Machine Learning

Exercise

Fabio Vandin

January 9, 2024

Exercise

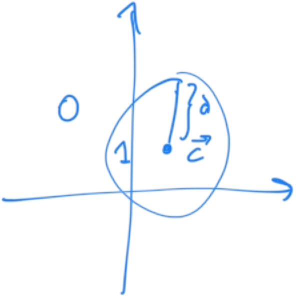
Consider the classification problem with $\mathcal{X} = \mathbb{R}^2, \mathbb{Y} = \{0, 1\}$.
Consider the hypothesis class $\mathcal{H} = \{h_{(\mathbf{c},a)}, \mathbf{c} \in \mathbb{R}^2, a \in \mathbb{R}\}$ with

$$h_{(\mathbf{c},a)}(\mathbf{x}) = \begin{cases} 1 & \text{if } ||\mathbf{x} - \mathbf{c}|| \leq a \\ 0 & \text{otherwise} \end{cases}$$

Find the VC-dimension of $\mathcal{H}$.

In $\mathbb{R}^2$

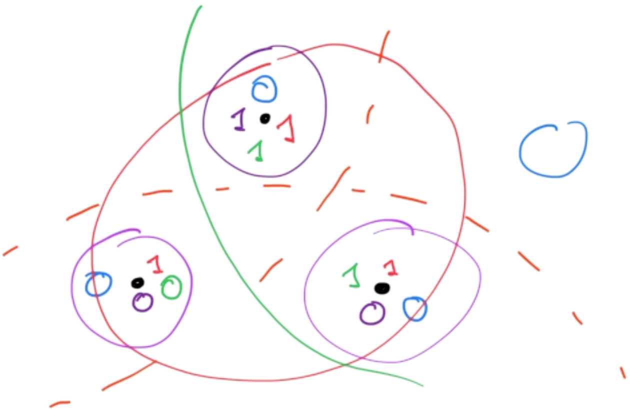
$h_{(\vec{c}, a)}$



Solution

$\mathrm{VCdim}(\mathcal{H}) = 3$

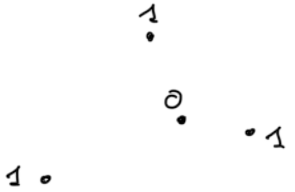
i) $\mathrm{VCdim}(\mathcal{H}) \geq 3$: we need to show a set of 3 instances (vectors in $\mathbb{R}^2$) that is shattered by $\mathcal{H}$



ii) $VC\dim(\mathcal{H}) \le 3$: we need to show that there is no set of 4 instances that can be shattered by $\mathcal{H}$.

Consider an arbitrary set of 4 instances. Then there are 3 cases:

i) 3 instances constitute a triangle and 4th instance is inside the triangle



$\Rightarrow$ impossible to obtain from $\mathcal{H}$

$\Rightarrow$ the set cannot be shattered

ii) 3 instances constitute a triangle and 4th instance is outside the triangle

Assign label 1 to instances on the "longest diagonal" and 0 to the other instances.

$\Rightarrow$ impossible to obtain

$\Rightarrow$ the set cannot be shattered

iii) 

$\Rightarrow$ impossible to obtain

$\Rightarrow$ the set cannot be shattered

$$\Rightarrow VC\dim(\mathcal{H}) \leq 3$$

Therefore $VC\dim(\mathcal{H}) = 3$

Exercise

Assuming we have the following dataset ($x_i \in \mathbb{R}^2$) and by solving the SVM for classification we get the corresponding optimal dual variables:

i	x_i^T	y_i	α_i^*
1	[0.2 -1.4]	-1	0
2	[-2.1 1.7]	1	0
3	[0.9 1]	1	0.5
4	[-1 -3.1]	-1	0
5	[-0.2 -1]	-1	0.25
6	[-0.2 1.3]	1	0
7	[2.0 -1]	-1	0.25
8	[0.5 2.1]	1	0

Answer to the following:

(A) Which are the support vectors?

