

The background of the slide features a dark blue, high-tech aesthetic. On the left, a human hand is shown in profile, with the index finger pointing towards the center. The background is filled with abstract digital elements: concentric circles, gear-like patterns, and various data visualization elements like bar charts and line graphs, all rendered in shades of teal and light blue. The overall composition suggests a theme of artificial intelligence, machine learning, or advanced technology.

Support Vector Machines

Cláudia Antunes

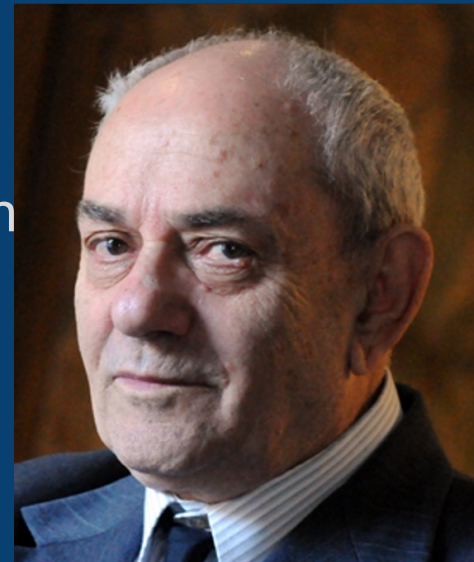
Instituto Superior Técnico – Universidade de Lisboa

LEARNING THEORY



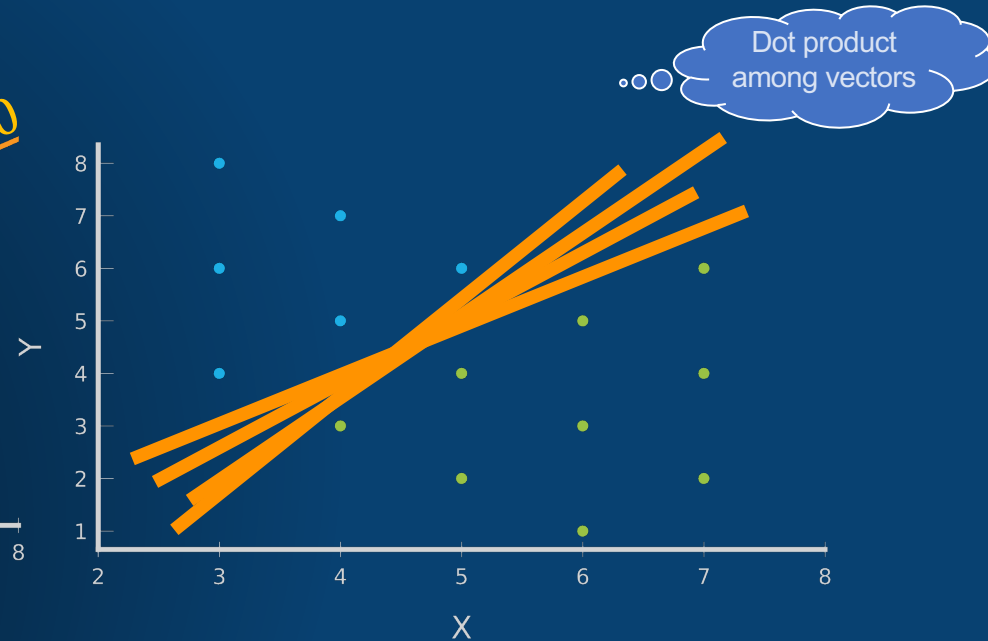
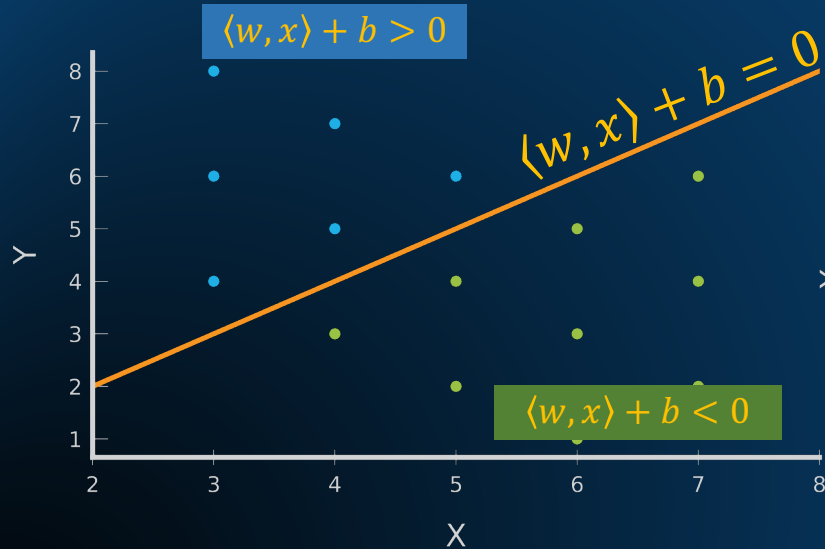
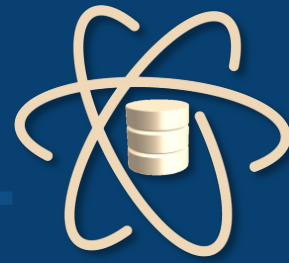
1971 – Vapnik-Chervonenkis dimension and the convergency of the learning process

