Machine Learning

Exercise

Fabio Vandin

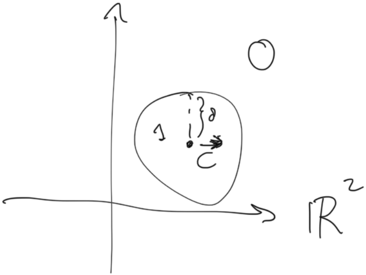
November 25, 2022

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 $\mathcal{H}$.



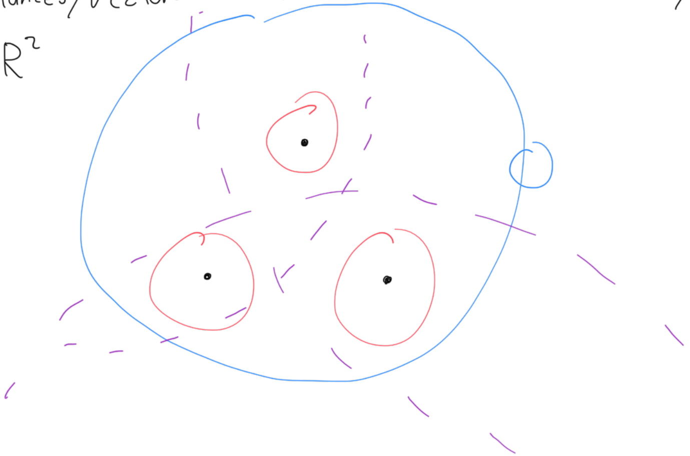
Solution

$VC\dim(\mathcal{H}) = 3 \rightarrow 12$

$VC\dim(\mathcal{H}) = 4 \rightarrow 10$

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

$X = \mathbb{R}^2$



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

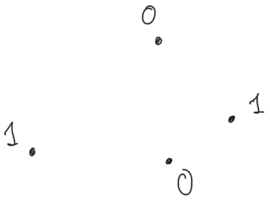
Consider an arbitrary set of 4 instances. Then there are 3 cases:

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

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

$\Rightarrow$ the set cannot be shattered

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



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

$\Rightarrow$ impossible to obtain

$\Rightarrow$ the set cannot be shattered

iii) 

$\Rightarrow$ impossible to obtain

$\Rightarrow$ the set cannot be shattered

From i), ii), iii) $\Rightarrow$ there is no set of 4 points that can be shattered by $\mathcal{H}$

$\Rightarrow \text{VCdim}(\mathcal{H}) \leq 3$

$\Rightarrow \text{VCdim}(\mathcal{H}) = 3$

Exercise

Let

$$\mathcal{H}_d = \{h_{\mathbf{w}}(\mathbf{x}) : h_{\mathbf{w}}(\mathbf{x}) = \text{sign}(< \mathbf{w}, \mathbf{x} >)\}$$

where $\mathcal{X} = \mathbb{R}^d$.

Prove that $VCdim(\mathcal{H}_d) = d$.

Solution We need to prove that $VCdim(\mathcal{H}_d) \geq d$ and that $VCdim(\mathcal{H}_d) \leq d$.

i) $VCdim(\mathcal{H}_d) \geq d$. We need to show a set of d vectors in $\mathbb{R}^d$ that is shattered by $\mathcal{H}_d$.

Consider $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_d\}$ with $\vec{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} i$, $\forall 1 \leq i \leq d$

This set is shattered by $\mathcal{H}_d$: we need to show that for every labeling $y_1, y_2, \dots, y_d$, where y_i is the label of $\vec{e}_i$, with $y_i \in \{-1, 1\}$, there is an hypothesis in $\mathcal{H}_d$ that assigns such labels to the set.

Consider an arbitrary labeling $y_1, y_2, \dots, y_d$: consider the hypothesis $h_{\vec{w}}$ where $\vec{w} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix}$. We have that for every i, with $1 \leq i \leq d$:

$$h_{\vec{w}}(\vec{e}_i) = \text{sign}(<\vec{w}, \vec{e}_i>) = \text{sign}\left(< \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix}, \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} > \right) = \text{Sign}(y_i) = y_i$$

ii) $VCdim(\mathcal{H}_d) \leq d$: we need to show that no set of $d+1$ vectors in $\mathbb{R}^d$ can be shattered by $\mathcal{H}_d$.

Consider an arbitrary set $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_{d+1}\}$ with $\vec{x}_i \in \mathbb{R}^d$ for $1 \leq i \leq d+1$.

They cannot be linearly independent $\Rightarrow \exists\, a_1, a_2, \dots, a_{d+1}$ with $a_i \in \mathbb{R}$, $1 \leq i \leq d+1$, such that:

$-$ not all a_i's are 0 $(\star)$

$\Rightarrow \sum_{i=1}^{d+1} a_i \vec{x}_i = 0$ $(\star\star)$

Define : $I = \{ i : a_i > 0 \}$. Note that it cannot be

$J = \{ j : a_j < 0 \}$ that $I = \emptyset = J$ (due to $(\star)$)

There are 3 cases: i) $I \neq \emptyset \neq J$; ii) $I \neq \emptyset = J$; iii) $I = \emptyset \neq J$

Case i) we are assuming $I \neq \emptyset \neq J$. Then

$$\underbrace{\sum_{i\in I} a_i \vec{x}_i = \sum_{j\in J} |a_j| \vec{x}_j}_{(\star\star\star)}$$

$$\sum_{i=1}^{d+1} a_i \vec{x}_i = \sum_{i\in I} a_i \vec{x}_i + \sum_{j\in J} a_j \vec{x}_j = 0 \quad (\text{by } \star\star)$$

$$\Longleftrightarrow \sum_{i\in I} a_i \vec{x}_i = \sum_{j\in J} -a_j \vec{x}_j$$

$$\Longleftrightarrow \sum_{i\in I} a_i \vec{x}_i = \sum_{j\in J} |a_j| \vec{x}_j$$

Assume that $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_{d+1}\}$ is shattered by $\mathcal{H}_d$:
must exists $\vec{w}$ such that

$$\langle \vec{w}, \vec{x}_i \rangle > 0 \quad \forall\, i \in I$$

$$\langle \vec{w}, \vec{x}_j \rangle < 0 \quad \forall\, j \in J$$

$$0 < \sum_{i \in I} \underbrace{a_i}_{\vee\, 0} \underbrace{\langle \vec{w}, \vec{x}_i \rangle}_{\vee\, 0} = \langle \overbrace{\sum_{i \in I} a_i \vec{x}_i}^{(\ast\ast)}, \vec{w} \rangle$$

$$= \langle \sum_{j \in J} |a_j| \vec{x}_j, \vec{w} \rangle$$

$$= \sum_{j \in J} \underbrace{|a_j|}_{\vee\, 0} \underbrace{\langle \vec{x}_j, \vec{w} \rangle}_{\wedge\, 0}$$

$$< 0$$

$\Rightarrow$ contradiction

Case ii) : $I \neq \emptyset = J$: same steps lead to

$$0 < \ldots \leq 0 \Rightarrow \text{contradiction}$$

Case iii) : $I = \emptyset \neq J$: same steps lead to

$$0 \leq \ldots < 0 \Rightarrow \text{contradiction}$$

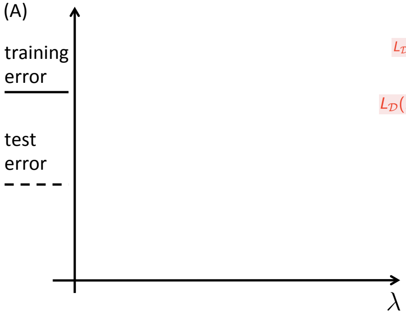
$\square$

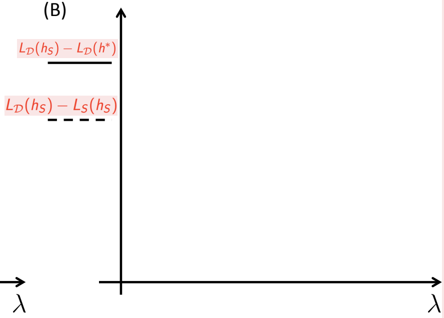
Exercise

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.