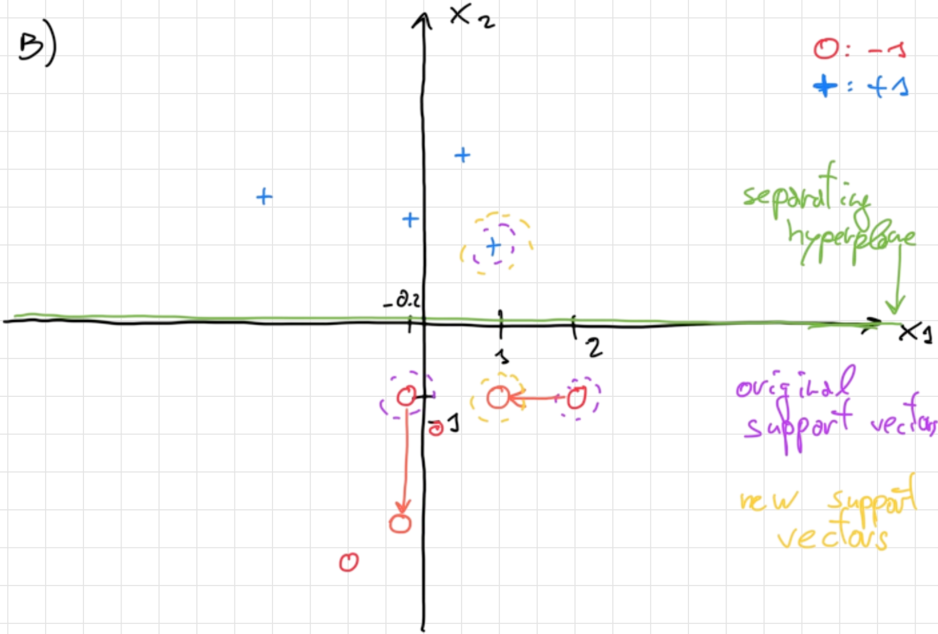
(B) Draw a schematic picture reporting the data points (approximately) and the optimal separating hyperplane, and mark the support vectors. Would it be possible, by moving only two data points, to obtain the SAME separating hyperplane with only 2 support vectors? If so, draw the modified configuration (approximately).

Solution

1) The support vectors are the ones for which $\alpha_i^* \neq 0$, that is:

$$[0.9, 1]^T, [-0.2, -1]^T, [2.0, -1]^T$$

B)

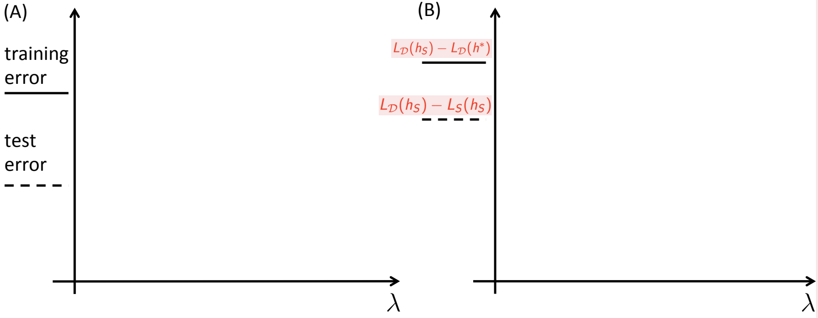

x2
x1
O: -1
+: +1
-0.2
1
2
-1
separating hyperplane
original support vectors
new support vectors

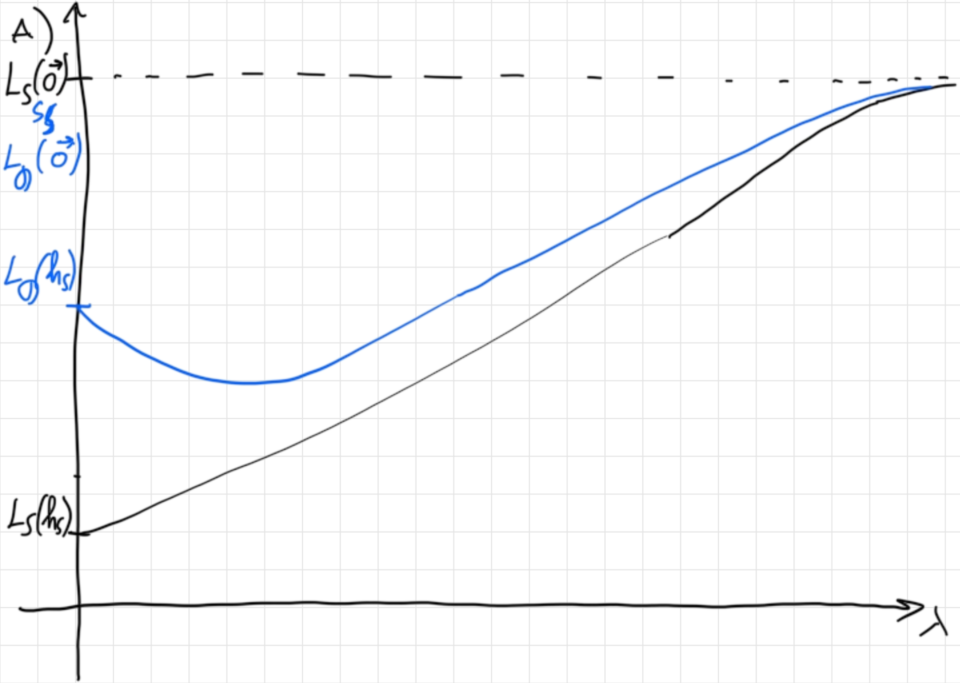
Exercise

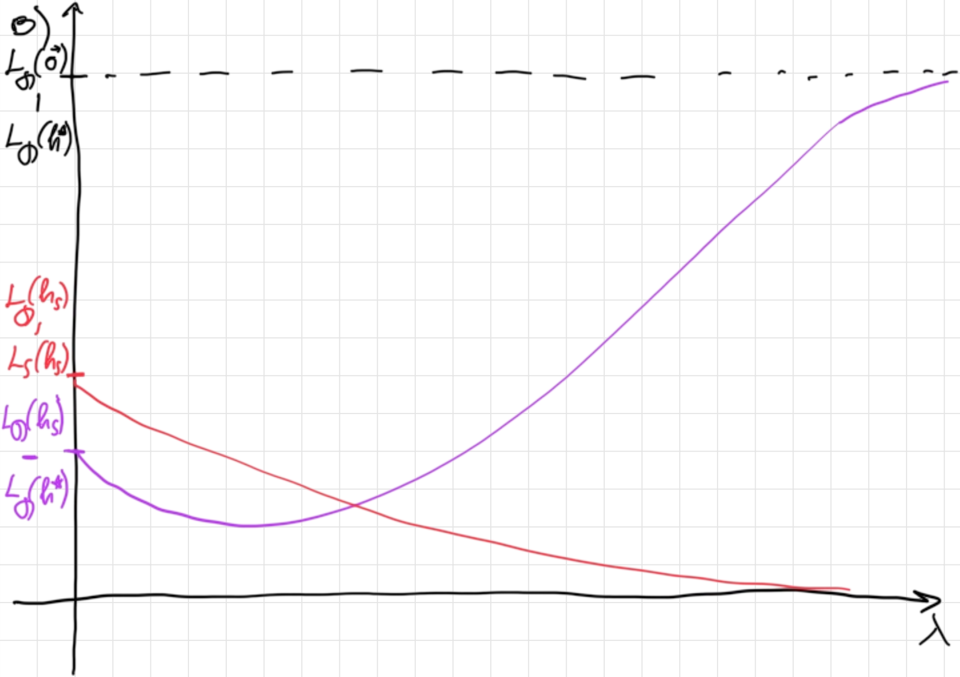
Consider the ridge regression problem $\arg\min_{\mathbf{w}} \lambda||\mathbf{w}||^2 + \sum_{i=1}^{m}(\langle \mathbf{w}, \mathbf{x_i} \rangle - y_i)^2$. Let: h_S be the hypothesis obtained by ridge regression with training set S; h^* be the hypothesis of minimum generalization error among all linear models.

(A) Draw, in the plot below, a *typical* behaviour of (i) *the training error* and (ii) *the test/generalization error* of h_S as a function of λ.

(B) Draw, in the plot below, a *typical* behaviour of (i) $L_{\mathcal{D}}(h_S) - L_{\mathcal{D}}(h^*)$ and (ii) $L_{\mathcal{D}}(h_S) - L_S(h_S)$ as a function of λ.

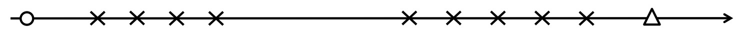



A)
$L_S(\vec{0})$
s_ξ
$L_\sigma(\vec{0})$
$L_\sigma(h_S)$
$L_S(h_S)$
λ


$L_{\mathcal{D}}(\vec{0})$
$L_{\mathcal{D}}(h^*)$
$L_{\mathcal{D}}(h_S)$
$L_S(h_S)$
$L_{\mathcal{D}}(h_S)$
$L_{\mathcal{D}}(h^*)$
λ

Exercize

Draw (approximately) the solution (clusters and centers) found by Lloyd algorithm for the 2 clusters ($k = 2$) problem, when the data ($x_i \in \mathbb{R}$) are the crosses in the figure below and the algorithm is initialised with center values indicated with the circle ($\circ$, cluster 1) and triangle ($\triangle$, cluster 2) shown in the figure.



Solution: