Support Vector Machines

Cláudia Antunes

Instituto Superior Técnico – Universidade de Lisboa

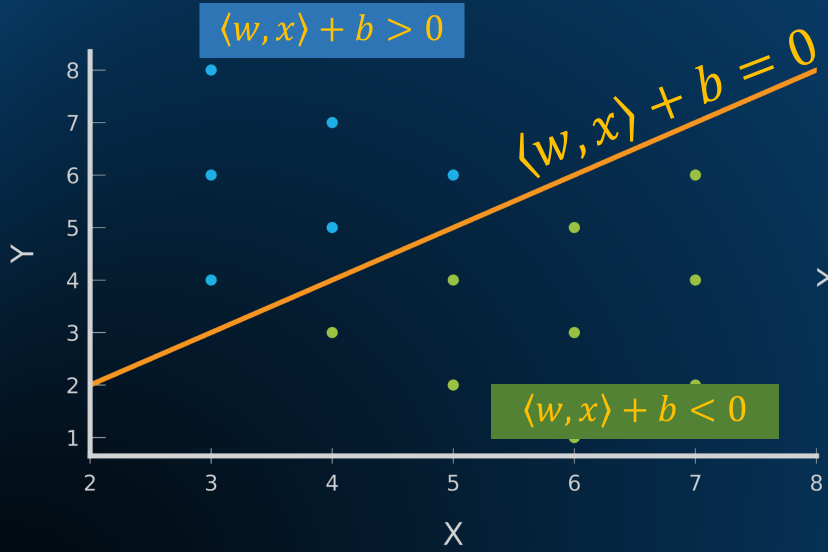
Learning Theory

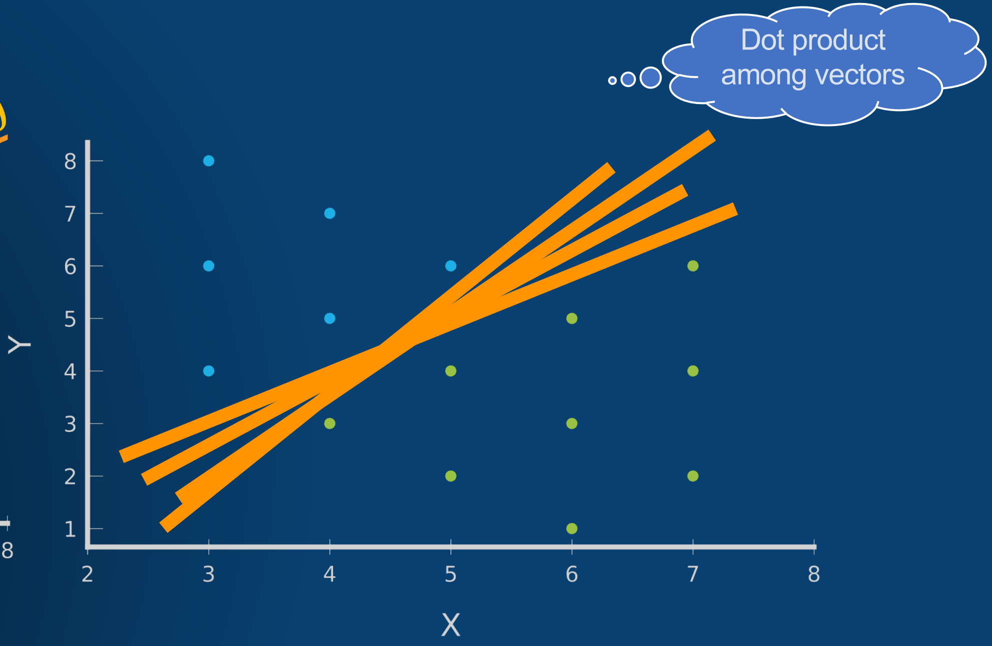
1971 – Vapnik-Chervonenkis dimension and th
convergency of the learning process