1995 – Support Vector Machines



Vladimir Vapnik

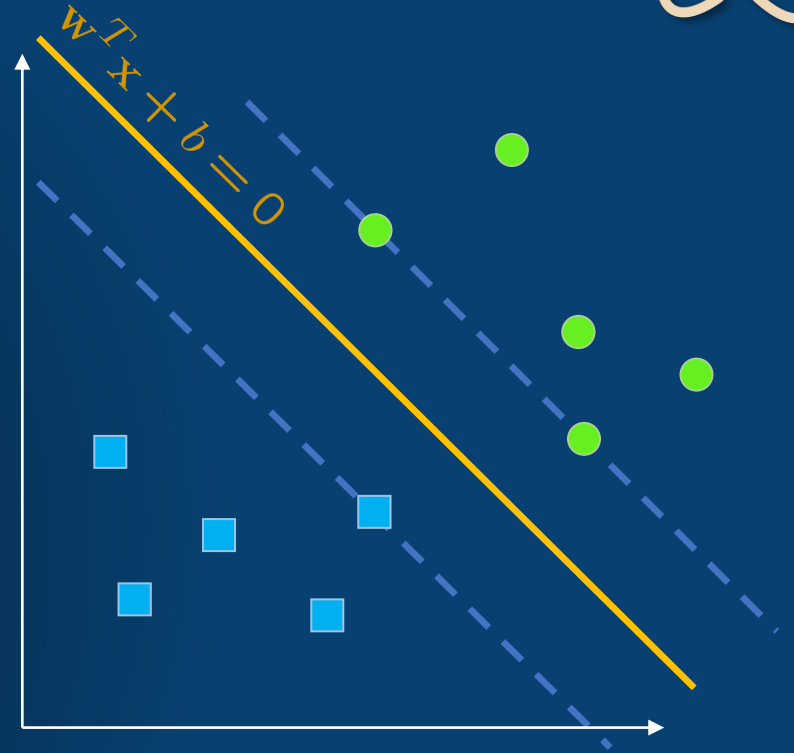
LINEAR SEPARABILITY



MAXIMUM-MARGIN HYPERPLANE



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



THE OPTIMIZATION PROBLEM



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

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

