Support Vector Classification

The kernel trick is to substitute $k(\mathbf{x}, \mathbf{x}_i)$ in place of $(\mathbf{x} \cdot \mathbf{x}_i)$. The decision function becomes:

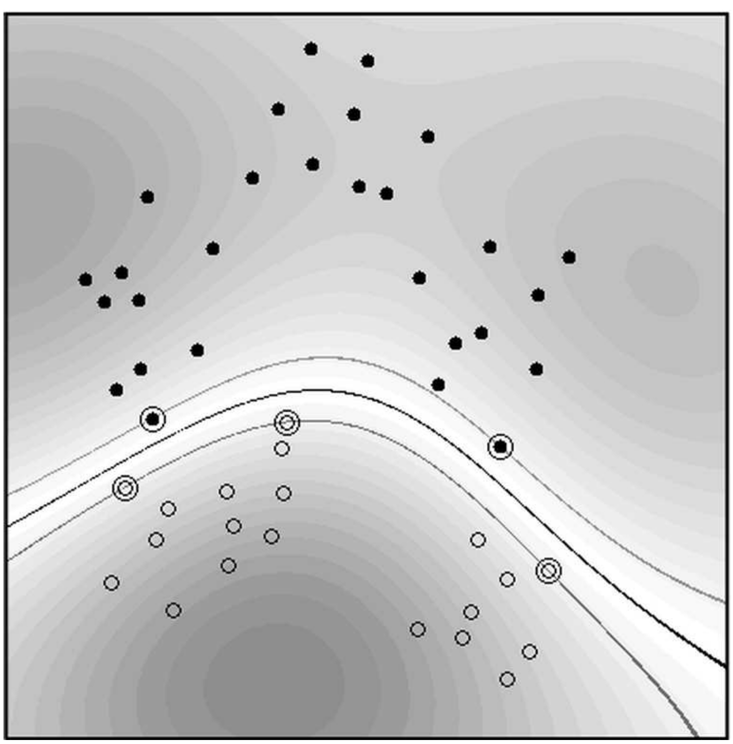
$$f_a(\mathbf{x}) = \operatorname{sign}\left(b + \sum_{i=1}^{m} y_i \alpha_i k(\mathbf{x}, \mathbf{x}_i)\right)$$

The α_i weights are found by solving:

maximize $\sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$

subject to $\alpha_i \geq 0$ and $\sum_{i=1}^{m} \alpha_i y_i = 0$

The support vectors are $\{\mathbf{x}_i \mid \alpha_i > 0\}$.
They satisfy $y_i f_a(\mathbf{x}_i) \leq 1$.



Soft Margin Classification

It might not be possible or desirable to satisfy:

$$y_i\left((\mathbf{w}\cdot\mathbf{x}_i)+b\right)\geq 1$$

To allow violations, an error term can be added:

$$y_i\left((\mathbf{w}\cdot\mathbf{x}_i)+b\right)\geq 1-\xi_i$$

where $\xi_i \geq 0$ and the problem is to:

$$\text{minimize } \|\mathbf{w}\|^2/2 + C\sum_{i=1}^{m}\xi_i$$

where C is chosen by the user. In the end, the only change is $0 \leq \alpha_i \leq C$

ν-Parameterization

Another type of soft margin classifier satisfies:

$$y_i\left((\mathbf{w}\cdot\mathbf{x}_i)+b\right)\geq \rho-\xi_i$$

where $\rho > 0$ is a free parameter, and solves:

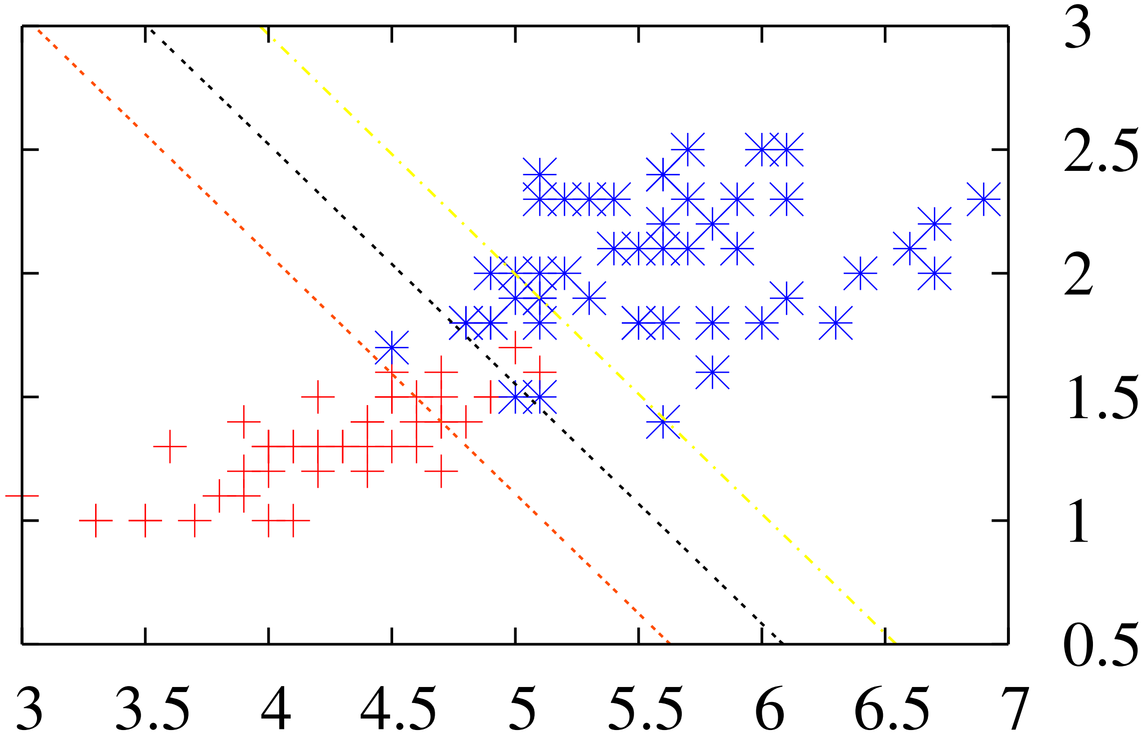
$$\text{minimize } \sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j k(\mathbf{x}_i,\mathbf{x}_j)$$

$$\text{subject to } 0 \leq \alpha_i \leq 1/(\nu m),$$

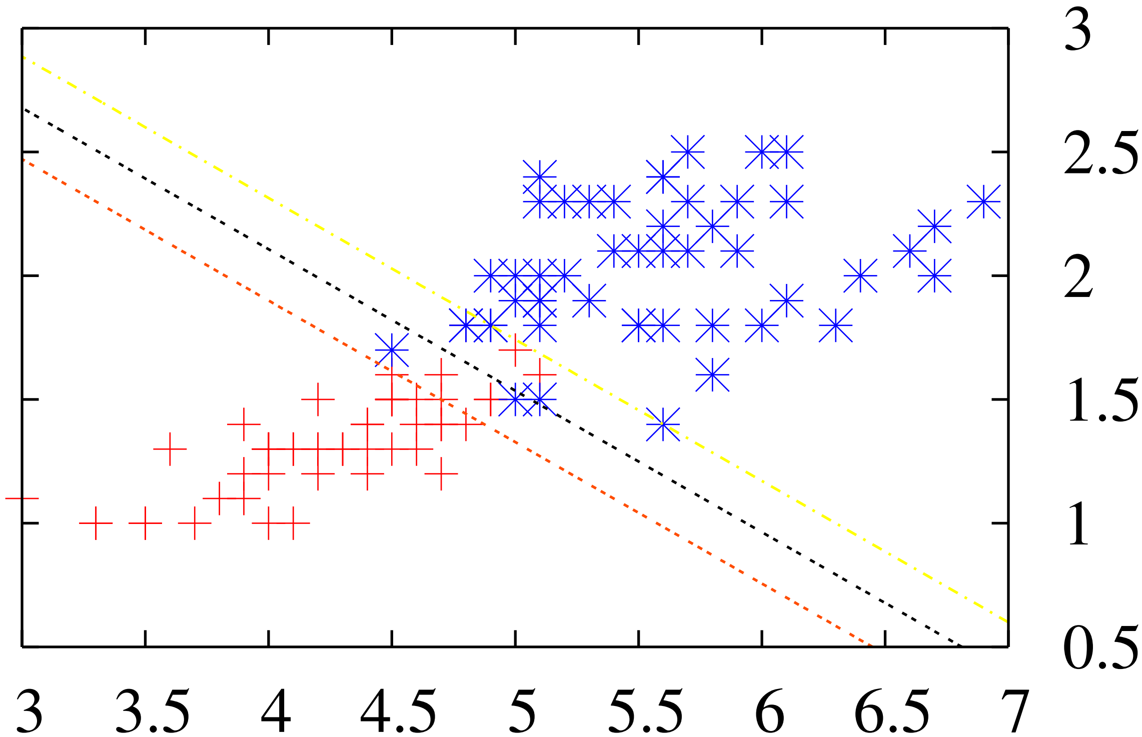
$$\sum_{i=1}^{m}\alpha_i y_i = 0, \text{ and } \sum_{i=1}^{m}\alpha_i = 1$$

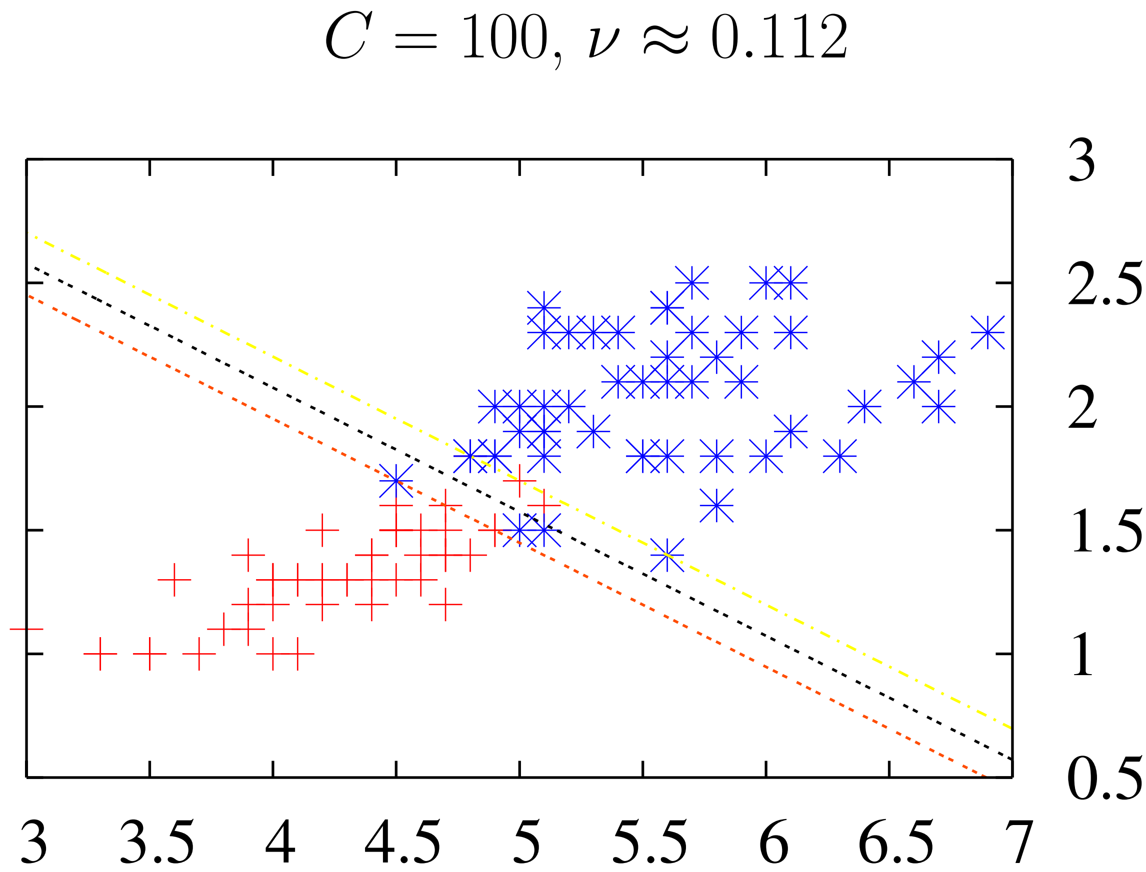
$\nu \in (0,1)$ means at least $m\nu$ support vectors.

$C = 1,\ \nu \approx 0.23$



$C = 10,\ \nu \approx 0.140$





$C = 100,\ \nu \approx 0.112$
3
2.5
2
1.5
1
0.5
3
3.5
4
4.5
5
5.5
6
6.5
7