1995 – Support Vector Machines

Vladimir Vapnik

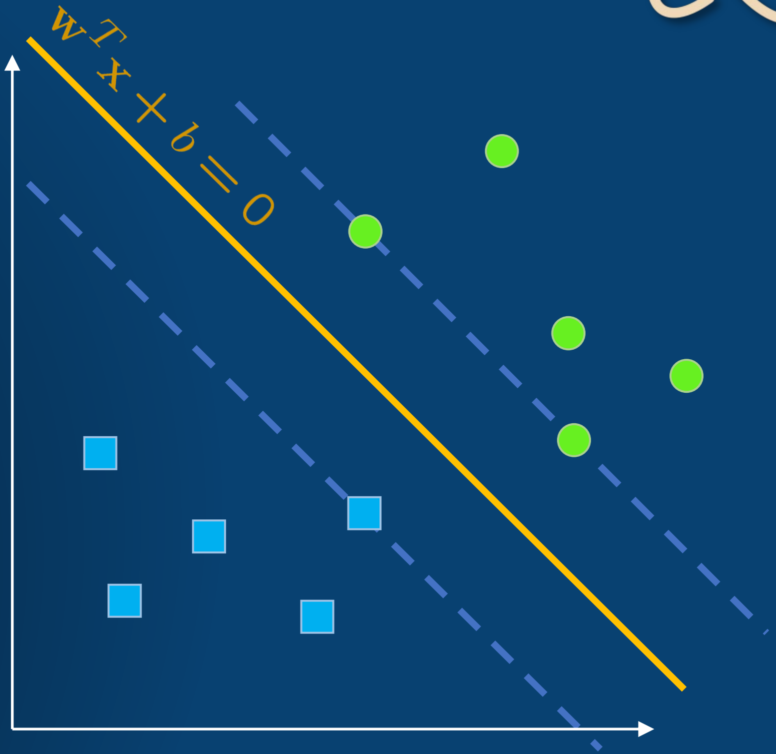
Linear Separability





Maximum–Margin Hyperplane

$$f(x) = sgn(\langle w, x \rangle + \mathbf{b})$$



The Optimization Problem

$$y_i(w^T x_i + b) \geq 1 \; \forall x_i \in D$$

$$min_{b,\vec{w}} \frac{1}{2} \|\vec{w}\|^2 \; subject \; to \; y_i(\vec{w} \cdot \vec{x_i} - b) \geq 1 \; \forall x_i \in D$$

A quadratic programming optimization problem…

$$L(\underline{w}, b, \alpha) = \frac{1}{2} \| \underline{w} \|^2 - \sum_{i=1}^{N} \alpha_i (y_i [\underline{w}^T \underline{x}_i + b] - 1)$$