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

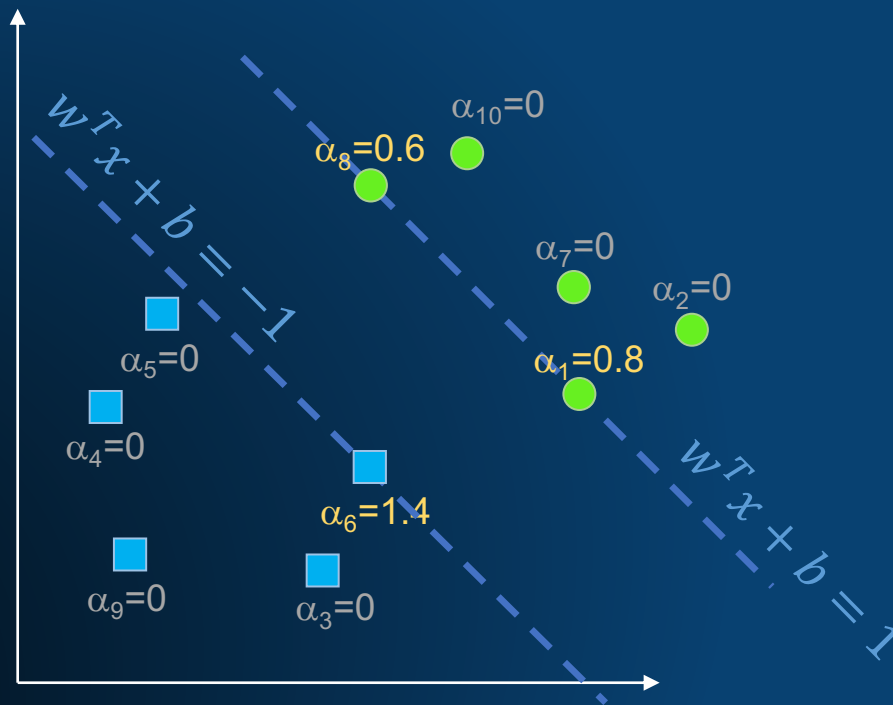


$$y_i(w^T x_i + b) \geq 1 \quad \forall x_i \in D$$

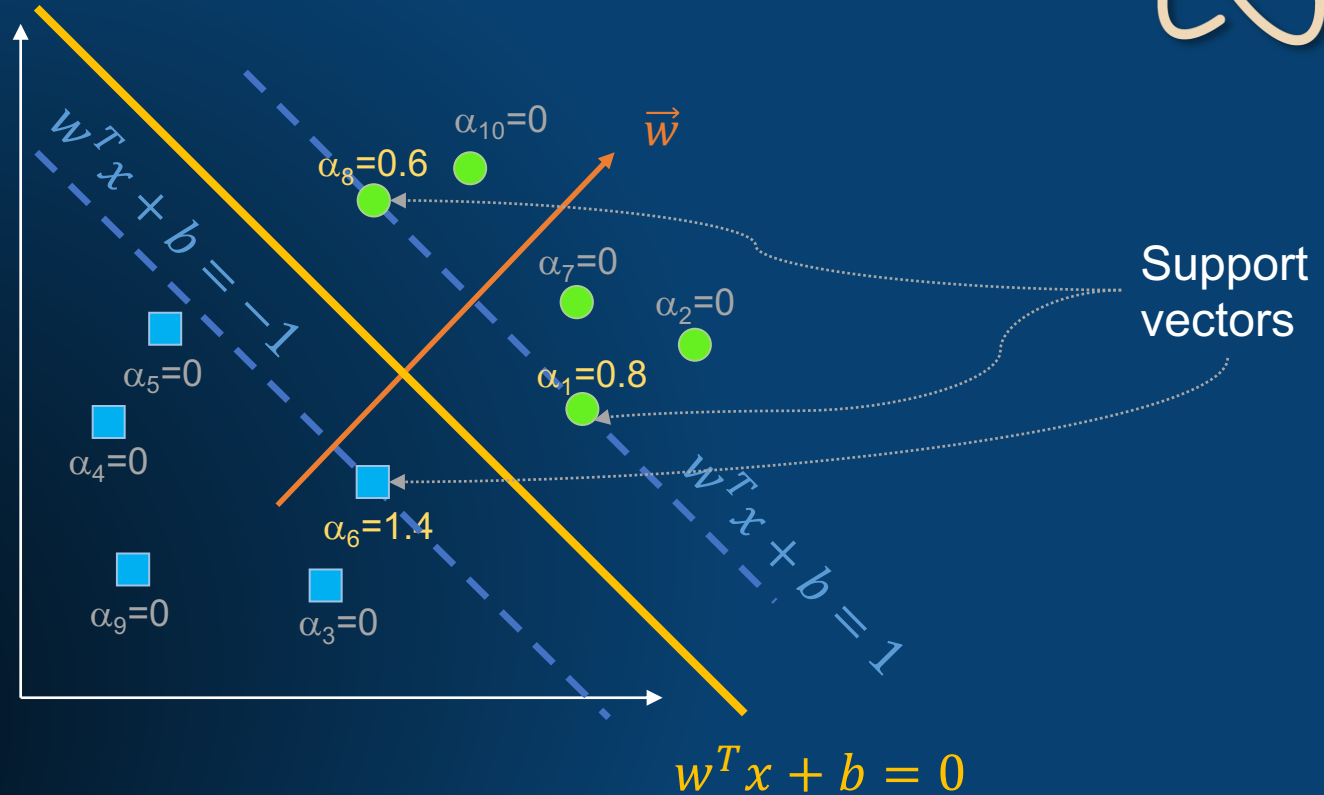
$$\min_{b, \vec{w}} \frac{1}{2} \|\vec{w}\|^2 \text{ subject to } y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1 \quad \forall x_i \in D$$