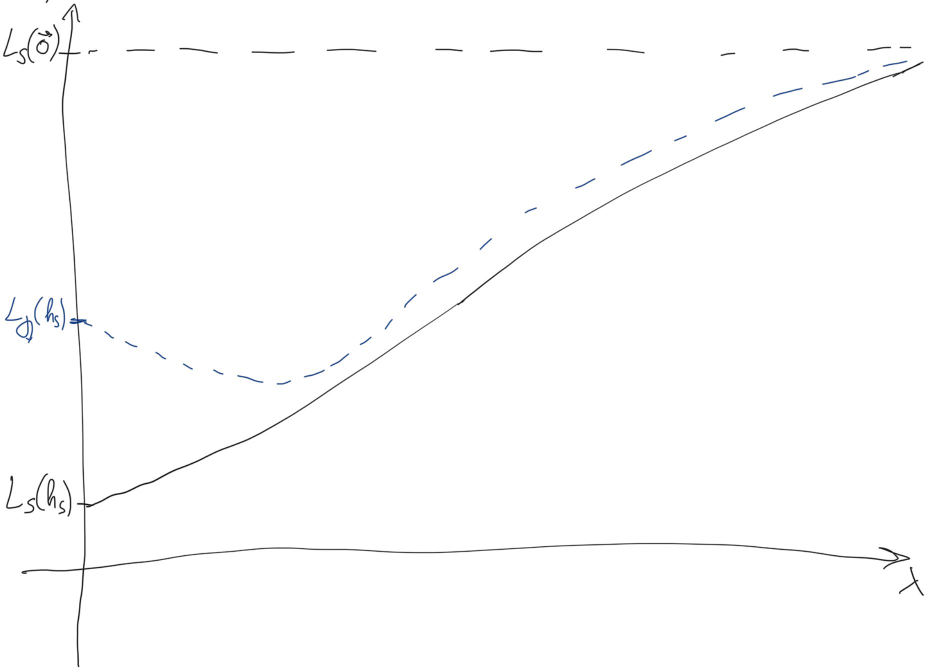
(A) Draw, in the plot below, a *typical* behaviour of (i) *the training error* and (ii) *the test/generalization error* of h_S as a function of λ.

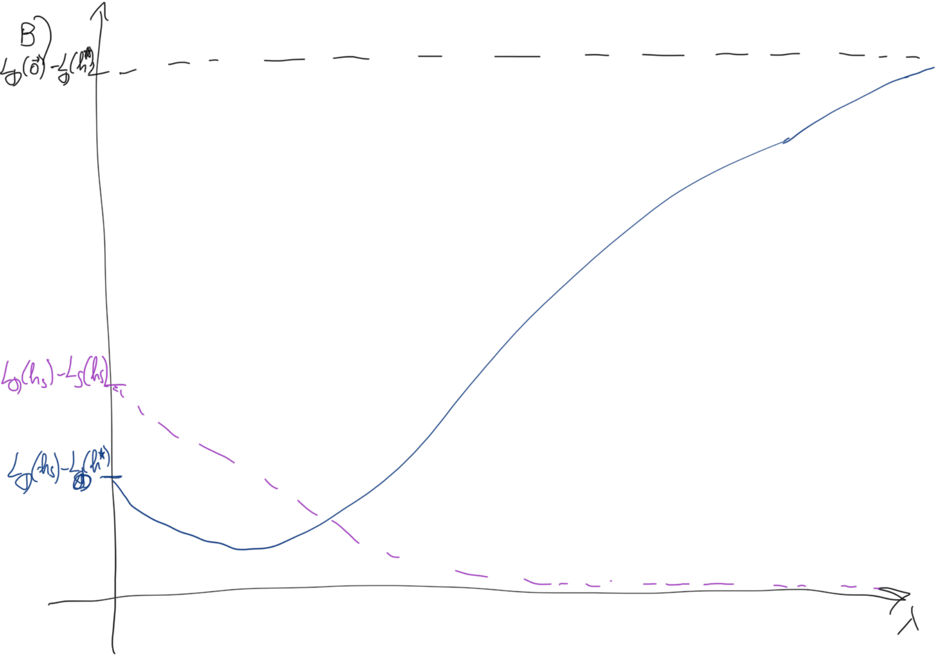
(B) Draw, in the plot below, a *typical* behaviour of (i) $L_{\mathcal{D}}(h_S) - L_{\mathcal{D}}(h^*)$ and (ii) $L_{\mathcal{D}}(h_S) - L_S(h_S)$ as a function of λ.





A)





B)
$L_D(\vec{0}) - L_D(h^*)$
$L_D(h_S) - L_S(h_S)$
$L_D(h_S) - L_D(h^*)$
λ