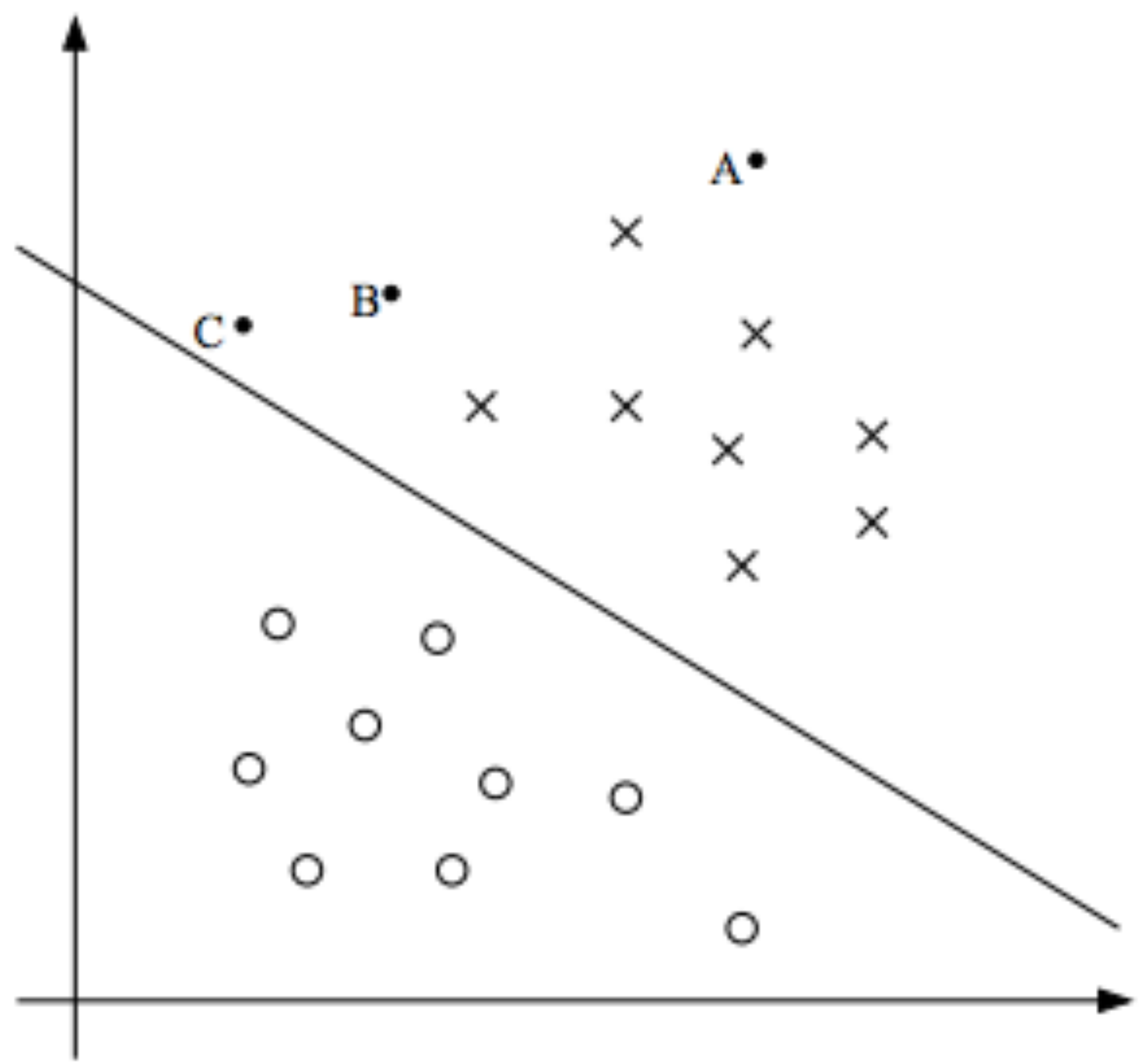
SVMs

Separating Hyperplane



Hypothesis

$$h_{w,b}(x) = g(w^T x + b).$$

1 if z >= 0

-1 otherwise

Functional and Geometric Margins

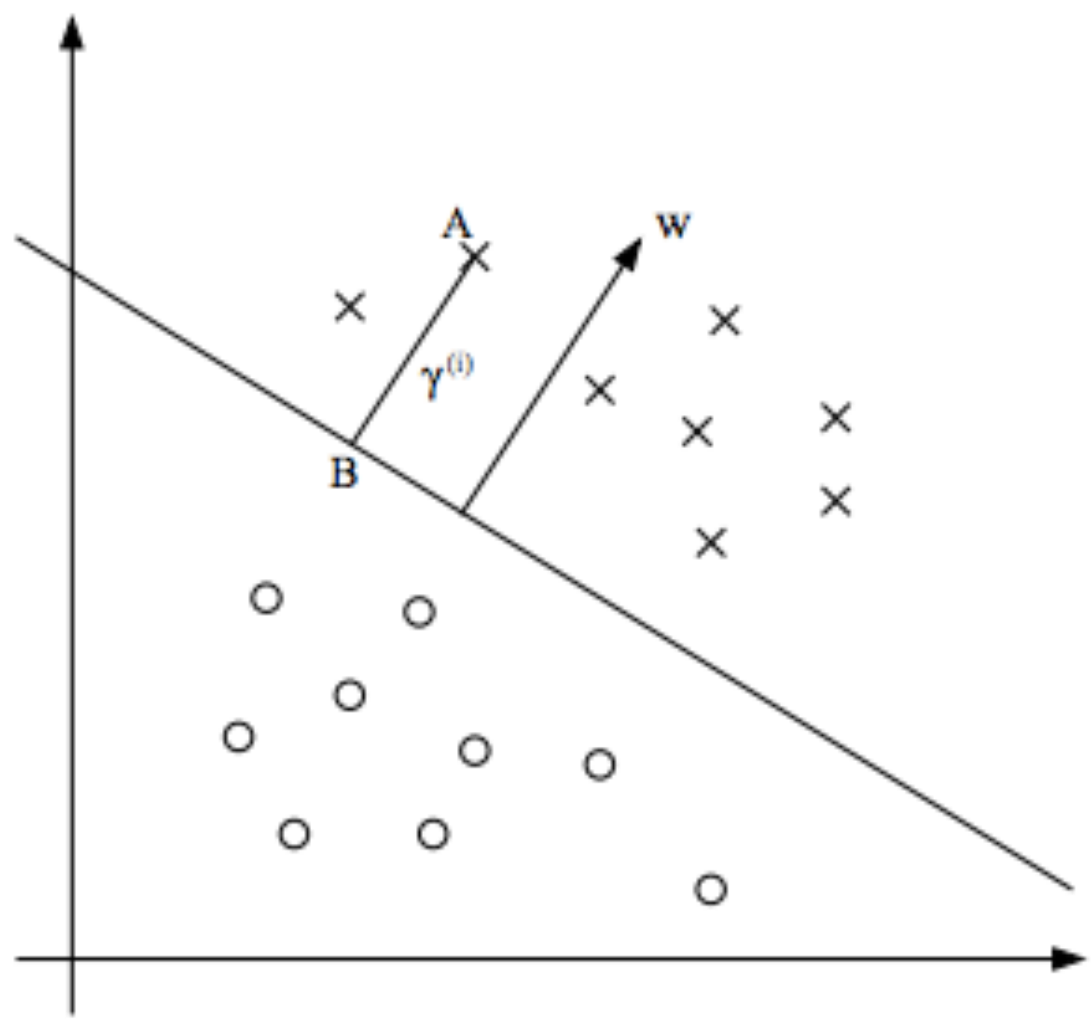
Functional Margin

$$\hat{\gamma}^{(i)} = y^{(i)}(w^T x + b).$$

When is our prediction we correct?

Smallest over training set: $\hat{\gamma} = \min_{i=1,\ldots,m} \hat{\gamma}^{(i)}.$

Geometric Margin



Why is **w** perpendicular to the hyperplane?

We want to find the distance gamma

A is the point x(i)

So B is given by: $x^{(i)} - \gamma^{(i)} \cdot w/||w||.$

We also know that the hyperplane satisfies:

$$w^T x + b = 0.$$

So let's sub in our point B to solve for gamma:

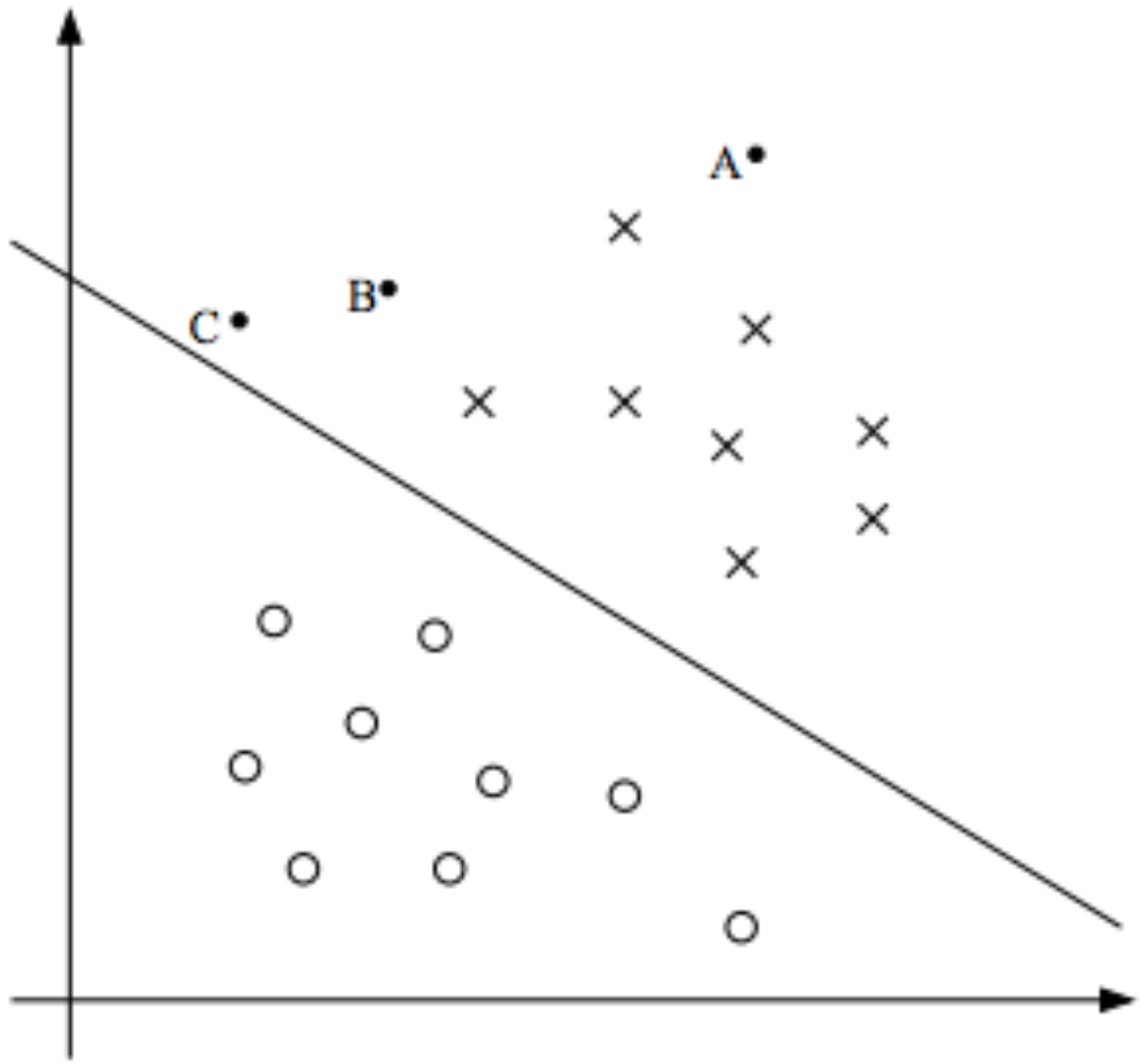
$$w^T \left(x^{(i)} - \gamma^{(i)} \frac{w}{||w||} \right) + b = 0.$$

$$\gamma^{(i)} = \frac{w^T x^{(i)} + b}{||w||} = \left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||}.$$

For y +/-

$$\gamma^{(i)} = y^{(i)} \left(\left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||} \right)$$

The Optimal Margin Classifier



Naive problem

$$
\begin{aligned}
\max_{\gamma,w,b} \quad & \gamma \\
\text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq \gamma, \;\; i = 1, \ldots, m \\
& ||w|| = 1.
\end{aligned}
$$

ugly (non-convex)!

Convert to functional margin

$$
\begin{aligned}
\max_{\hat{\gamma},w,b} \quad & \frac{\hat{\gamma}}{||w||} \\
\text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, \;\; i = 1, \ldots, m
\end{aligned}
$$

Let $\hat{\gamma} = 1.$

And minimise equivalently

$$\begin{aligned} \min_{\gamma,w,b} \quad & \frac{1}{2}||w||^2 \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, m \end{aligned}$$

Works on off-the-shelf software!

Kernels

Can we map our features to more interesting ones

$$\phi(x) = \begin{bmatrix} x \\ x^2 \\ x^3 \end{bmatrix}$$

Kernel given feature mapping

$$K(x, z) = \phi(x)^T \phi(z)$$

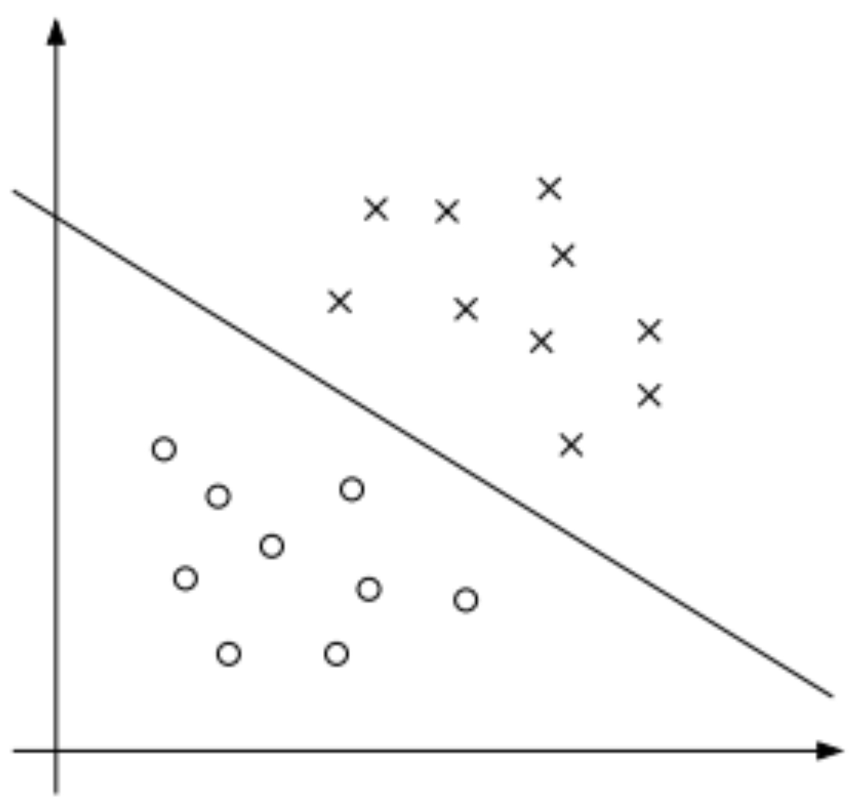
Essentially a distance function anywhere we had {x, y} in our algorithm

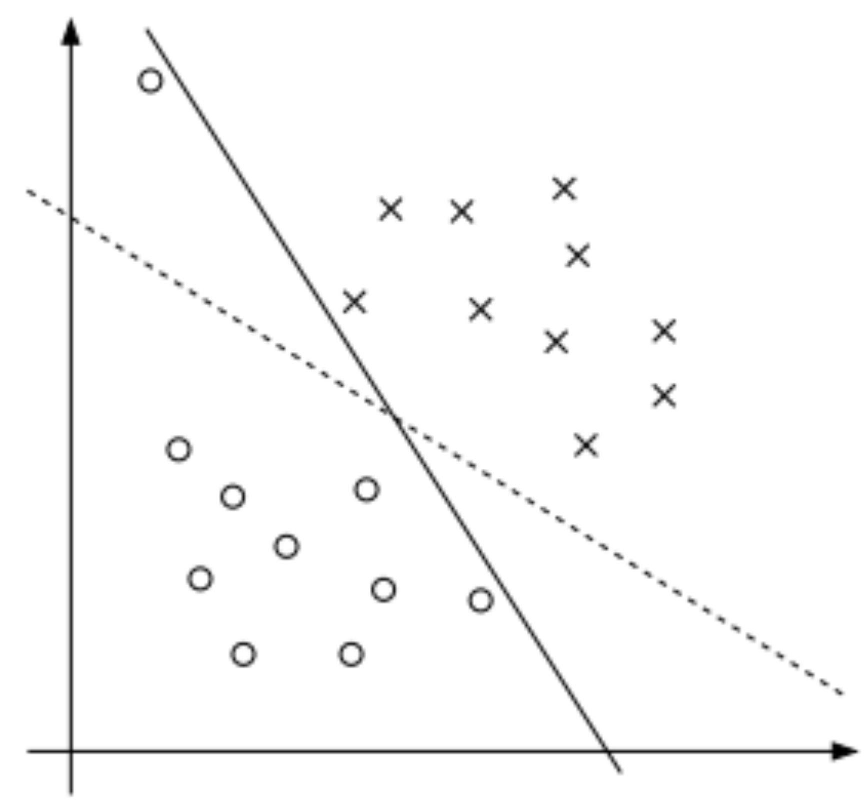
Infinite dimensional Gaussian kernel

$$K(x, z) = \exp\left(-\frac{||x - z||^2}{2\sigma^2}\right)$$

How to choose?
http://stats.stackexchange.com/questions/18030/how-to-select-kernel-for-svm

Regularisation, Separability





Regularisation

$$\begin{aligned}
\min_{\gamma,w,b} \quad & \frac{1}{2}||w||^2 + C\sum_{i=1}^{m}\xi_i \\
\text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \quad i = 1, \ldots, m \\
& \xi_i \geq 0, \quad i = 1, \ldots, m.
\end{aligned}$$

Exercise

- Go over the tutorial, it's great: http://scikit-learn.org/stable/modules/svm.html
- Try to apply it to your data set in a way (if impossible, try clustering)