Reading Club: Support Vector Machines

Christian Brosch

24. Mai 2006

The XOR Example

Input vector $\vec{x}$	Desired response d
$(-1, -1)$	-1
$(-1, +1)$	$+1$
$(+1, -1)$	$+1$
$(+1, +1)$	-1

Kernel

$$K(\vec{x}, \vec{x}_i) = (1 + \vec{x}^T \vec{x}_i)^2$$

with $\vec{x} = [x_1, x_2]^T$ and $\vec{x}_i = [x_{i1}, x_{i2}]^T$

Thus, the inner-product kernel can be expressed as

$$K(\vec{x}, \vec{x}_i) = 1 + x_1^2 x_{i1}^2 + 2x_1 x_2 x_{i1} x_{i2} + x_2^2 x_{i2}^2 + 2x_1 x_{i1} + 2x_2 x_{i2}$$

With $K(\vec{x}, \vec{x}_i) = \vec{\phi}(\vec{x})\vec{\phi}(\vec{x}_i)$ one gets

$$\vec{\phi}(\vec{x}) = [1, x_1^2, \sqrt{2}x_1 x_2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2]^T$$

and similarly

$$\vec{\phi}(\vec{x}_i) = [1, x_{i1}^2, \sqrt{2}x_{i1} x_{i2}, x_{i2}^2, \sqrt{2}x_{i1}, \sqrt{2}x_{i2}]^T$$

Gram Matrix

- for the input vectors $\vec{x}_i, i = 1, \ldots, 4$ we get the Gram Matrix

$$\mathbf{K} = \begin{pmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{pmatrix}$$

- The dual problem $Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j K(x_i, x_j)$ therefore evaluates to

$$Q(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}(9\alpha_1^2 - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_3 + 2\alpha_1\alpha_4 + 9\alpha_2^2$$
$$+2\alpha_2\alpha_3 - 2\alpha_2\alpha_4 + 9\alpha_3^2 - 2\alpha_2\alpha_4 + 9\alpha_4^2$$

Optimization

Optimizing $Q(\alpha)$ with respect to the Lagrange multipliers yields the following set of simultaneous equations:

$$\begin{aligned} 9\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 &= 1 \\ -\alpha_1 + 9\alpha_2 + \alpha_3 - \alpha_4 &= 1 \\ -\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4 &= 1 \\ \alpha_1 - \alpha_2 - \alpha_3 + 9\alpha_4 &= 1 \end{aligned}$$

From dual solution to optimum weight vector

- Solution of the simultaneous equations:

$$\alpha_{o,1} = \alpha_{o,2} = \alpha_{o,3} = \alpha_{o,4} = \frac{1}{8}$$

- all four input vectors are support vectors!

- The optimum value of $Q(\alpha)$ is:

$$Q_o(\alpha) = \frac{1}{4}$$

Duality Theorem

- If the primal problem has an optimal solution, the dual problem also has an optimal solution, and the corresponding optimal values are equal

- Hence, we can write

$$\frac{1}{2}\|\vec{w}_o\|^2 = \frac{1}{4}$$

or

$$\|\vec{w}_o\| = \frac{1}{\sqrt{2}}$$

Getting the optimum weight vector (1)

- Remember: With

$$\frac{\partial J(\vec{w}, b, \alpha)}{\partial \vec{w}} = 0$$

  we got

$$\vec{w} = \sum_{i=1}^{N} \alpha_i d_i \vec{x}_i$$

- Now we have to do it with the optimum values of α_i and with the image of $\vec{x}_i$ in the feature space, so we get

$$\vec{w}_o = \sum_{i=1}^{N} \alpha_{o,i} d_i \vec{\phi}(\vec{x}_i)$$

Getting the optimum weight vector (2)

$$\vec{w}_o = \frac{1}{8}[-\vec{\phi}(\vec{x}_1) + \vec{\phi}(\vec{x}_2) + \vec{\phi}(\vec{x}_3) - \vec{\phi}(\vec{x}_4)]$$

$$= \frac{1}{8}\left[-\begin{bmatrix} 1 \\ 1 \\ \sqrt{2} \\ 1 \\ -\sqrt{2} \\ -\sqrt{2} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ -\sqrt{2} \\ 1 \\ -\sqrt{2} \\ \sqrt{2} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ -\sqrt{2} \\ 1 \\ \sqrt{2} \\ -\sqrt{2} \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ \sqrt{2} \\ 1 \\ \sqrt{2} \\ \sqrt{2} \end{bmatrix} \right] = \begin{bmatrix} 0 \\ 0 \\ \frac{-1}{\sqrt{2}} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Getting the optimal hyperplane

- The optimal hyperplane is defined by

$$\vec{w}_o^T \vec{\phi}(\vec{x}) = 0$$

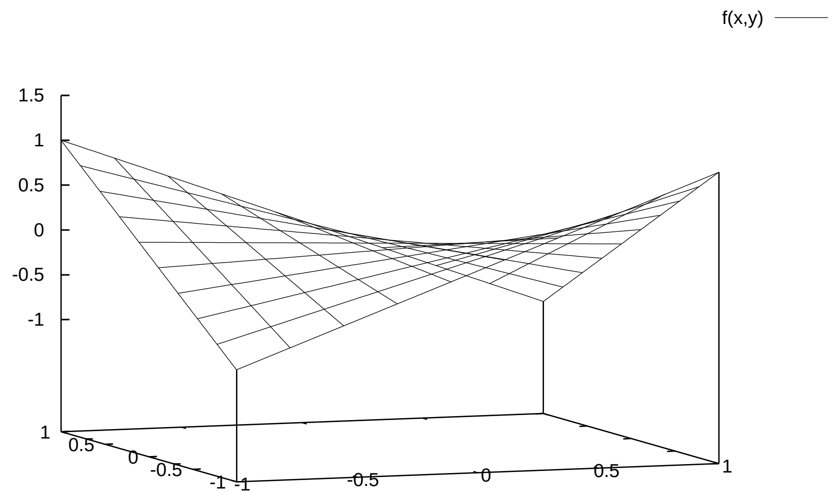
- That is,

$$\left[0, 0, \frac{-1}{\sqrt{2}}, 0, 0, 0\right] \begin{bmatrix} 1 \\ x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \end{bmatrix} = 0$$

- which reduces to

$$-x_1x_2 = 0$$

Decision Surface



Polynomial form of SVM for XOR example