Sistemas Inteligentes

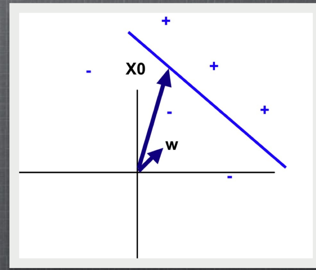
T2. AA 2 / L2 ML 2. SVM

Antonio Bahamonde
Departamento de informática

Universidad de Oviedo en Gijón

SVM: SUPPORT VECTOR MACHINES

Scalar Products
Optimization



Feature maps

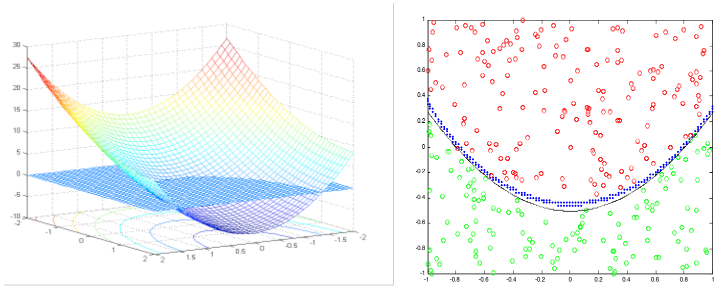
Consider fitting cubic functions $y = \theta_3x^3+ \theta_2x^2+ \theta_1x + \theta_0$. We can view it as a linear function over a different set of feature variables

$$\phi : \mathbb{R} \rightarrow \mathbb{R}^4 \qquad \phi(x) = \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix} \in \mathbb{R}^4 \qquad \theta_3 x^3 + \theta_2 x^2 + \theta_1 x + \theta_0 = \theta^T \phi(x)$$

A *cubic* function of the variable x (**input attributes**) can be viewed as a linear function over the variables $\phi(x)$ (**features variables**)

ϕ is a **feature map**, which maps the attributes to the features.

Polinomial decisions



$$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^p$$

SVM: notations

- $y \in \{-1, 1\}$ (instead of $\{0, 1\}$)
- the parameters w and b instead of the vector θ
- the classifier

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

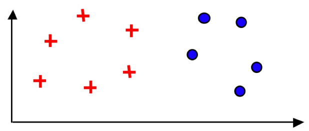
where $g(z) = 1$ if $z \geq 0$, and $g(z) = -1$ otherwise

Scalar products and geometry

Support Vector Machines

- Classifying with scalar products
- Maximizing margin

$S = \{(\mathbf{x_i}, y_i) : \boldsymbol{x_i} \in \mathbf{H},\ y_i \in \{+1, -1\};\ i = 1, \ldots, m\}$



Scalar product

- **Definition.-**

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} = \mathbf{x}^T \mathbf{y}$$

- **Properties.-**

(1) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
(2) $\langle \mathbf{x}, (\mathbf{y}+\mathbf{z}) \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$
(3) $\langle a\,\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, a\,\mathbf{y} \rangle = a \langle \mathbf{x}, \mathbf{y} \rangle$

Scalar product (2)

Gives rise to a norm

$$\|\boldsymbol{x}\|^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle$$

and then a metric (distance)

$$(d_H(\boldsymbol{x}, \boldsymbol{x}'))^2 = \|\boldsymbol{x} - \boldsymbol{x}'\|^2 = \langle \boldsymbol{x} - \boldsymbol{x}', \boldsymbol{x} - \boldsymbol{x}' \rangle$$
$$= \langle \boldsymbol{x}, \boldsymbol{x} \rangle - 2\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + \langle \boldsymbol{x}', \boldsymbol{x}' \rangle$$

Scalar product (3)

Geometric meaning

$$\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^n \qquad \langle \boldsymbol{x}, \boldsymbol{x}' \rangle = \|\boldsymbol{x}\| \cdot \|\boldsymbol{x}'\| \cdot \cos(\boldsymbol{x}, \boldsymbol{x}')$$

if two vectors have norm 1, their scalar product ranges from:

- +1 (same) to
- -1 (opposite).
- 0 (perpendicular)

Scalar product (4)

Geometric meaning

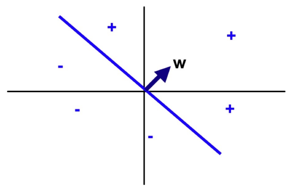
If we do not consider their lengths, the scalar product of 2 vectors measures their **similarity**

$$\langle \boldsymbol{x}, \boldsymbol{x}' \rangle = \|\boldsymbol{x}\| \cdot \|\boldsymbol{x}'\| \cdot \cos(\boldsymbol{x}, \boldsymbol{x}')$$

Hyperplanes and scalar product

In a n-dimensional space, an **hyperplane** is a subspace of dimension n-1.

