

Support Vector Machines cheatsheet

By J  r  my Fix, Herv   Frezza-Buet

Kernels

The separator reads

$$h_{w,b}(x) = \sum_i w_i k(x_i, x) + b$$

Linear	$k(x, x') = \langle x, x' \rangle$
RBF	$k_\sigma(x, x') = \exp(-\frac{\ x - x'\ ^2}{2\sigma^2})$
Polynomial	$k_{\gamma,c,d}(x, x') = (\gamma \langle x, x' \rangle + c)^d$

Classification SVM : C-SVC

Primal optimization problem:

$$\begin{aligned} & \underset{w,b,\xi}{\operatorname{argmin}} \frac{1}{2} \|w\|^2 + C \mathbf{1}^T \xi \\ & \text{subject to } \begin{cases} y_i h_{w,b}(x_i) \geq 1 - \xi_i & , \forall (x_i, y_i) \in S \\ \xi_i \geq 0 & , \forall i \end{cases} \end{aligned}$$

Dual optimization problem:

$$\begin{aligned} & \underset{\alpha}{\operatorname{argmin}} \frac{1}{2} \alpha^T Q \alpha - \mathbf{1}^T \alpha \\ & \text{subject to } \begin{cases} y^T \alpha = 0 \\ 0 \leq \alpha_i \leq C, \forall i \end{cases} \end{aligned}$$

$$Q_{i,j} = y_i y_j k(x_i, x_j)$$

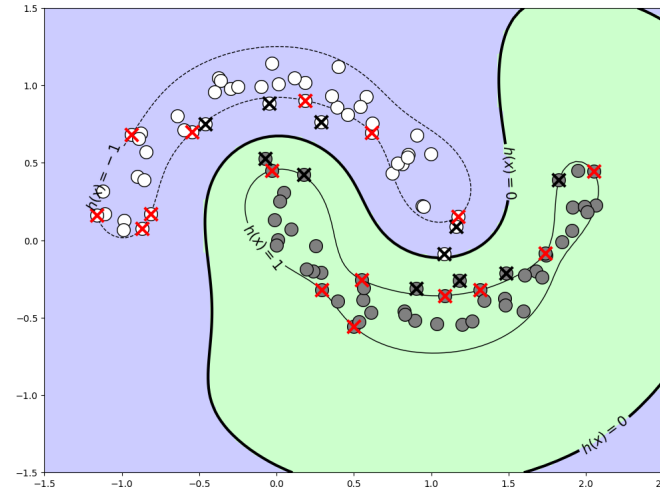
The optimal separator reads :

$$h(x) = \sum_i \alpha_i y_i k(x_i, x) + b$$

For the optimal separator, we have either :

- $\alpha_i = 0$ for **non support vector** x_i , the associated sample does not contribute to the definition of the separating hyperplane,
- $0 < \alpha_i < C$ for **support vectors** x_i just on the margin. These samples are associated with $\xi_i = 0$ and $y_i h_{w,b}(x_i) = 1$
- $\alpha_i = C$ for **support vectors** x_i which involve non null slack variables; they are "on the wrong side" of the margin. These samples are associated with $\xi_i > 0$ and $y_i h_{w,b}(x_i) = 1 - \xi_i$

On the figure below, we see a bi-class problem solved with a C-SVC with $C=1$. The positive samples are drawn with a white dot; The negative samples with a gray dot.



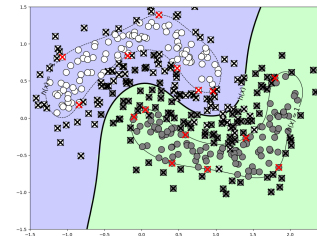
The support vectors are the samples indicated with a cross. A red cross indicates the samples on the margin. A black cross indicates the samples with non null slack variables ξ . Only the support vectors (i.e. the samples indicated with a cross) contribute to the definition of the separator.

C-SVC examples

400 samples, noise=0.25

$C = 0.1, \sigma = 0.4$

With a small C, a lot of vectors can be support vectors



100 samples, noise=0.1

$C = 1000, \sigma = 0.4$

It costs a lot to add slack variables. The separator is really stuck to the data.

