Advanced Machine Learning - Assignment 1

Adrian Iordache - Artificial Intelligence

April 25, 2021

Exercise 1

Statement:

Give an example of a finite hypothesis class $\mathcal{H}$ with $VCdim(\mathcal{H}) = 2021$. Justify your choice.

Solution:

From the Seminar 3 we know that the class of Boolean Conjunctions over the variables $x_1, ..., x_d$ ($d \geq 2$) has $VCdim(\mathcal{H}^d_{con}) = d$. (1)

Also, each literal of the conjunction has 2 possible values, from that $|\mathcal{H}^d_{con}| = 2^d$ (finite). (2)

From (1) and (2) we can choose $\mathcal{H}^{2021}_{con}$ as a finite hypothesis class which has $VCdim(\mathcal{H}^{2021}_{con}) = 2021$.

Exercise 2

Statement:

Consider $\mathcal{H}_{balls}$ to be the set of all balls in $\mathbf{R}^2$:

$\mathcal{H}_{balls} = \left\{ \mathrm{B}(x, r), x \in \mathbb{R}^2, r \geq\ 0 \right\}$, where $\mathrm{B}(x, r) = \left\{ y \in \mathbb{R}^2 \mid \|y - x\|_2 \leq r \right\}$

As mentioned in the lecture, we can also view $\mathcal{H}_{balls}$ as the set of indicator functions of the balls $\mathrm{B}(x, r)$ in the plane: $\mathcal{H}_{balls} = \left\{ h_{x,r} : \mathbb{R}^2 \longrightarrow \{0, 1\}, h_{x,r} = \mathbf{1}_{\mathrm{B}(x,r)}, x \in \mathbb{R}^2, r > 0 \right\}$.

Can you give an example of a set A in $\mathbf{R}^2$ of size 4 that is shattered by $\mathcal{H}_{balls}$? Give such an example or justify why you cannot find a set A of size 4 shattered by $\mathcal{H}_{balls}$.

Solution:

There is NO set A of size 4 in $\mathbf{R}^2$ that can be shattered by $\mathcal{H}_{balls}$.

Based on Radon's theorem any set of 4 points $\{a, b, c, d\}$ in $\mathbf{R}^2$ can be partitioned into two disjoint sets whose convex hulls don't have empty intersection.

We consider two cases:

- **Case 1**: Without loss of generality, $\{a, b, c, d\}$ are partitioned into $\{a, b, c\}$ and $\{d\}$, and the convex hull of $\{a, b, c\}$ contains $\{d\}$. So any hypothesis from $\mathcal{H}_{balls}$ that contains the triple $\{a, b, c\}$ has to include the singleton {d}. In other words, we cannot realize $\{a, b, c\}$ without including {d}.

- **Case 2**: Without loss of generality, $\{a, b, c, d\}$ are partitioned into $\{a, b\}$ and $\{c, d\}$, and the line segment $a - b$ intersects with the line segment $c - d$. It is not possible for both diagonals to be separated from the endpoints of the respective other diagonal by a closed ball. We can select at most one diagonal segment when the endpoints of the other diagonal are far apart, but we can't select the second diagonal without including the first one, as displayed in the image below (Case 2.1). When the diagonal segments have approximately the same lengths we cannot select either one individually without including one other endpoint. (Case 2.2)

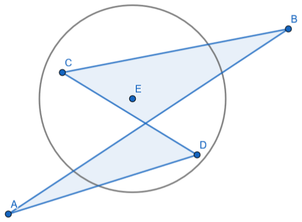

Figure 1: Case 2.1

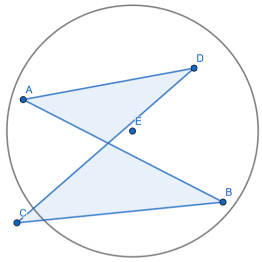

Figure 2: Case 2.2

In the previous cases we did not considered when all 4 points are collinear. The reason for that is based on the fact that in Lecture 6 we proved that $\mathcal{H}_{balls}$ cannot shatter a set 3 collinear points $\mathbf{R}^2$.

Exercise 3

Statement:

Let $\mathcal{X} = \mathbf{R}^2$ and consider $\mathcal{H}_\alpha$ the set of concepts defined by the area inside a right triangle ABC with two catheti AB and AC parallel to the axes (O_x and O_y) and with $AB/AC = \alpha$ (fixed constant > 0). Consider the realizability assumption. Show that the class $\mathcal{H}_\alpha$ can be (ε, δ) - PAC learned by giving an algorithm A and determining an upper bound on the sample complexity $m_H(\varepsilon, \delta)$ such that the definition of PAC-learnability is satisfied.

Solution:

We consider the following definition for $\mathcal{H}$:

$$\mathcal{H} = \left\{ h_{\mathrm{A,P_{BC}}} : \mathbb{R}^2 \longrightarrow \{0, 1\} \mid h_{\mathrm{A,P_{BC}}}(Z) = \begin{cases} 1, & (x_Z \geq x_A) \wedge (y_Z \geq y_A) \wedge (x_Z + y_Z \leq x_{P_{BC}} + y_{P_{BC}}) \\ 0, & \text{otherwise} \end{cases}, Z \in \mathbb{R}^2 \right\}$$

Where A is the right point in triangle and P_{BC} is a given point on the hypotenuse.

For $\mathcal{H}$ to be **PAC learnable** needs to:

- $\exists m_{\mathcal{H}} : (0, 1)^2 \longrightarrow N$
- $\exists A$ learning algorithm such that:
  - $\forall \varepsilon \forall \delta \forall f \forall D$ we have that $\underset{S \sim D^m}{\mathrm{P}}[L_{h^\cdot,D}(h_S) \leq \epsilon] \geq \delta \Longleftrightarrow \underset{S \sim D^m}{\mathrm{P}}[L_{h^\cdot,D}(h_S) > \epsilon] < \delta$

where $\varepsilon, \delta \in (0,1)$, $f \in \mathcal{H}$ labeling function, D distribution over $\mathbf{R}^3$
S training set $|S| = m \geq m_{\mathcal{H}(\varepsilon,\delta)}$ examples sampled i.i.d from D and labeled by f
such that $A(S) = h_S$.

We are under the realizability assumption, so there exists a labeling function $f \in \mathcal{H}, f = h^* = h_{A^*,P^*_{BC}}$ that labels the training data.

Consider a training set $S = \left\{ (x_1, y_1), (x_2, y_2), \dots, (x_m, y_m) \middle| \begin{array}{l} y_i = h^*_{A^*,P^*_{BC}}(x_i), \\ x_i \in \mathbb{R}^2, x_i = (x_{i1}, x_{i2}) \end{array} \right\}$

Take LA the learning algorithm that gets the training set S and outputs $LA(S) = h_S = h_{A_S,P_{BC_S}}$, where $A_S = (x_{min1}, x_{min2})$ such that:

$$x_{min1} = \min_{\substack{i \in \overline{1,m} \\ y_i = 1}} x_{i1}$$

$$x_{min2} = \min_{\substack{i \in \overline{1,m} \\ y_i = 1}} x_{i2}$$

and $P_{BC_S} = (x_{i1}, x_{i2})$ such that:

$$i = \arg\max_{\substack{i \in \overline{1,m} \\ y_i = 1}} x_{i1} + x_{i2}$$

If the set S does not contain any positive samples then we choose the point $A = P_{BC} = (0,0)$.

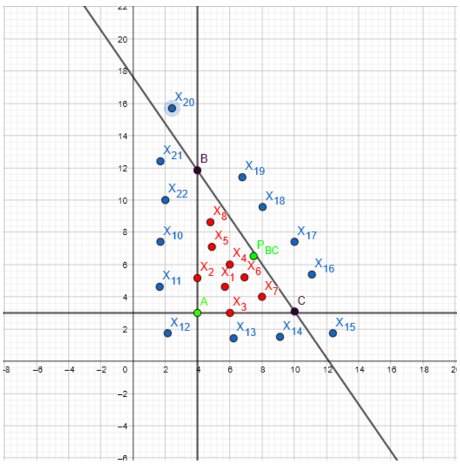

Figure 3: Visualization of a $h_{A_S,P_{BC_S}}$

The main idea of LA

This learning algorithm uses the as boundaries A = the "smallest" point in the set S and P_{BC} = the "farthest" point in S, which is placed on the hypotenuse of the right triangle ABC.

From construction, A is ERM, meaning that $L_S(h_S) = 0$ (h_S doesn't make any errors on the training set S).

Notation: $T^* = h_{A^*,P^*_{BC}}$ and $T_S = h_{A_S,P_{BC_S}}$.

We can make the observation that h_S makes errors in the region $T^* \setminus T_S$ assigning the label 0 to points that should get label 1.

Let $\varepsilon > 0$, $\delta > 0$ and D a distribution over R. We want to find how many number of training samples $m \geq m_{\mathcal{H}(\varepsilon,\delta)}$ do we need such that:

$$\underset{S \sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) \leq \epsilon] \geq \delta \iff \underset{S \sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) > \epsilon] < \delta$$

We can distinguish 2 cases:

Case 1:

If $\mathcal{D}(T^*) = \underset{x \sim D}{\mathrm{P}}[x \in T^*] \leq \varepsilon$ then in this case

$L_{h^*,\mathcal{D}}(h_S) = \underset{x \sim D}{P}[h_S(x) \neq h^*(x)] = \underset{x \sim D}{P}[x \in T^* \setminus T_S] \leq \underset{x \sim D}{P}[x \in T^*] \leq \varepsilon$ so we have that

$\underset{S \sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) \leq \epsilon] = 1$ (this happens every time). ✓

Case 2:

If $\mathcal{D}(T^*) = \underset{x\sim D}{\mathrm{P}}[x \in T^*] > \varepsilon$ then in this case

We consider three regions along the sides of the triangle T_S, named R_1, R_2, R_3 with $\mathcal{D}(R_i) = \underset{x\sim D}{\mathrm{P}}[x \in \mathrm{R_i}] = \frac{\varepsilon}{3}$.

If $T_S \cap R_1 \neq \emptyset$ or $T_S \cap R_2 \neq \emptyset$ or $T_S \cap R_3 \neq \emptyset$ then $\underset{S\sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) > \epsilon] = 0$ ✓

In order to have $L_{h^*,D}(h_S) > \epsilon$, we need that T_S will not intersect at least one region R_i.

Let $F_i = \{S \sim D^m \mid T_S \cap R_i = \emptyset, i \in 1 \ldots 3\} \implies$

$$\underset{S\sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) > \epsilon] \leq \sum_{n=1}^{3} \underset{S\sim D^m}{\mathrm{P}}[F_i] = 3 \cdot \left(1 - \frac{\epsilon}{3}\right)^m \leq 3 \cdot e^{-\frac{\varepsilon}{3}\cdot m}$$

In this case we have $3 \cdot e^{-\frac{\varepsilon}{3}\cdot m} < \delta \implies m \geq m_{\mathcal{H}(\varepsilon,\delta)} = \frac{3}{\varepsilon}\log\frac{3}{\delta}$

From the above cases $\implies$ $\mathcal{H}$ can be (ε, δ) - PAC learned by LA with sample complexity $m \geq m_{\mathcal{H}(\varepsilon,\delta)} = \frac{3}{\varepsilon}\log\frac{3}{\delta}$

Exercise 4

Statement:

Consider $\mathcal{H}$ to be the class of all centered in origin sphere classifiers in the 3D space. A centered in origin sphere classifier in the 3D space is a classifier h_r that assigns value 1 to a point if and only if it is inside the sphere with radius $r > 0$ and center given by the origin $\mathbf{O}(0, 0, 0)$. Consider the realizability assumption.

a. show that the class $\mathcal{H}$ can be (ε, δ) - PAC learned by giving an algorithm A and determining an upper bound on the sample complexity $m_H(\varepsilon, \delta)$ such that the definition of PAC-learnability is satisfied.

b. compute VCdim($\mathcal{H}$).

Solution:

a) We consider the following definition for $\mathcal{H}$:

$$\mathcal{H} = \left\{ h_{\mathrm{r}} : \mathbb{R}^3 \longrightarrow \{0,1\} \mid h_{\mathrm{r}}(x) = \begin{cases} 1, & \sqrt{a^2+b^2+c^2} \leq r, \\ 0, & \text{otherwise} \end{cases}, x = (a,b,c) \in \mathbb{R}^3, \mathrm{r} \in \mathbb{R}_+^* \right\}$$

For $\mathcal{H}$ to be **PAC learnable** needs to:

- $\exists m_{\mathcal{H}} : (0,1)^2 \longrightarrow N$
- $\exists A$ learning algorithm such that:
  - $\forall\varepsilon\forall\delta\forall f\forall D$ we have that $\underset{S\sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) \leq \epsilon] \geq \delta \iff \underset{S\sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) > \epsilon] < \delta$

  where $\varepsilon, \delta \in (0,1)$, $f \in \mathcal{H}$ labeling function, D distribution over $\mathbf{R}^3$
  S training set $|S| = m \geq m_{\mathcal{H}(\varepsilon,\delta)}$ examples sampled i.i.d from D and labeled by f such that $A(S) = h_S$.

Consider a training set $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$. We are under the realizability assumption, so there exists a labeling function $f \in \mathcal{H}, f = h^* = h_{r^*} = SP(O(0,0,0), r^*) = SP(O, r^*) \in \mathcal{H}$ with $r^* > 0$, that labels the training data.

Take A the learning algorithm that gets the training set S and outputs $h_S = A(S)$ = the "smallest" sphere containing all positive examples in S.

Formal definition for A:

$$A(S) = h_S = h_{r_S} = \max_{\substack{i \in \overline{1,m} \\ y_i = 1}} \|x_i\|_2 = \max_{\substack{i \in \overline{1,m} \\ y_i = 1 \\ x_i = (a_i, b_i, c_i)}} \sqrt{a_i^2 + b_i^2 + c_i^2} \tag{4.1}$$

If there are no positive points with label 1 in S will define the algorithm to output the sphere centered in origin of radius 0, assigning label 0 to all points in S.
($\nexists i \in 1..m$ such that $y_i = 1 \longrightarrow A(S) = h_0$)

From construction, A is ERM, meaning that $L_S(h_S) = 0$.

As a short proof for the statement above we can assume that $L_S(h_S) \neq 0$, from that we can distinguish 2 possible cases:

1) $\exists i \in 1..m$ with $(x_i, y_i) \in S$ such that $h_S(x_i) = 0$ and $y_i = 1$
2) $\exists i \in 1..m$ with $(x_i, y_i) \in S$ such that $h_S(x_i) = 1$ and $y_i = 0$

Case 1: From this assumption results $\exists i \in 1..m$ with $(x_i, y_i) \in S$ such that $\|x_i\|_2 > r_S$ and $\|x_i\|_2 \leq r^* \implies \|x_i\|_2 \in (r_S, r^*]$. We assume that $\exists i \in 1..m$ with $(x_i, y_i) \in S$ such that $\|x_i\|_2 \in (r_S, r^*]$. This is impossible because if there would be such a sample labeled positive by the function f the learning algorithm A would have choose r_S at least to be $\|x_i\|$ $(\perp) \implies \nexists i \in 1..m$ with $(x_i, y_i) \in S$ such that $h_S(x_i) = 0$ and $y_i = 1$.

Case 2: From this assumption results $\exists i \in 1..m$ with $(x_i, y_i) \in S$ such that $\|x_i\|_2 \leq r_S$ and $\|x_i\|_2 > r^* \implies \perp$ (from the realizability condition) $\implies \nexists i \in 1..m$ with $(x_i, y_i) \in S$ such that $h_S(x_i) = 1$ and $y_i = 0$

From (1), (2) $\implies A$ is ERM, meaning that $L_S(h_S) = 0$.

Let $\varepsilon > 0$, $\delta > 0$ and D a distribution over R. We want to find how many number of training samples $m \geq m_{\mathcal{H}(\varepsilon,\delta)}$ do we need such that: $\underset{S\sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) \leq \epsilon] \geq \delta \iff \underset{S\sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) > \epsilon] < \delta$

Case 1: if $D(SP(O, r^*)) \leq \varepsilon$ then $\underset{S\sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) > \epsilon] = 0$ ✓

Case 2: if $D(SP(O, r^*)) > \varepsilon$ then

Build a sphere $SP(O, z)$ such that $D(SP(O, r^*)/SP(O, z)) = \varepsilon$.

- **Case 2.1:** If $SP(O, r_S) \cap SP(O, z) \neq \emptyset$ then $\underset{S\sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) > \epsilon] = 0$ ✓
- **Case 2.2:** Else $SP(O, r_S) \cap SP(O, z) = \emptyset$ then let $F = \{S \sim D^m \mid SP(O, r_S) \cap SP(O, z) = \emptyset\}$ $\implies \underset{S\sim D^m}{\mathrm{P}}[L_{h^*,D}(h_S) > \epsilon] \leq \underset{S\sim D^m}{\mathrm{P}}[F]$ = the probability that $SP(O, r_S)$ will not intersect $SP(O, r^\cdot) \cap SP(O, z)$ (where the training set S is chosen i.i.d.)
  In this case we have $(1-\varepsilon)^m \leq e^{-em} < \delta \implies m \geq m_{\mathcal{H}(\varepsilon,\delta)} > \lceil \frac{1}{\varepsilon} \log \frac{1}{\delta} \rceil$

From the above cases $\implies$ $\mathcal{H}$ can be (ε, δ) - PAC learned by A with sample complexity $m \geq m_{\mathcal{H}(\varepsilon,\delta)} > \lceil \frac{1}{\varepsilon} \log \frac{1}{\delta} \rceil$

b) To show that the VC-dimension of a hypothesis class $\mathcal{H}$ is d, we need to prove:

- $\exists C$ such that $|C| = d$ and C is shattered by $\mathcal{H}$. $(VCdim(\mathcal{H}) \geq d)$
- $\forall C$ such that $|C| = d + 1$ and C is NOT shattered by $\mathcal{H}$. $(VCdim(\mathcal{H}) < d + 1)$

We will start from the assumption that $VCdim(\mathcal{H}) = 1$ and proof that
$VCdim(\mathcal{H}) \geq 1$ and $VCdim(\mathcal{H}) < 2$.

Let $C = \{c \mid c \in \mathrm{R}^3\}$ and $|C| = 1$ we can distinguish 2 cases:

Case 1: for $y_c = 0$ we can choose the hypothesis h_{r_C} with $r_C \in (||c||, \infty)$

Case 2: for $y_c = 1$ we can choose the hypothesis h_{r_C} with $r_C \in (0, ||c||]$

From the cases above we can see that the hypothesis class $\mathcal{H}$ can learn any labeling function for a set C containing a single sample $\implies$ $VCdim(\mathcal{H}) \geq 1$ (1).

Now let $C = \{c_i \mid c_i \in \mathrm{R}^3, i \in 1 \dots 2\}$, $|C| = 2$ and without loss of generality we can consider $||c_1||_2 < ||c_2||_2$.

It is trivial to prove that the hypothesis class $\mathcal{H}$ can achieve the following labeling functions: $(0, 0), (1, 1), (1, 0)$. Next we will show that the label $(0, 1)$ cannot be achieved by the $\mathcal{H}$.

Let's assume that $\exists h_z \in \mathcal{H}$ such that $h_z(c_1) = 0$ and $h_z(c_2) = 1 \implies ||c_1||_2 > r_z$ and $||c_2||_2 \leq r_z \implies ||c_2||_2 \leq r_z < ||c_1||_2$, but we know that $||c_1||_2 < ||c_2||_2 \implies \perp \implies \nexists h_z \in \mathcal{H}$ such that $h_z(c_1) = 0$ and $h_z(c_2) = 1$. So $VCdim(\mathcal{H}) < 2$. (2)

From (1), (2) we have that $VCdim(\mathcal{H}) = 1$.

Exercise 5

Statement:

Consider $\mathcal{H} = \left\{ h_\theta : \mathbb{R} \longrightarrow \{0,1\} \mid h_\theta(x) = \mathbf{1}_{[\theta,\theta+1]\cup[\theta+2,\infty)}, \theta \in \mathbb{R} \right\}$. Compute $VCdim(\mathcal{H})$.

Solution:

$$\mathcal{H} = \left\{ h_\theta : \mathbb{R} \longrightarrow \{0,1\} \mid h_\theta(x) = \begin{cases} 1, & x \in [\theta, \theta+1] \cup [\theta+2, \infty) \\ 0, & \text{otherwise} \end{cases}, \theta \in \mathbb{R} \right\}$$

Figure 4: Visualization of hypothesis class $\mathcal{H}$

Lets consider a set with three points $Z = \{z_i \mid i \in 1..3, z_i \in R\}$ we will choose the points z_1, z_3 to be symmetric with respect to z_2 with ± 0.75, where $z_2 \in R$.

We will try to proof that Z is shattered by $\mathcal{H}$ resulting in $VCdim(H) \geq |Z| \geq 3$.

Label (0, 0, 0): To achieve this labeling we can choose $\theta \in (z_2 + 0.75, \infty)$. (e.g.: $\theta = z_3 + 1$)

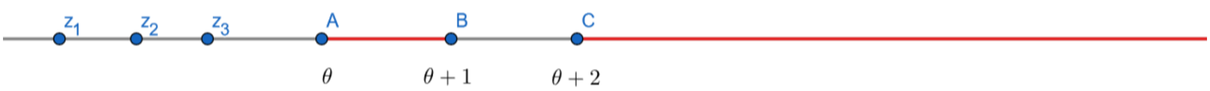

Figure 5: Label Case $(0,0,0)$

Label (0, 0, 1): To achieve this labeling we can choose $\theta \in (z_2, z_2 + 0.75]$. (e.g.: $\theta = z_2 + 0.25$)

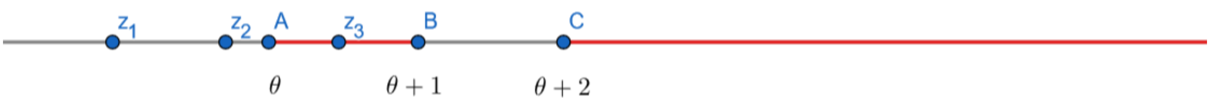

Figure 6: Label Case $(0,0,1)$

Label (0, 1, 0): To achieve this labeling we can choose $\theta \in (z_2 - 0.75, z_2 - 0.25)$. (e.g.: $\theta = z_2 - 0.5$)

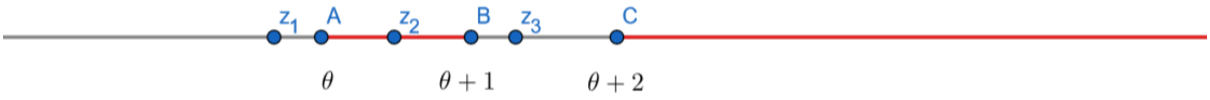

Figure 7: Label Case (0, 1, 0)

Label (1, 0, 0): To achieve this labeling we can choose $\theta \in (z_2 - 1.25, z_2 - 1)$. (e.g.: $\theta = z_2 - 1.2$)

Figure 8: Label Case (1, 0, 0)

Label (0, 1, 1): To achieve this labeling we can choose $\theta \in (z_2 - 2.75, z_2 - 2]$. (e.g.: $\theta = z_2 - 2.5$)

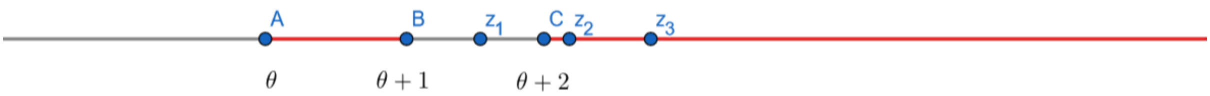

Figure 9: Label Case (0, 1, 1)

Label (1, 0, 1): To achieve this labeling we can choose $\theta \in [z_2 - 1.75, z_2 - 1.25]$. (e.g.: $\theta = z_2 - 1.5$)

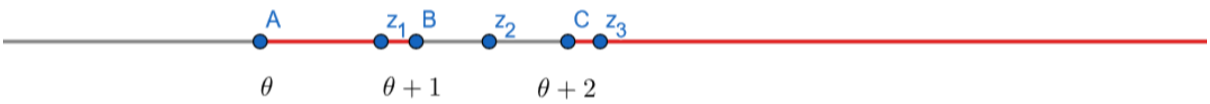

Figure 10: Label Case (1, 0, 1)

Label (1, 1, 0): To achieve this labeling we can choose $\theta \in (z_2 - 1, z_2 - 0.75]$. (e.g.: $\theta = z_2 - 0.85$)

Figure 11: Label Case $(1, 1, 0)$

Label (1, 1, 1): To achieve this labeling we can choose $\theta \in (z_2 - \infty, z_2 - 2.75]$. (e.g.: $\theta = z_2 - 5$)

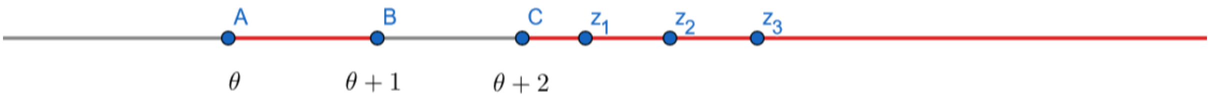

Figure 12: Label Case $(1, 1, 1)$

The algorithm used for finding lower and upper boundaries for θ is the following:

- fix θ on the number line (e.g.: $\theta = 1$)
- fix z_2 on the number line depending on the case you are trying to solve
- based on the symmetry with respect z_2 check if z_1 and z_3 respect the case constraints
- solve for x the equation $\theta = z_2 - x$ for different values and identify the boundaries

As we can see above all $2^{|Z|} = 2^3 = 8$ possible labels can be learned by the hypothesis class $\mathcal{H}$.
$\implies VCdim(\mathcal{H}) \geq 3$. (1)

Lets consider a set with four points $Z = \{z_i \mid i \in 1..4, z_i \in \mathrm{R}\}$, without loss of generality we will consider that $z_1 < z_2 < z_3 < z_4$. We will prove that the label $(1, 0, 1, 0)$ **can't** be learned.

Suppose that $\exists h_z \in \mathcal{H}$ that can accomplish the labels $(1, 0, 1, 0) \implies$

- $h_z(z_1) = 1 \implies z_1 \in [\theta_z, \theta_z + 1]$ or $z_1 \in [\theta_z + 2, \infty)$
- $h_z(z_2) = 0 \implies z_2 \in (-\infty, \theta_z)$ or $z_2 \in (\theta_z + 1, \theta_z + 2)$
- $h_z(z_3) = 1 \implies z_3 \in [\theta_z, \theta_z + 1]$ or $z_3 \in [\theta_z + 2, \infty)$
- $h_z(z_4) = 0 \implies z_4 \in (-\infty, \theta_z)$ or $z_4 \in (\theta_z + 1, \theta_z + 2)$

We have $z_4 \in (-\infty, \theta_z)$ or $z_4 \in (\theta_z + 1, \theta_z + 2) \implies z_4 < \theta z + 2$ and $z_1 < z_2 < z_3 < z_4 \implies z_1 < \theta z + 2$.

From $z_1 < \theta z + 2$ and $z_1 \in [\theta_z, \theta_z + 1]$ or $z_1 \in [\theta_z + 2, \infty)$
$\implies z_1 \in [\theta_z, \theta_z + 1]$.

Now $z_1 \geq \theta_z$ and $z_1 < z_2 < z_3 < z_4 \implies z_2 \geq \theta_z \implies z_2 \in (\theta_p + 1, \theta_p + 2)$

Now $z_2 > \theta_z + 1$ and $z_1 < z_2 < z_3 < z_4 \implies z_3 > \theta_z + 1 \implies z_3 \in (\theta_p + 2, \infty)$

Now $z_3 > \theta_z + 2$ and $z_1 < z_2 < z_3 < z_4 \implies z_4 > \theta_z + 2 \implies \perp \implies$
$\nexists h_z \in \mathcal{H}$ that can accomplish the labels $(1, 0, 1, 0) \implies VCdim(\mathcal{H}) < 4$. (2)

From (1), (2) $\implies VCdim(\mathcal{H}) = 3$.

Exercise 6

Statement:

Let $\mathcal{X}$ be an instance space and consider $\mathcal{H} \subseteq \{0, 1\}^X$ a hypothesis space with finite VC dimension. For each $x \in \mathcal{X}$, we consider the function $z_X : \mathcal{H} \longrightarrow 0, 1$ such that $z_x(h) = h(x)$ for each $h \in \mathcal{H}$. Let $Z = \{z_x : \mathcal{H} \longrightarrow \{0, 1\}, x \in \mathcal{X}\}$. Prove that $VCdim(Z) < 2^(VCdim(H) + 1)$.

Solution: