Vector Machine is to find the best splitting boundary between data.
And thus, we will have two classes:

-Positive class: above hyperplane
-Negative class: below hyperplane

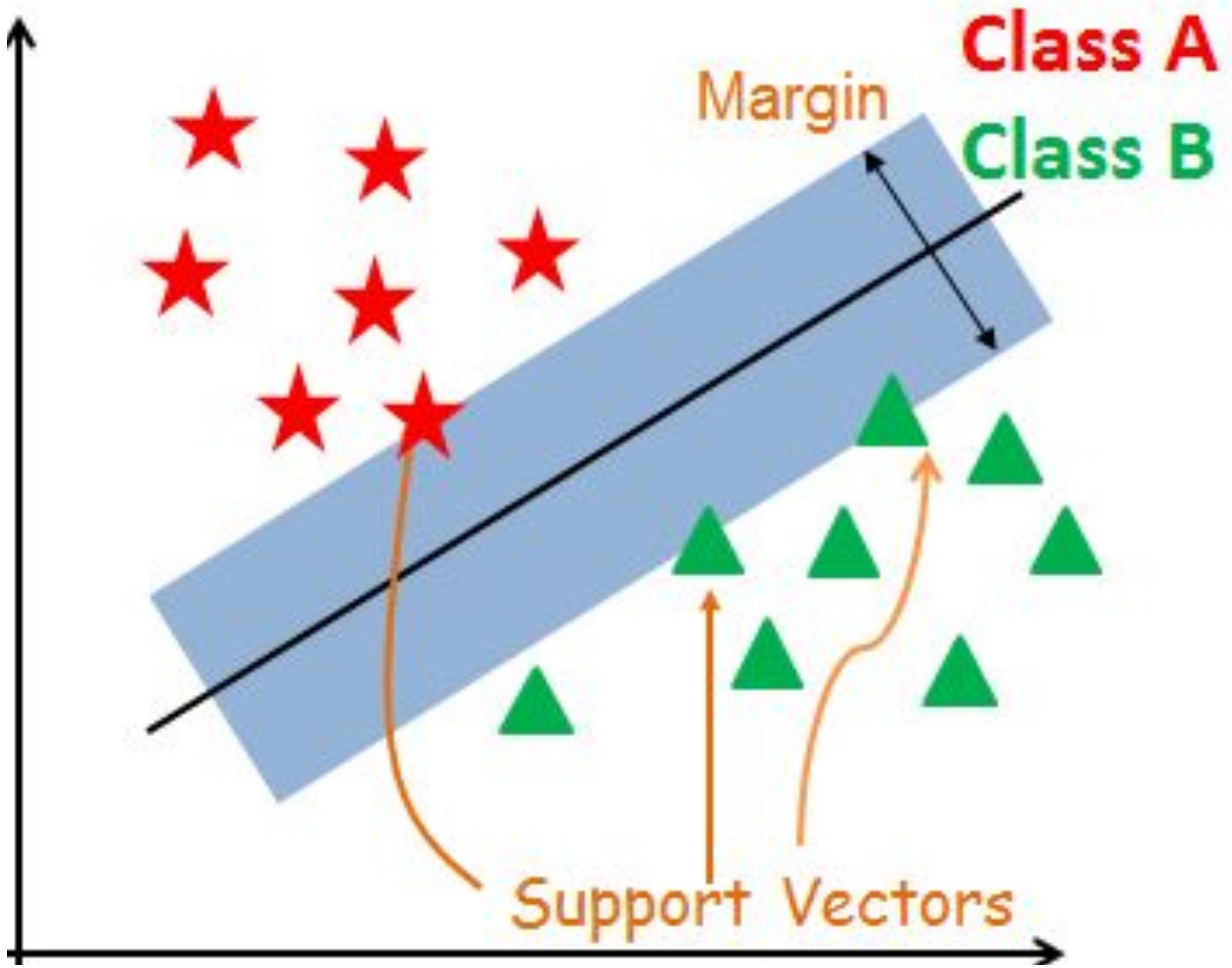


SVM Decision Boundary:

Support Vectors

Support vectors are the data points, which are closest to the hyperplane. These points will define the separating line better by calculating margins. These points are more relevant to the construction of the classifier.

SVM Decision Boundary:



SVM Decision Boundary:

Hyperplane

A hyperplane is a decision plane which separates between a set of objects having different class memberships.

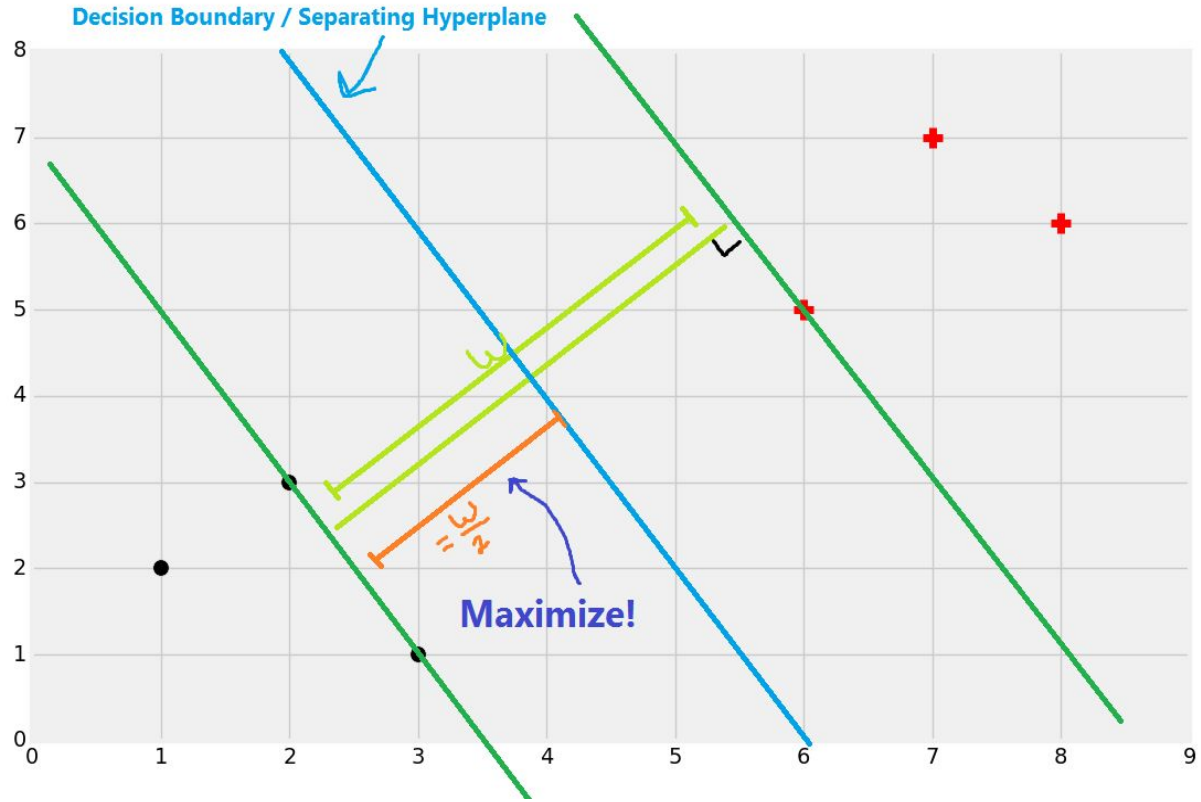
SVM Decision Boundary:

Margin

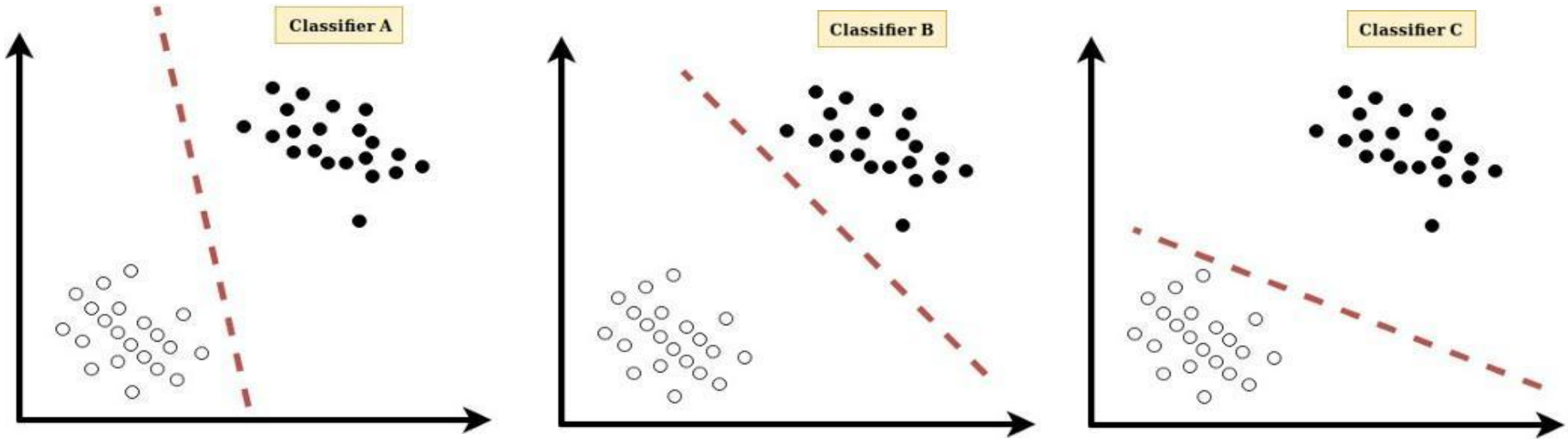
A margin is a gap between the two lines on the closest class points. This is calculated as the perpendicular distance from the line to support vectors or closest points. If the margin is larger in between the classes, then it is considered a good margin, a smaller margin is a bad margin.

SVM Decision Boundary:

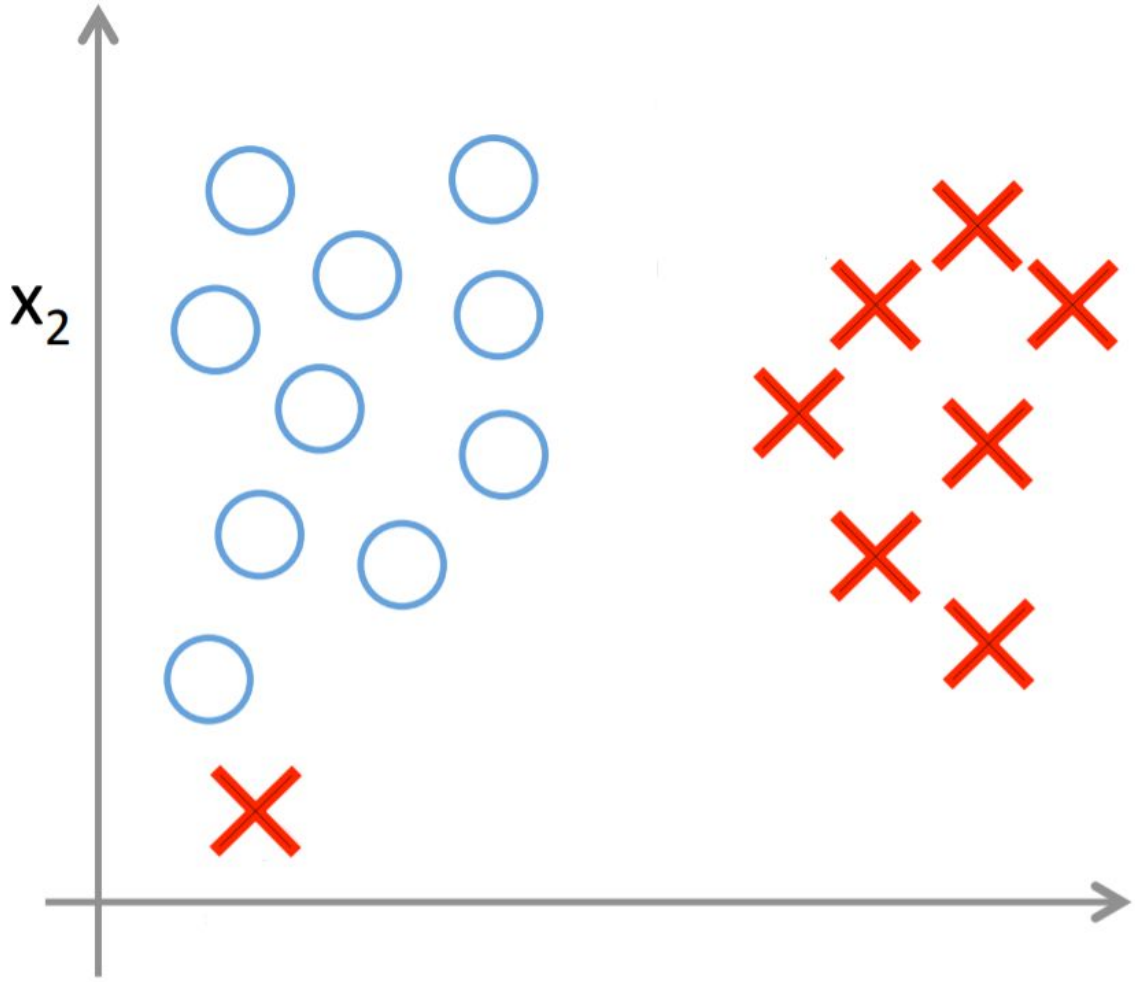
The best separating hyperplane is defined as the hyperplane that contains the "widest" margin between support vectors.



SVM Decision Boundary:



Large margin classifier in presence of outliers

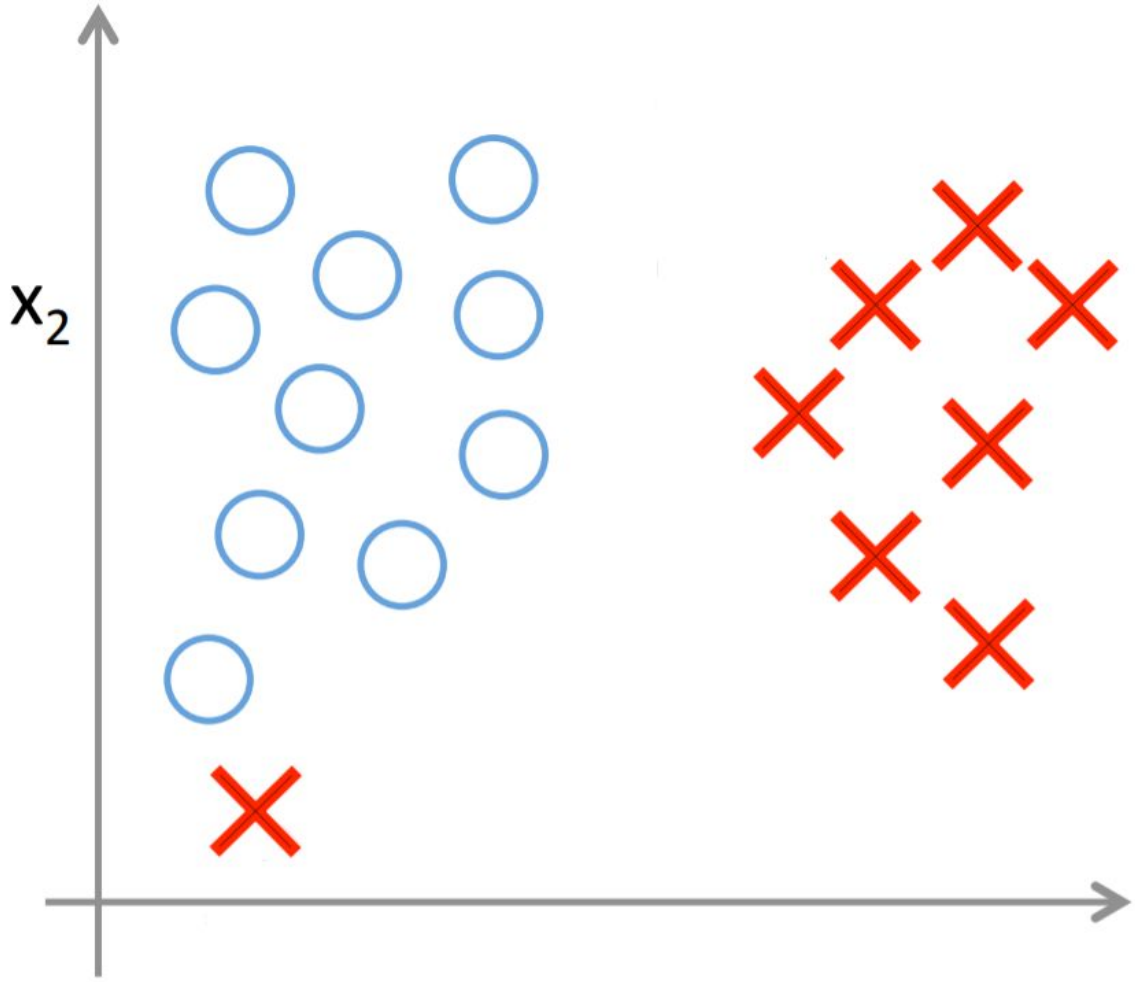


Large margin classifier in presence of outliers

$$\min_{\theta} \frac{1}{m} \left[\sum_{i=1}^{m} y^{(i)} \left(-\log h_\theta(x^{(i)}) \right) + (1 - y^{(i)}) \left((-\log(1 - h_\theta(x^{(i)})) \right) \right] + \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2$$

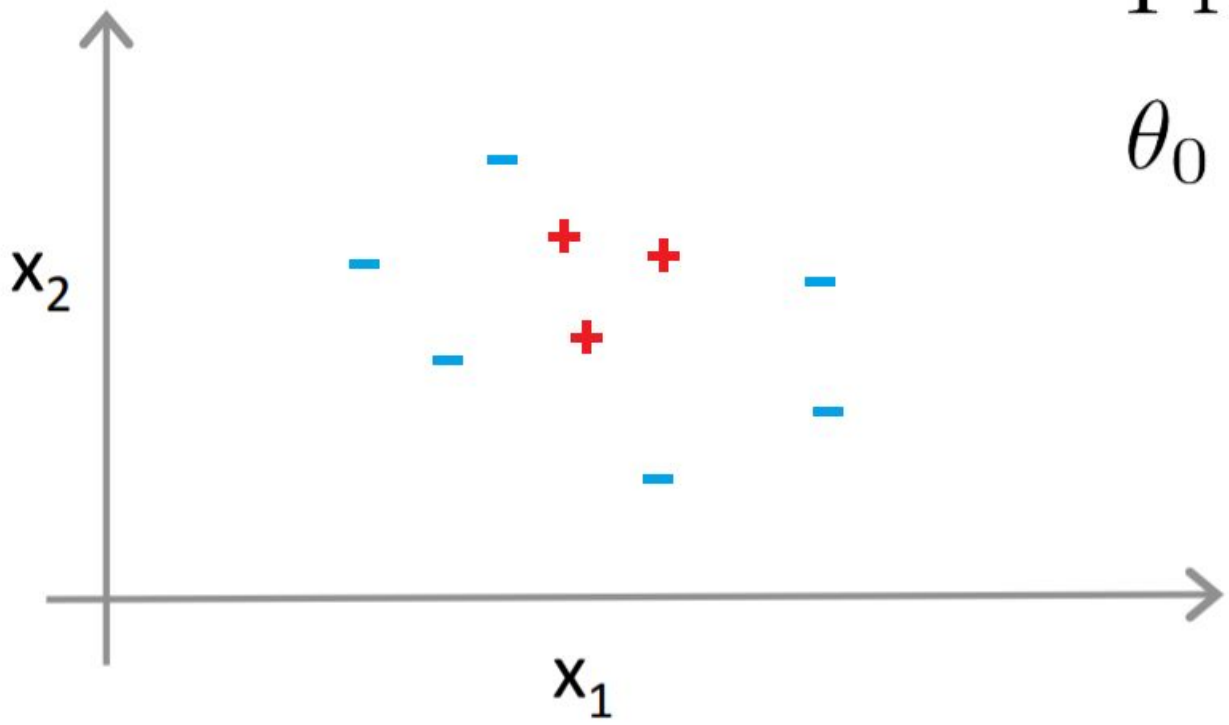
$$\min_{\theta} C \sum_{i=1}^{m} \left[y^{(i)} cost_1(\theta^T x^{(i)}) + (1 - y^{(i)}) cost_0(\theta^T x^{(i)}) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_j^2$$

Large margin classifier in presence of outliers



Kernels

Kernel

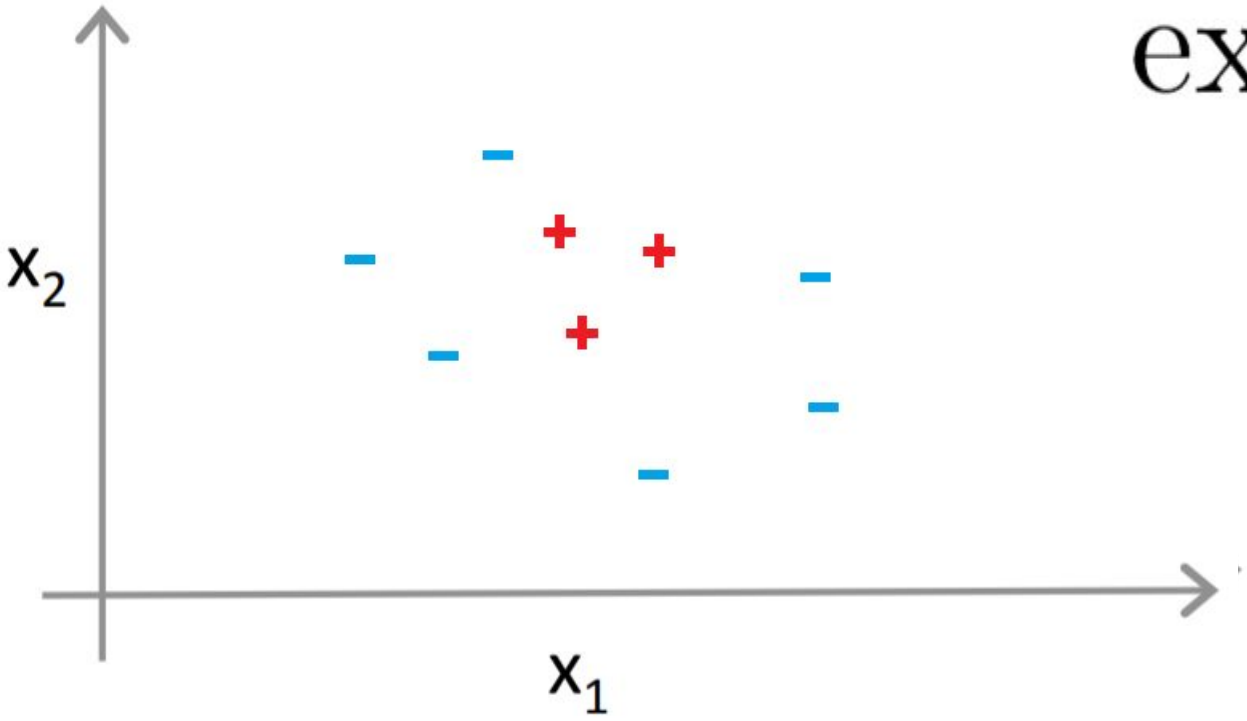


$$\text{Predict } y = 1 \text{ if}$$

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \cdots \geq 0$$

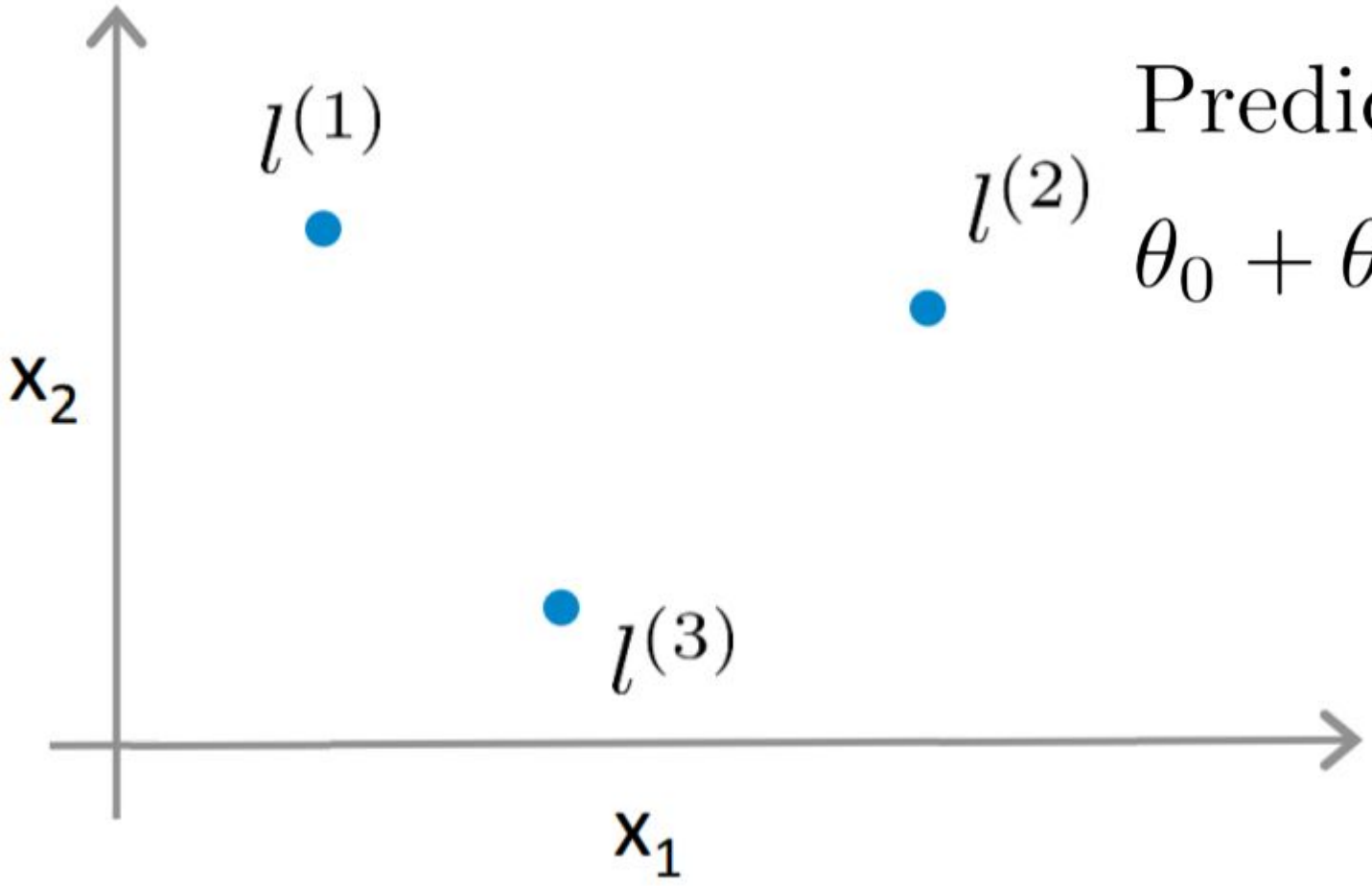
Kernels

Kernel



$$\exp\left(-\frac{\|x-l^{(1)}\|^2}{2\sigma^2}\right)$$

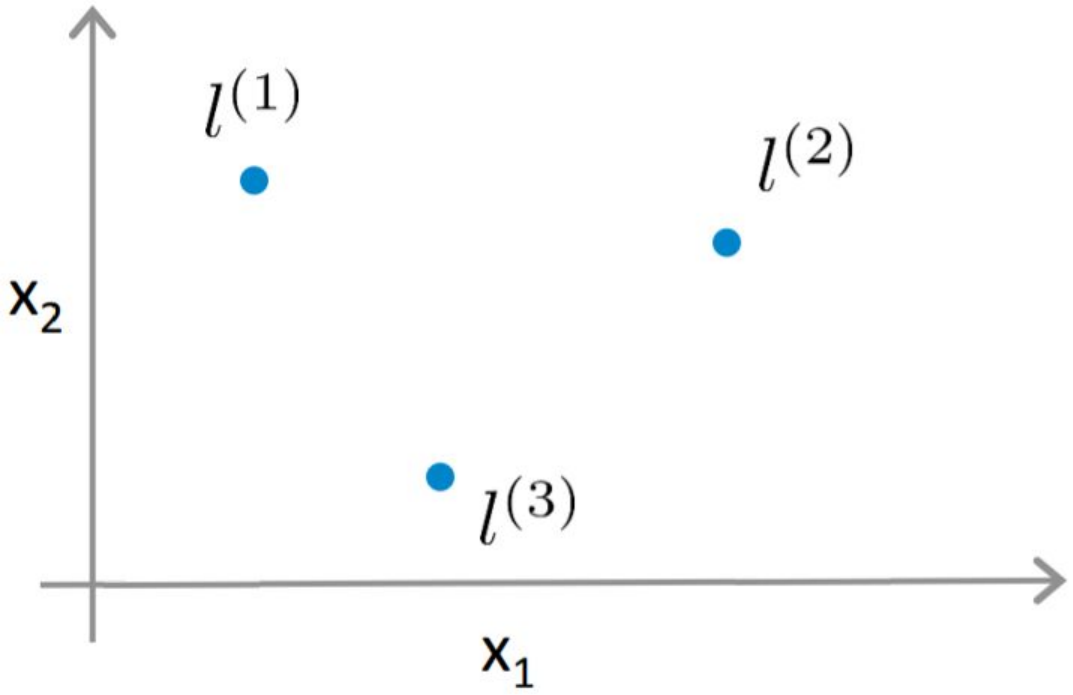
Kernels



Predict "1" when

$$\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$$

Kernels



Predict "1" when

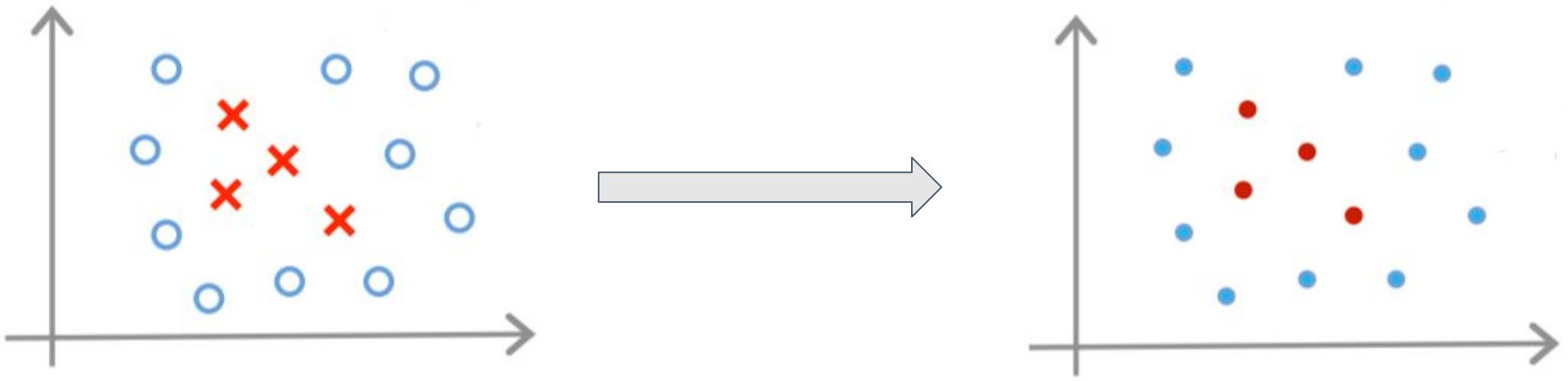
$$\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$$

-0.5 1 1 0

Kernels

Predict $y = 1$ if $\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$

Where to get $l^{(1)}, l^{(2)}, l^{(3)}, \ldots$?



Kernels

other types of kernel:

- Polynomial kernel
- String kernel
- chi-square kernel
- histogram
- intersection kernel

SVM parameters

use SVM software package (e.g. liblinear, libsvm,...)

need to chose:

- **C**
- **kernel**

more parameters:

σ, degree, gamma, ……

Goodbye