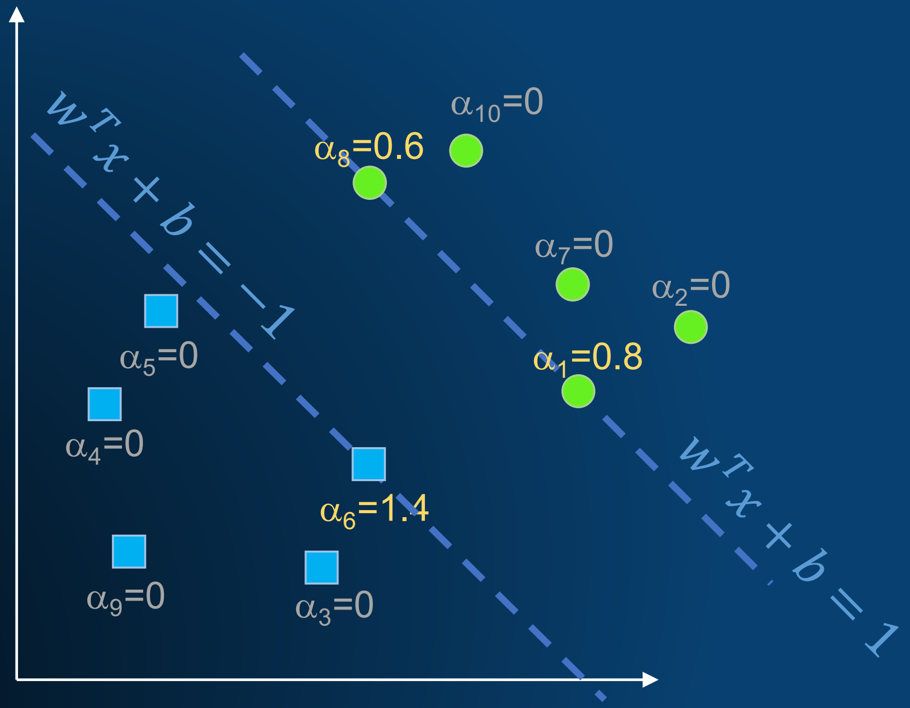
The Optimization Problem

$$y_i(w^T x_i + b) \geq 1 \; \forall x_i \in D$$

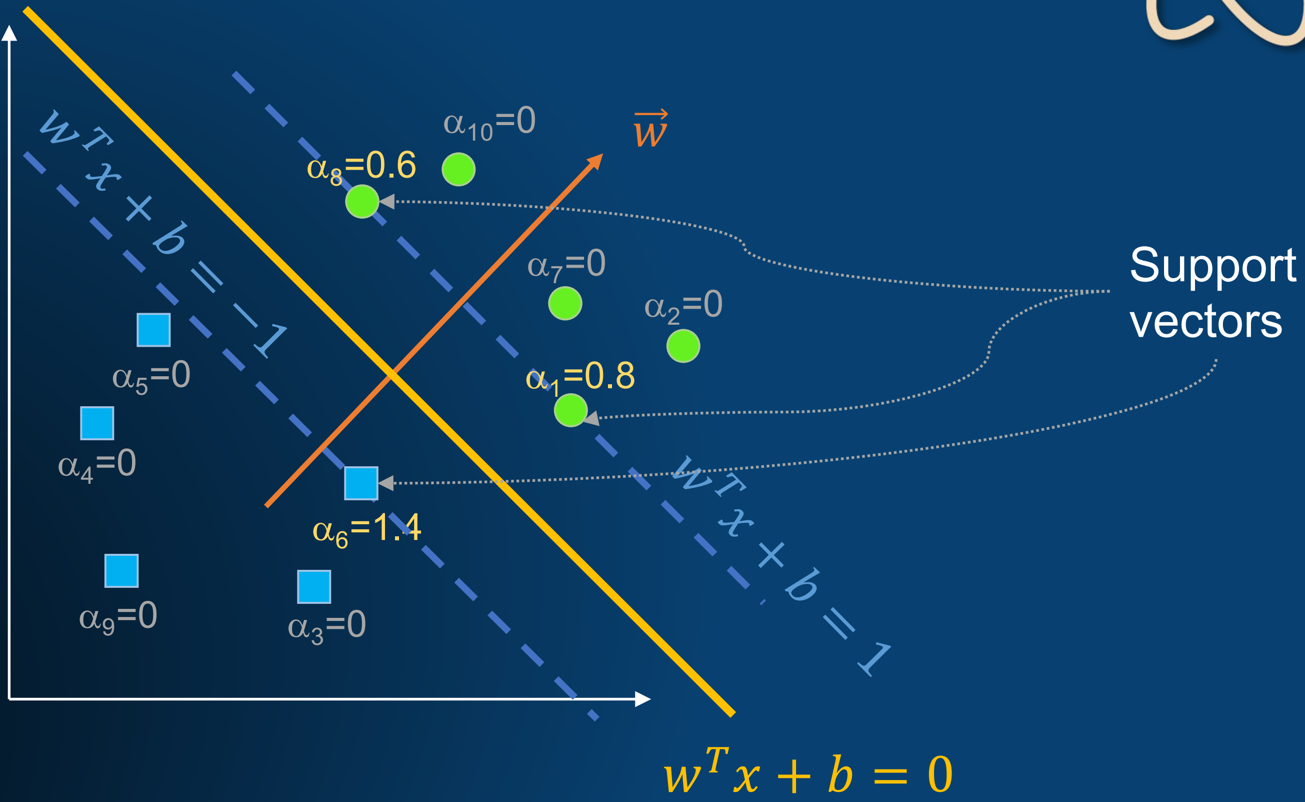
$$min_{b,\vec{w}} \frac{1}{2} ||\vec{w}||^2 \; subject \; to \; y_i(\vec{w} \cdot \vec{x_i} - b) \geq 1 \; \forall x_i \in D$$

$$\boxed{f(x) = sgn(\langle w, x \rangle + \mathbf{b}) = sgn(\sum_{x_i \in D} \alpha_i y_i \langle x, x_i \rangle + \mathbf{b})}$$

A Geometrical Interpretation

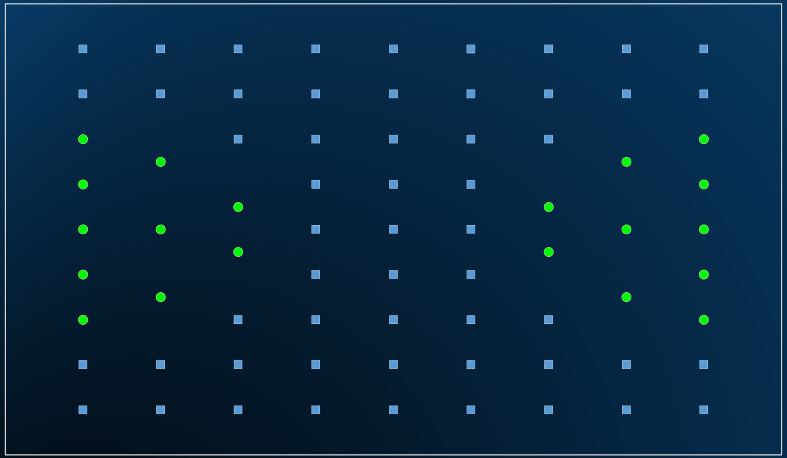


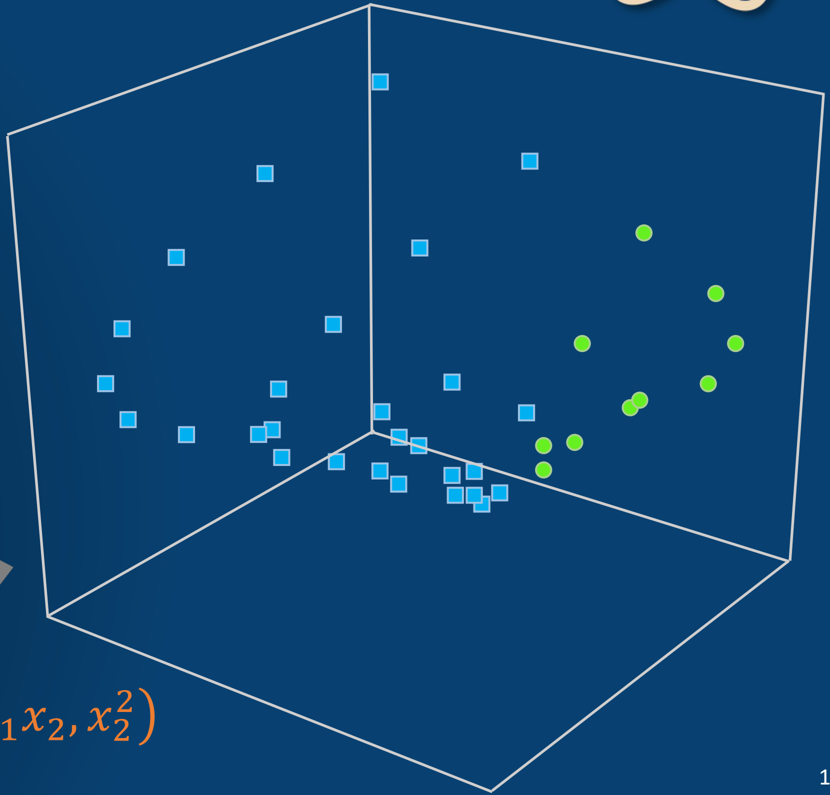
A Geometrical Interpretation



What if data is non-linearly separable?

Change of Space





$$\phi(\vec{x}) = \left(x_1^2, -\sqrt{2}x_1x_2, x_2^2\right)$$

Change of Space

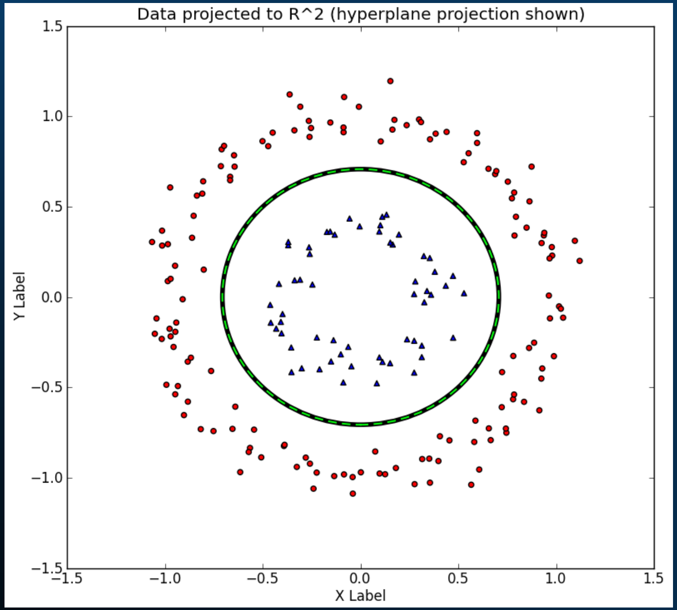
$x, z \in \mathbb{R}^2$

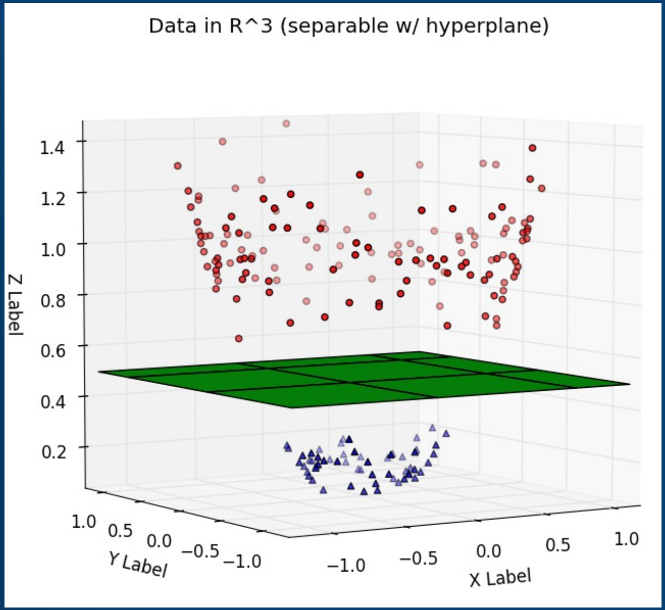
$$
\begin{aligned}
&\langle x, z \rangle^2 \\
&= (x_1 z_1 + x_2 z_2)^2 \\
&= x_1^2 z_1^2 + 2 x_1 z_1 x_2 z_2 + x_2^2 z_2^2 \\
&= \left\langle \left(x_1^2, \sqrt{2} x_1 x_2, x_2^2\right), \left(z_1^2, \sqrt{2} z_1 z_2, z_2^2\right) \right\rangle \\
&= \langle \phi(x), \phi(z) \rangle
\end{aligned}
$$

Kernels

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$$

Polynomial Kernel 2nd degree





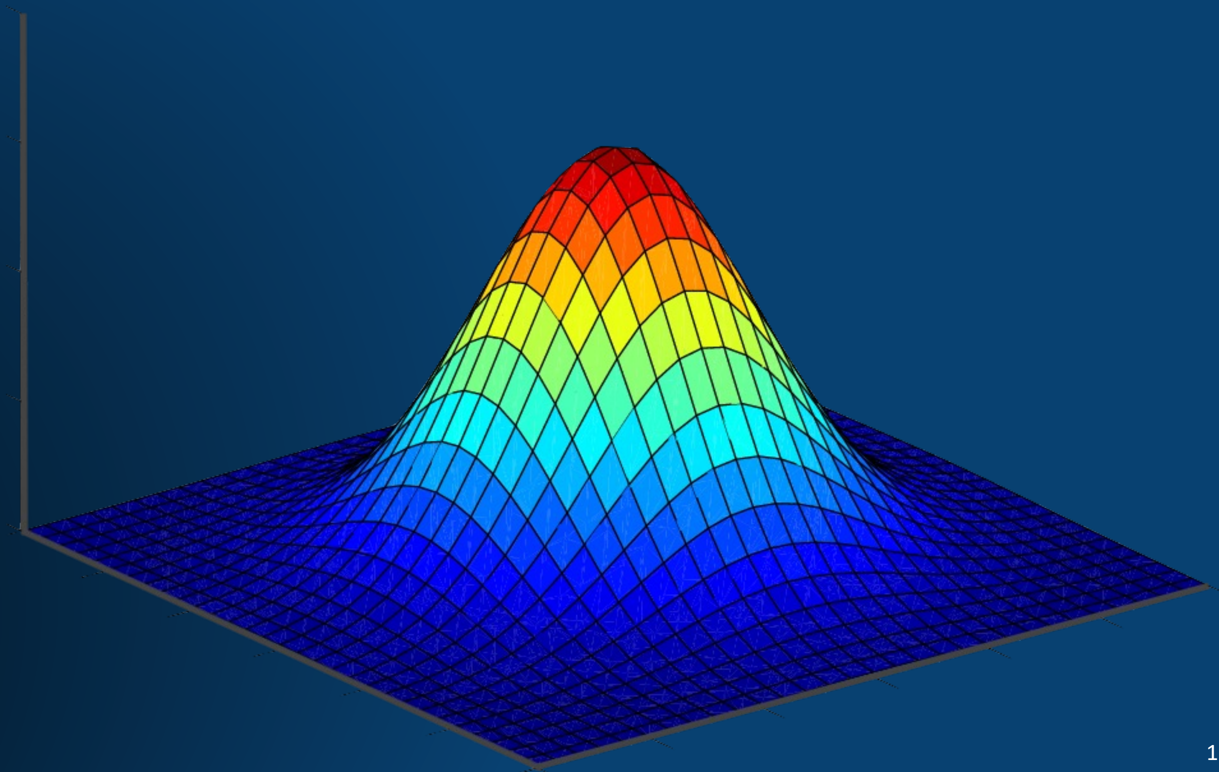
$$K(x_i, x_j) = (x_i^T x_j + c)^d$$

http://www.eric-kim.net/eric-kim-net/posts/1/kernel_trick.html

Gaussian Kernel

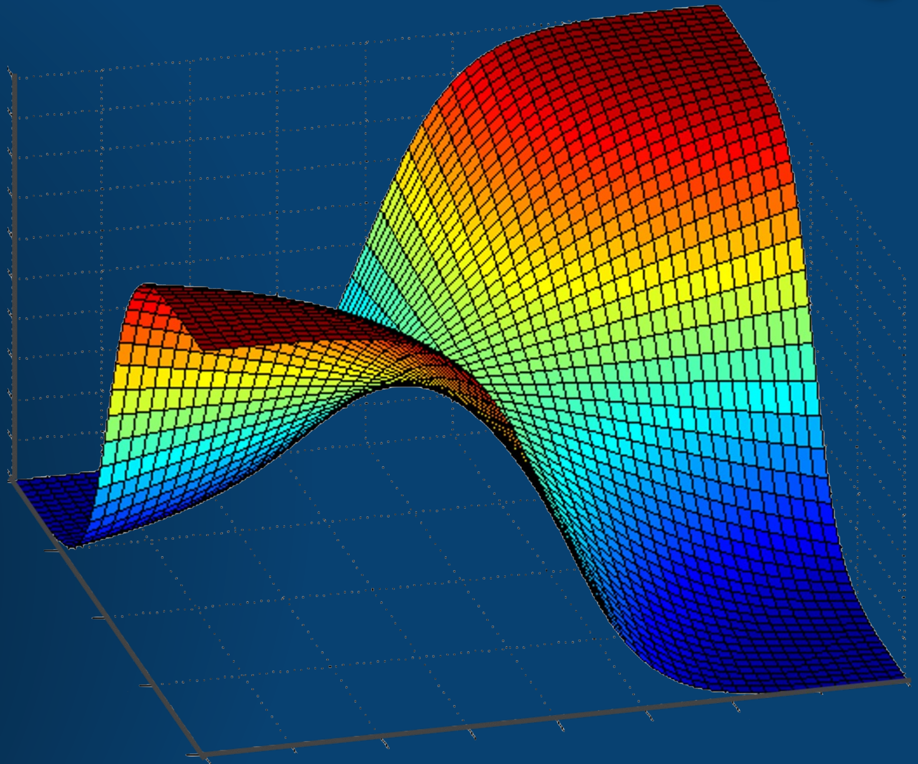
$$K(x_i, x_j) = e^{-\gamma \left\| x_i - x_j \right\|^2}$$

$$\gamma = \frac{1}{2\sigma^2}$$



Sigmoid Kernel

$$K(x_i, x_j) = tanh(ax_i^T x_j + r)$$



Classifier with Kernels

$$f(x) = sgn(\sum_{x_i \in D} \alpha_i y_i \langle x, x_i \rangle + \mathbf{b})$$

$$f(x) = sgn(\sum_{x_i \in D} \alpha_i y_i \langle \phi(x), \phi(x_i) \rangle + \mathbf{b})$$

$$\boxed{f(x) = sgn(\sum_{x_i \in D} \alpha_i y_i K(x, x_i) + \mathbf{b})}$$

Thank you!

claudia.antunes@tecnico.ulisboa.pt