They are determined by one vector (**director vector**) perpendicular to the hyperplane



Equation:

<x,w> = 0

Splits the space in 2 regions

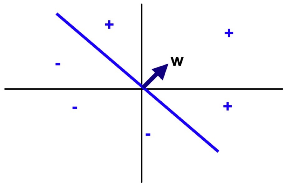
Pos= {**x**: <**x,w**> ≥ **0**}

Neg={**x**: <**x,w**> **< 0**}

Hyperplanes and scalar product

In a n-dimensional space, an **hyperplane** is a subspace of dimension n-1.

They are determined by one vector (**director vector**) perpendicular to the hyperplane



Equation:

$\langle x, w\rangle = 0$

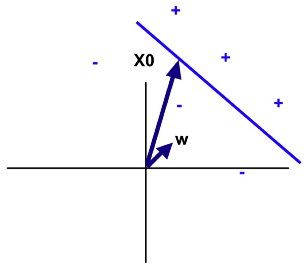
Splits the space in 2 regions

Pos= {**x**: <**x**,**w**> ≥ **0**}

Neg={**x**: <**x**,**w**> < **0**}

Affine hyperplane

Def.- **hyperplanes** translated to a point x_0 in the space



Equation:

$\langle x-x_0, w\rangle = \langle x, w\rangle + b = 0$

Splits the space in 2 regions

Pos= {**x**: $\langle x-x_0, w\rangle \geq 0$}

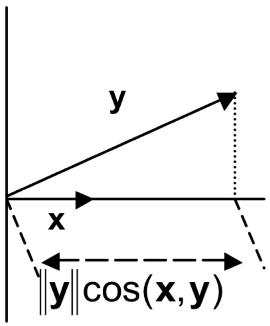
Neg= {**x**: $\langle x-x_0, w\rangle < 0$}

Distance to hyperplanes

$$if \quad \|x\| = 1, \quad then \quad \langle x, y\rangle = \|y\| cos(x,y)$$

The scalar product is

- the length of the projection of **y** in the direction of **x**
- the distance of ***y*** to the hyperplane determined by(perpendicular to) ***x***



Distance to affine hyperplanes

The **distance** from **x** ∈ H to an hyperplane

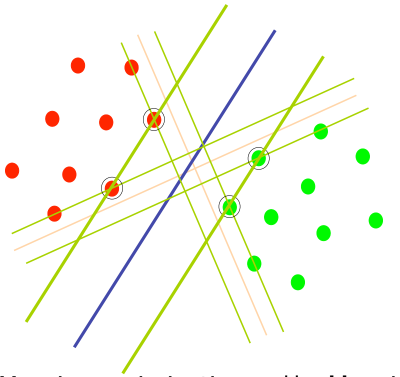
<w,x> + b = 0

$$d(x, Hyperplane) = \frac{\langle w, x\rangle + b}{\|w\|}$$

The **split** in two regions can be obtained by

y = sign (<***w***, ***x***> + b) = sign(d(***x***, Hyperplane))

The optimization of SVM

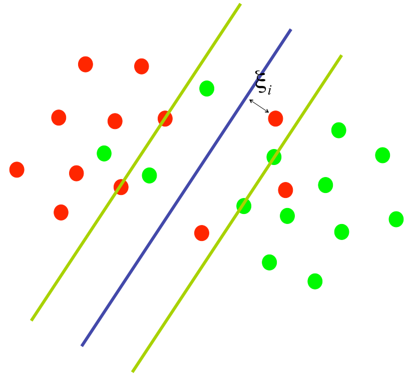


margin $= \frac{1}{\|\mathbf{w}\|}$

Margin maximization = ||**w**|| minimization

SVM in non separable datasets

Soft margin



SVM: convex optimization

To find the optimal soft-margin hyperplane =
Solving a quadratic convex optimization problem subject to some constraints

$$\begin{aligned} \min_{\mathbf{w},\xi,b} \quad & \frac{1}{2}\langle \mathbf{w}, \mathbf{w} \rangle + C\sum_{i=1}^{m} \xi_i, \\ \text{s.t.} \quad & y_i(\langle \mathbf{w}, \phi(\mathbf{x}_i)\rangle + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \qquad i = 1, \ldots, m \end{aligned}$$

SVM: convex optimization

Solving a quadratic convex optimization:

- There are *very* fast optimizers

- Convex means:
  - there exist one global optimum and
  - no local optima (this is not the case in ANN)

Métodos Kernel y Máquinas de Vectores Soporte

Juan José del Coz Velasco Oscar Luaces Rodríguez