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

A GEOMETRICAL INTERPRETATION



A GEOMETRICAL INTERPRETATION

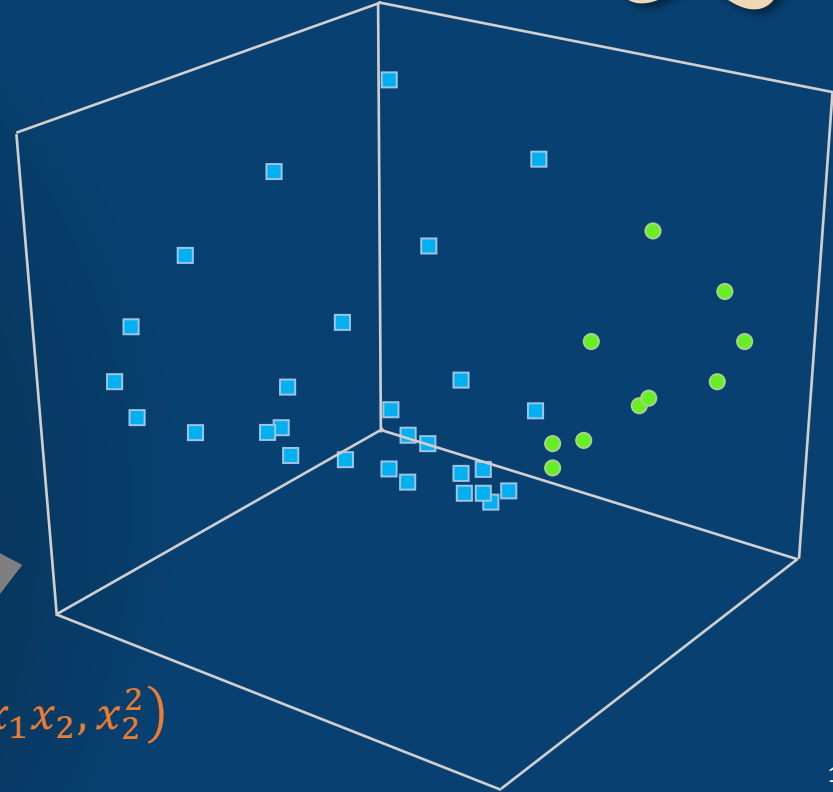
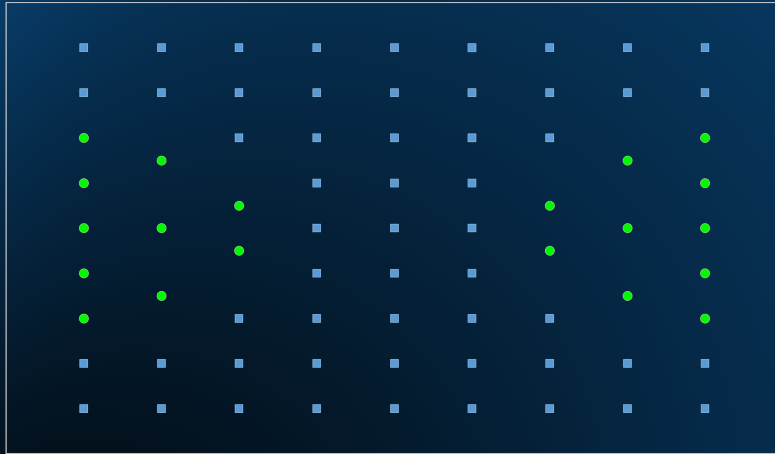






**What if data is
non-linearly separable?**

CHANGE OF SPACE



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

CHANGE OF SPACE



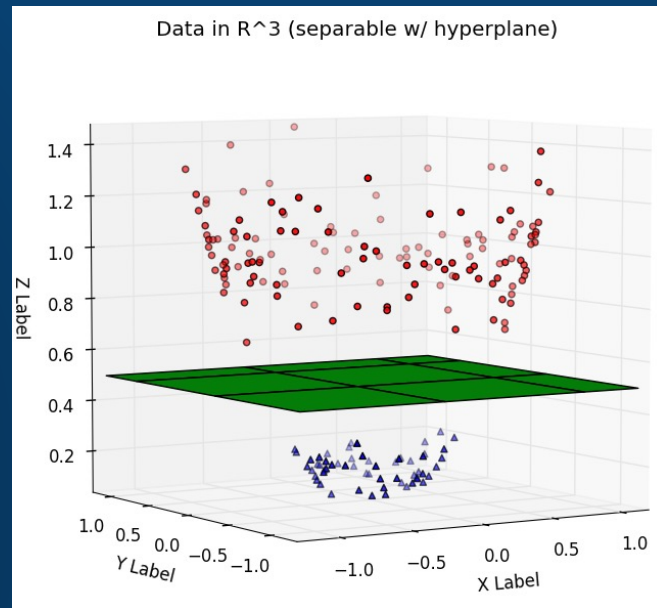
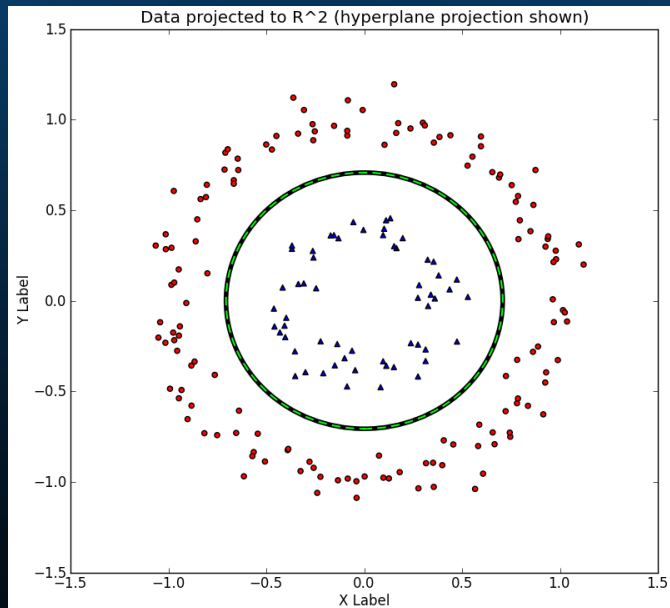
$$x, z \in \mathbb{R}^2$$

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



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

POLYNOMIAL KERNEL 2ND DEGREE



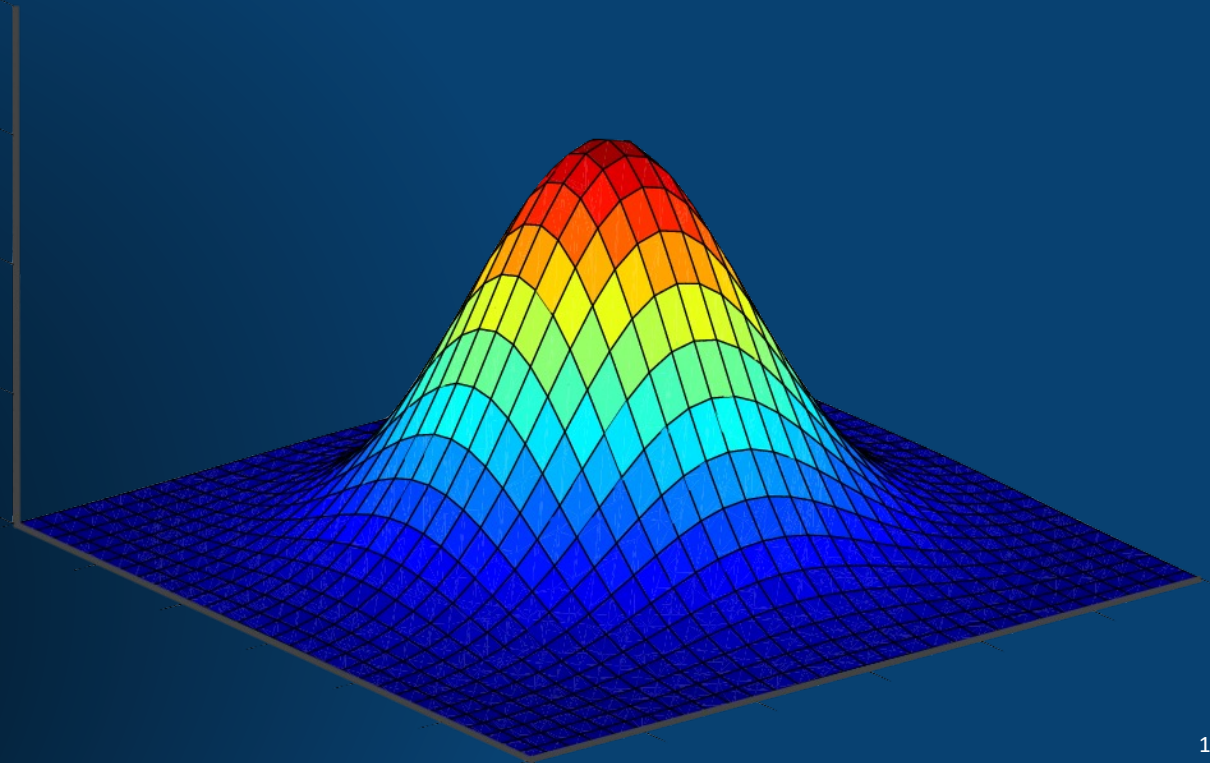
$$K(x_i, x_j) = (x_i^T x_j + c)^d$$

GAUSSIAN KERNEL



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

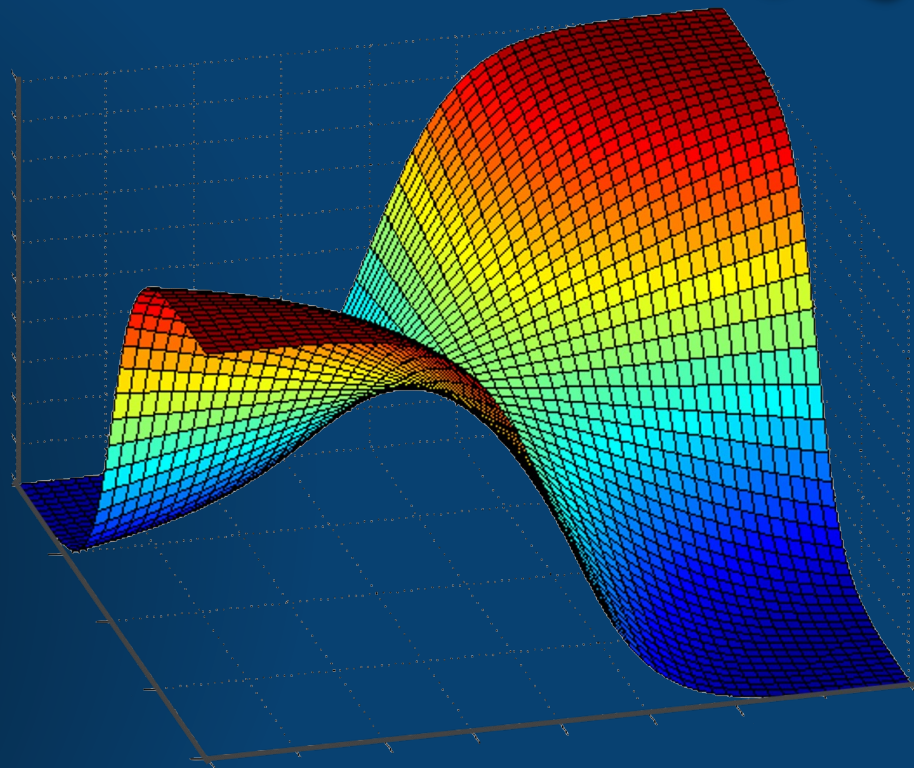
$$\gamma = \frac{1}{2\sigma^2}$$



SIGMOID KERNEL



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



CLASSIFIER WITH KERNELS



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

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

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



*Thank
you!*

