Support Vector Machines cheatsheet

By Jérémy Fix, Hervé Frezza-Buet

Kernels

The separator reads

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

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

Classification SVM : C-SVC

Primal optimization problem:

$$\underset{w,b\xi}{\text{argmin}} \frac{1}{2}\|w\|^2 + C\mathbf{1}^T\xi$$

$$\text{subject to} \begin{cases} y_i h_{w,b}(x_i) \geq 1 - \xi_i & ,\forall(x_i,y_i) \in S \\ \xi_i \geq 0 & ,\forall i \end{cases}$$

Dual optimization problem:

$$\underset{\alpha}{\text{argmin}} \frac{1}{2}\alpha^T Q \alpha - \mathbf{1}^T\alpha$$

$$\text{subject to} \begin{cases} y^T\alpha = 0 \\ 0 \leq \alpha_i \leq C, \forall i \end{cases}$$

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

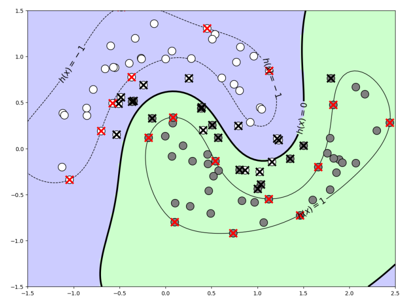
The optimal separator reads :

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

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 $y_i h_{w,b}(x_i) = 1 - \xi_i$

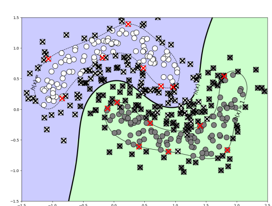
On the figure below, we see a bi-class problem solved with a C-SVC with C=1. The positive samples are drawn with a white dot; The negative samples with a gray dot. 100 samples were used for training.



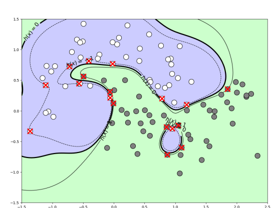
The support vectors are the samples indicated with a cross. A red cross indicates the samples on the margin ($\xi_i = 0$). A black cross indicates the samples with non null slack variables ($\xi_i > 0$). Only the support vectors (i.e. the samples indicated with a cross) contribute to the definition of 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=0.25
$C = 1000, \sigma = 0.4$
It costs a lot to add slack variables. The separator is really stuck to the data.



Classification SVM : ν-SVC

Primal optimization problem:

$$\underset{w,b,\xi,\rho}{\text{argmin}} \frac{1}{2}\|w\|^2 - \nu\rho + \frac{1}{|S|}\mathbf{1}^T\xi$$

$$\text{subject to} \begin{cases} y_i h_{w,b}(x_i) \geq \rho - \xi_i & ,\forall(x_i,y_i) \in S \\ \rho \geq 0 & \\ \xi_i \geq 0 & ,\forall i \end{cases}$$

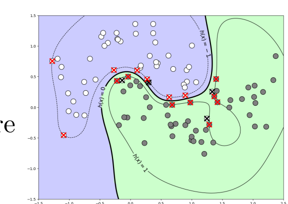
Dual optimization problem:

$$\underset{w,b,\xi,\rho}{\text{argmin}} \frac{1}{2}\alpha^T Q \alpha$$

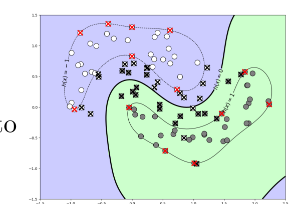
$$\text{subject to} \begin{cases} y^T\alpha = 0 \\ 0 \leq \alpha_i \leq \frac{1}{|S|}, \quad \forall i \\ \mathbf{1}^T\alpha = \nu \end{cases}$$

ν-SVC examples

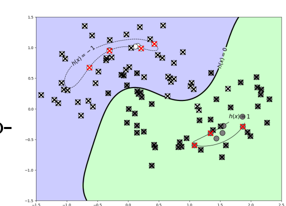
100 samples, noise=0.25
$\nu = 1000, \sigma = 0.4$
Only 10% of the samples are support vectors.



100 samples, noise=0.25
$\nu = 20\%, \sigma = 0.4$
More samples contribute to the decision boundary..



100 samples, noise=0.25
$\nu = 90\%, \sigma = 0.4$
Most of the samples are support vectors.



100 samples, noise=0.25

$\nu = 20\%, \sigma = 0.2$

Only 20% of the samples are support vectors but the σ is so small that the decision wiggles around the samples.