Machine Learning from Data – IDC

HW6 – Theory

This assignment includes question related to learning theory.

1. a.(10 pts) Let X be some infinite space of instances. Compute the VC-dimension of the following hypothesis space: $H = \{h: X \rightarrow \{-1,+1\}, |x: h(x) = -1| \leq 100\}$ The hypothesis space contains hypotheses that can return -1, 100 times or less.

   Answer: we know that the VC dimension is the maximum number of points in a possible hypothesis that we can shatter in different arrangements.
   This is that for any set of size 100 and any dichotomy, there exists a separation that classifies the set in 2 types of labels.(-1, and +1 in this case)
   Now we prove the question by saying that we can find a set of size bigger than 100, let say 101, but this actually don't exist.

   Then let's say that exist a set of 101 points, that can be shatter by the hypothesis h.
   Let's choose the instance where all labels receive the value -1, now by the definition of H, there are not and hypotheses h in the hypotheses space H , where the number of labels are bigger or equal than 100, so this will prove that we can't reach a set of points with 101 , so, the hypothesis space contains hypotheses that can return -1, 100 times or less and by that the VC dimension (H) = 100

   b.(10 pts) Give an example of an instance space X and a space H binary hypotheses on X, such that: $VC(H) = 2019$

   Answer: from the previous part a, and example can be :

   $H = \{h: X \rightarrow \{-1,+1\}, |x: h(x) = -1| \leq 2019\}$, VC(H) = 2019

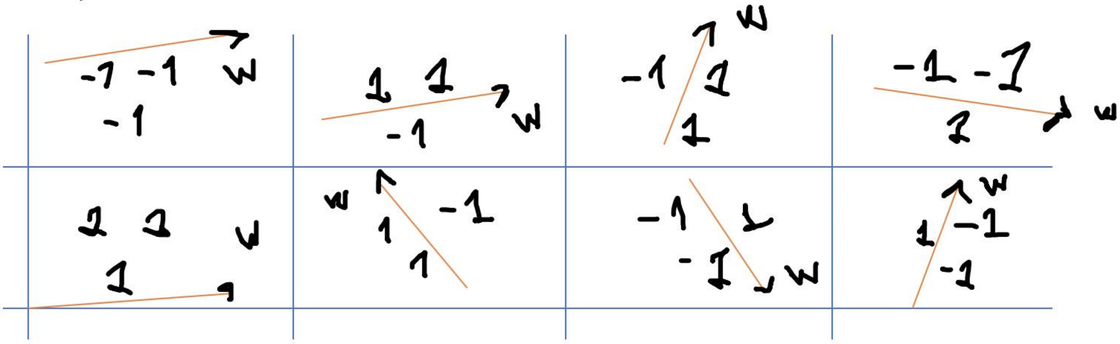
c. (20 pts) Consider the hypotheses space of all linear classifiers in the plain. That is, let

X = R² and then:

$$H = \left\{ h: \exists\, w_1, w_2, b \in \mathbb{R}\ s.t\ h(x_1, x_2) = \begin{cases} +1 & w_1x_1 + w_2x_2 + b > 0 \\ -1 & w_1x_1 + w_2x_2 + b \leq 0 \end{cases} \right\}$$

Show that VC(H) = 3 by performing the following steps.

1. 1) Find a set of size 3 that H shatters.



Here is a set of size 3 where H shatters.

2) Show that no set of size 4, $A = (z_1, z_2, z_3, z_4), z_i \in \mathbb{R}$ can be shattered by H.
Guidance: First prove the following lemma:
Lemma 1: Suppose a linear classifier h obtains prediction $y \in \{-1, +1\}$ on a set of points $z, z' \in \mathbb{R}^2$ $(h(z) = h(z') = y)$. Then it also obtains the same prediction on any intermediate point. Namely,

$$\forall \alpha \in [0,1] \quad h\big((1-\alpha)z + \alpha z'\big) = y$$

And use it in each of the following 3 possible cases:
a) The convex hull of A forms a line.
b) The convex hull of A forms a triangle.
c) The convex hull of A forms a quadrilateral.

We will prove the lemma 1: lets say we have , and hypothesis h, which classifies with the same criteria the points z, z' ∈ $\mathbb{R}^2$ with labels y = {-1,+1}. We will see that we have the same result for any point ,

let's say if h(z)=h(z')= y, then $\forall \alpha \in [0,1]$, will apply that h((1- α)z + αz') =y

let's define the point z= (x_1, y_1), z' = (x_2, y_2) , z'' = ((1- α)z + αz')

$z'' = ((1-\alpha)z + \alpha z') = z'' = ((1-\alpha)(x_1, y_1) + \alpha(x_2, y_2)) = ((1-\alpha)x1 + \alpha x_2, (1-\alpha)y1 + \alpha y_2)$

define $h(z'') = (x_3, y_3)$, and know we will calculate it:

$h(z'') = w_1x_3 + w_2y_3 + b = w_1((1-\alpha)x_1 + \alpha x_2) + w_2((1-\alpha)y_1 + \alpha y_2) + (1-\alpha+\alpha)b =$

$= w_1(1-\alpha)x_1 + w_2(1-\alpha)y_1 + (1-\alpha)b + w_1(\alpha)x_2 + w_2(\alpha)y_2 + (\alpha)b =$

$= (1-\alpha)(w_1x_1 + w_2y_1 + b) + \alpha(w_1x_2 + w_2y_2 + b) = (1-\alpha)h(z) + \alpha h(z') =$

$= (1-\alpha)y + \alpha y = y \rightarrow$ what we are looking for $h(z'') = y$

->Now we want to show that no set of size 4, $A=(z_1,z_2,z_3,z_4)$, $z_i \in R$ can be shattered by H.

We will use the lemma in the 3 different cases:

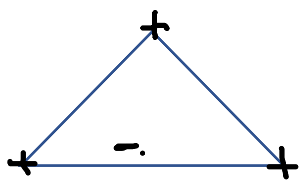
a) **The convex hull of A forms a line**, lets say z_1+ , z_2+, z_3-, z_4-.

Lets define $z_1 = z=1$, $z_2 = z'=1$ $z_3 = z''=0$

By $h(z'') = (1-\alpha)h(z) + \alpha h(z') = (1-\alpha)1 + \alpha(1) = 1$

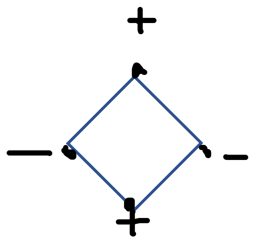
But we know that $h(z'') = 0$. Contradiction this dichotomy z_1+, z_2+, z_3-, z_4- don't satisfice the lemma 1

b) **The convex hull of A forms a triangle**.



The 3 points of the corners are 1+ , whit the dichotomy (+1,+1,+1,-1), in this case it will be impossible for z_4 in the middle of the triangle (all the area of the triangle is +4).

c) **The convex hull of A forms a quadrilateral**. Lets define the next square



in this case z_1,z_2,z_3,z_4 are on the corners, where the dichotomy of (-1, 1, -1, 1) is impossible since lemma 1 says that for all the points in the diagonal would be classified as its corners, in this case z_1,z_3 will be -1 and z_2,z_4 as +1 , but the intersection point will be classified both as +1 and -1, clearly a contradiction.

Then we prove that dim VC(H) < 4, and that $VC(H) \geq 3$, then VC(H) = 3

d. (20 pts) Consider the hypotheses space of all linear classifiers in d dimensional Euclidean space. That is, let X = R^d and then:

$$H = \left\{ h{:}\ \exists\ \overline{w} \in \mathbb{R}^d, b \in \mathbb{R}\ s.t\ h(\bar{x}) = \begin{cases} +1 & \overline{w}\bar{x} + b > 0 \\ -1 & \overline{w}\bar{x} + b \leq 0 \end{cases} \right\}$$

Show that VC(H) = d+1.

We will prove that $VC(H) \geq d + 1$ and that VC(H) < d+2, and with that we will show the claim

First part- $VC(H) \geq d + 1$

Let's define a group of instances X=[(0,0…0),(1,0…0)(1,1,…0)(1,1,…1)] of size d+1, we have dimension d, and hypothesis space H that shatters d+1 instances at least.

In the follow equation, we know that X is invertible, so we know that there exists a solution for it.

$$\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} b1 \\ \vdots \\ w1 \end{bmatrix} = \begin{bmatrix} y1 \\ \vdots \\ y(d+1) \end{bmatrix} \to \begin{bmatrix} b1 \\ \vdots \\ w1 \end{bmatrix} = \begin{bmatrix} y1 \\ \vdots \\ y(d+1) \end{bmatrix}$$

And by this we prove the first part. $VC(H) \geq d + 1$

Second part - VC(H) < d+2

We will show that there is a set of d + 2 instances that H can't shatter.
Assume a set of instances X with dimension d+2 over space dimension d+1, then in that case our set is linearly dependent.

Let's say there's is a concept c, where x_j the dependent instance has $h(x_j) = y_j$ and we classify it $y_j = -1$. For the rest of the instances, the sign of w with the inner product

$$h(x_i) = \text{sign}(x_i w) = sign(\alpha_i) = y_i$$

Finally, we can see that:

$$h(x_j) = sign(x_j w) = sign\left(\sum_{i \neq j} \alpha_i x_i\, w\right) > 0$$

but we already label $y_j = -1$ different of $h(x_j)$. So, it can't be VC(H) = d+2,

instead VC(H) < d+2

2. Let's look at this expression: $x_1 \wedge \overline{x_1} \wedge x_2 \wedge \overline{x_2} \wedge \ldots \wedge x_n \wedge \overline{x_n}$

For every positive instance [c(x) = true] we will remove the literals according to the following logic in order to save h as consistent:

If $x_i = 0\ \ then\ delete\ x_i\ from\ h\ and\ if\ \ x_i = 1\ then\ delete\ \overline{x_i}\ from\ h$, so we get just this literals h' = $x_i \ldots \wedge \ldots \wedge\ x_j \wedge \ldots \wedge\ \ \ldots \wedge\ x_m$

h' is consistent because for every negative instance at least one of them will be in h so h(x)=false for all negative instance.

For all positive instances h will give true because we removed all literals with label 0

The size of hypotheses space is upper bound by 3^n because every instance can be positive, negative or not at all. According to the formula shown in class we get

$$m \geq \frac{1}{\varepsilon}(ln3^n + ln\frac{1}{\delta})\ \geq \frac{1}{\varepsilon}(nln3 + ln\frac{1}{\delta})$$

So its polynomial for $\frac{1}{\varepsilon}, n, \frac{1}{\delta}$ and its PAC-learnable

3. the algorithm is as follow:

1. find the farthest point from (0,0)
2. find a straight line passes through this point and construct a straight-angled triangle

This line is h. every point below it its negative and every point above it its positive.

Let h (that line), C the concept line, δ, ε.

We can see that h is above or equal to C because if not there was a positive point above h which is impossible according to the way we built h.

Lets define h' the line who give us the error to ε. If h'<h then the probability to error is $< \varepsilon$ because the triangle there is bounded by h'.

If h'=h then se the error is ε.

if h' > h then $for\ m \geq \frac{ln\left(\frac{1}{\delta}\right)}{\varepsilon}$ we get :

$$error_D(h) = (1-\varepsilon)^m \leq e^{-\varepsilon m} \leq e^{-\varepsilon \frac{ln\left(\frac{1}{\delta}\right)}{\varepsilon}} = e^{-ln\left(\frac{1}{\delta}\right)} = e^{ln\delta} = \delta$$

Then, the probability will be bigger then ε is bounded by δ and its PAC-learnable