Advanced Machine Learning - Assignment 1

Iordachescu Anca Mihaela

April 2022

1 Exercise 1

The example of a finite hypothesis class $\mathcal{H}$ I chose is inspired from the set of linear classifiers in $\mathcal{HS}_0^{2022}$.

$$\mathcal{H} = \{h_w : \mathcal{R}^{2022} \to \{-1, 1\}, h_w(x) = sign(< w, e_i >)|w \in \mathcal{R}^{2022}\}$$

$$and$$

$$w_i = \begin{cases} 1, e_i \in B \\ -1, e_i \notin B \end{cases}$$

where w = $(w_1, w_2, ..., w_{2022})$ and A, B subsets of standard unit vectors in R^{2022}, A = $\{e_0, e_1, ..., e_{2022}\}$ and $B \subseteq A$. It is obvious that $\mathcal{H}$ is a finite hypothesis class because, due to the fact that $|A| = 2022$, the number of created subsets B with $B \subseteq A$ is finite.

In order to demonstrate that the VCdimension of $\mathcal{H}^{2022}$ is 2022, according to *Lecture 6*, I have to show that:

1. *There exists a set A of size 2022 that is shattered by $\mathcal{H}$.*

Let A be the set of standard unit vectors A = $\{e_0, e_1, ..., e_{2022}\}$. I want to prove that VCdim($\mathcal{H}$) $\geq$ 2022 by having A and showing that for each subset B of A there exists a function $h_B \in \mathcal{HS}_0^{2022}$ such that each element of B is labelled 1 while the elements of A-B are labelled -1.

Defining w as below, I can construct h_B as $h_B(e_i) = sign(< w, e_i >)$ that creates label 1 for elements from B (because the dot product is greater than 0) and creates label -1 for elements from A-B (because the dot product is less than 0) showing that there exists a set A of size 2022 that is shattered by H. Thus, *VCdim($\mathcal{H}$) $\geq$ 2022.*

2. *Every set A of size 2023 is not shattered by $\mathcal{H}$.*

I use the first property of the VCdim($\mathcal{H}$), presented in the *Lecture 7*, according to which VCdim($\mathcal{H}$) $\leq$log$_2$($|\mathcal{H}|$). Knowing that H shatters A, $|A| = 2022$ and there is only one hypothesis for each subset concerning A, the concluded result is $|\mathcal{H}_a| = |\mathcal{H}| = 2^{|A|} = 2^{2022}$.

Then, $VCdim(\mathcal{H}) \leq log_2(\mathcal{H}) \Leftrightarrow VCdim(\mathcal{H}) \leq log_2(2^{2022}) \Leftrightarrow VCdim(\mathcal{H}) \leq 2022 \Leftrightarrow VCdim(\mathcal{H}) < 2023$.

Demostrating (1) and (2), it can be proven that $VCdim(\mathcal{H}) = 2022$.

2 Exercise 2

Due to the fact that $\mathcal{H}$ has n elements and, therefore, it is a finite hypothesis class, I can use the first property of the VCdim($\mathcal{H}$), presented in the *Lecture 7*, according to which VCdim($\mathcal{H}$) $\leq$log$_2$($|\mathcal{H}|$) = $log_2(|n|)$. **(1)**

Due to the fact that $\mathcal{H}$ shatters a set C of $\frac{n}{2}$ points, this means that $VCdim(\mathcal{H}) \geq |C| = \frac{n}{2}$. **(2)**.

Using **(1)** and **(2)**, I can easily write the next inequality and find the biggest even number that satisfies it:

$$\frac{n}{2} \leq VCdim(\mathcal{H}) \leq log_2(n)$$

When n = 2 $\Rightarrow \frac{2}{2} \leq$VCdim(H) $\leq log_2(2)$ - satisfied.
When n = 4 $\Rightarrow \frac{4}{2} \leq$VCdim(H) $\leq log_2(4)$ - satisfied.
When n = 2p, p $\in N, p \geq 3 \Rightarrow \frac{n}{2} >log_2(n)$.

Looking at the relations written above, it is obvious that the biggest even number that satisfies the inequality is n = 4. So, I have to construct a $\mathcal{H}$ that has 4 elements that can shatters a set C of $\frac{4}{2}$ = 2 points.

For instance, I choose $\mathcal{H} = \{h_{0,2}, h_{0,0.5}, h_{5,0.5}, h_{0,6}\}$ where $h_{x,r} \in \mathcal{H}_{balls}$ and C = {A(0, 2), B(5,0)}. It is obvious that $\mathcal{H}$ shatters C because I can obtain every configuration of labelling:

1. label (0,0) $\rightarrow$ I am going to use $h_{0,0.5}$ since none of the points are inside the ball.

2. label (1,1) $\rightarrow$ I am going to use $h_{0,5}$ since both points are inside the ball.

3. label (0,1) $\rightarrow$ I am going to use $h_{5,0.5}$ since B is inside the ball and A is outside the ball.

4. label (1,0) $\rightarrow$ I am going to use $h_{0,2}$ since A is inside the ball and B is outside the ball.

3 Exercise 3

I am going to prove that VCdim($\mathcal{H}$) = 2, where $\mathcal{H}$ is the set of axis aligned rectangles with the center in origin O(0,0).

In order to demonstrate that the VCdimension of VCdim($\mathcal{H}$) = 2, according to *Lecture 6*, I have to show that:

1. *There exists a set A of size 2 that is shattered by $\mathcal{H}$.*

Let A be A = $\{A(0,5), B(4,0)\}$. I am going to prove that there exists a classifier $h \in \mathcal{H}$ that can achieve all the possible labellings of the set of points.

i. Both points have labels equal to 0(h(A) = h(B) = 0), so both of them should be outside a axis-aligned rectangle. I can choose the rectangle with the corner points: X(1,1), Y(-1,-1), Z(-1,1), W(1, -1) and the points A and B don't lie inside them because $5 > 1, 5 > -1$ and $4 > 1, 4 > -1$.

ii. Both points have labels equal to 1(h(A) = h(B) = 1), so both of them should be inside a axis-aligned rectangle. I can choose the rectangle with the corner points: X(7,7), Y(-7,-7), Z(-7,7), W(7, -7) and the points A and B lie inside them because their $-7 < 0 < 7, -7 < 5 < 7$ and $-7 < 0 < 7, -7 < 4 < 7$.

iii. A has label 0 and B has label 1 (h(A) = 0 and h(B) = 1), so A should lie inside a axis-aligned rectangle and B should lie outside it. I can choose the rectangle with the corners: X(5,1), Y(-5,-1), Z(-5,1), W(5, -1) because it is obvious that A lies outside it while B is inside it.

iv. B has label 0 and A has label 1 (h(A) = 1 and h(B) = 0), so A should lie inside a axis-aligned rectangle and B should lie ouside it. I can choose the rectangle with the corners: X(1,6), Y(-1,-6), Z(-1,6), W(1, -6) because it is obvious that A lies inside it while B is outside it.

Since there are 2 points, it is obvious that there can be $2^2 = 4$ cases of labelling and I have proved above that the axis-aligned rectangles can generate all the labelling, so $\mathcal{H}$ shatters A. Thus, VCdim($\mathcal{H}$) $\geq$ 2.(**1**)

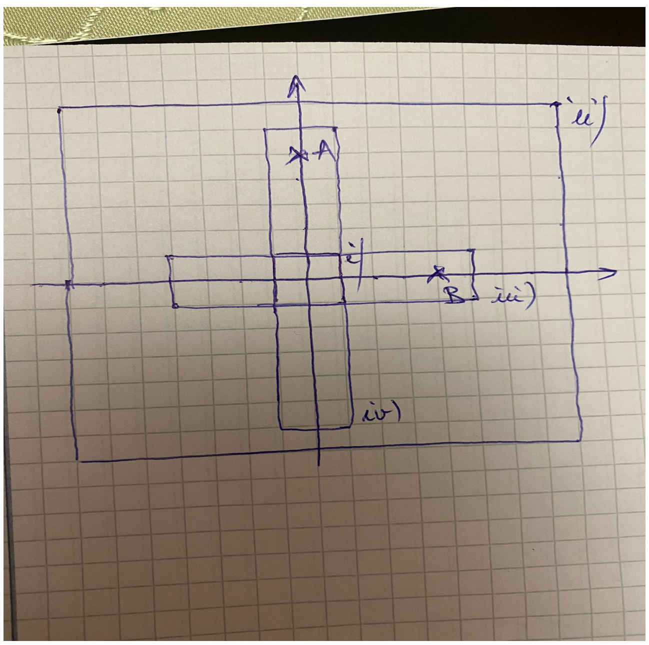

Figure 1. The axis-aligned rectangles for the set A(0,5) and B(4,0)

2. *Every set A of size 3 is not shattered by $\mathcal{H}$.*

I am going to prove that for any set of three points $D = \{A, B, C\}$, D cannot be shattered. Without lose of generality, I can use the set that consists of 3 points $A(x_0, y_0), B(x_1, y_1), C(x_2, y_2)$ where $|x_0| \leq |x_1| \leq |x_2|$ and {A, B, C} $\in R^2$. I will prove that, no matter the relations between y_0, y_1, y_2, there are at least one labelling that can't be realized.

i. Consider $|y_0| \leq |y_1| \leq |y_2|$. Using Reductio Ad Absurdum, I am going to prove that the labelling (1,0,1) can't be realised. Suppose that there exists a classifier $h \in \mathcal{H}$ that can achieve the (1, 0, 1) labelling for the set of 3 points. If h(C) is equal to 1, it means that C is inside a zero centered axis-aligned rectangle. However, since $|x_1| \leq |x_2|$ and $|y_1| \leq |y_2|$, it means that every rectangle that contains C also contains B $\Rightarrow$ h(B) would be 1. Contradiction (h(B) was supposed to be 0).

ii. Consider $|y_0| \leq |y_2| \leq |y_1|$. Using Reductio Ad Absurdum, I am going to prove that the labelling (0, 1, 1) can't be realised. Suppose that there exists a classifier $h \in \mathcal{H}$ that can achieve the (0, 1, 1) labelling for the set of 3 points. If h(A) is equal to 0, it means that A is outside a zero centered axis-aligned rectangle. However, since $|x_0| \leq |x_1|$ and $|y_0| \leq |y_1|$, it means that if A is outside the rectangle, B also has to be outside the rectangle $\Rightarrow$ h(A) would be 0. Contradiction (h(A) was supposed to be 1).

iii. Consider $|y_1| \leq |y_0| \leq |y_2|$. Using Reductio Ad Absurdum, I am going to prove that the labelling (1, 0, 1) can't be realised. Suppose that there exists a classifier $h \in \mathcal{H}$ that can achieve the (1, 0, 1) labelling for the set of 3 points. If h(B) is equal to 0, it means that B is outside a zero centered axis-aligned rectangle. However, since $|x_1| \leq |x_2|$ and $|y_1| \leq |y_2|$, it means that if B is outside the rectangle, C also has to be outside the rectangle $\Rightarrow$ h(C) would be 0. Contradiction (h(C) was supposed to be 1).

iv. Consider $|y_1| \leq |y_2| \leq |y_0|$. Using Reductio Ad Absurdum, I am going to prove that the labelling (0, 0, 1) can't be realised. Suppose that there exists a classifier $h \in \mathcal{H}$ that can achieve the (0, 0, 1) labelling for the set of 3 points. If h(B) is equal to 0, it means that B is outside a zero centered axis-aligned rectangle. However, since $|x_1| \leq |x_2|$ and $|y_1| \leq |y_2|$, it means that if B is outside the rectangle, C also has to be outside the rectangle $\Rightarrow$ h(C) would be 0. Contradiction (h(C) was supposed to be 1).

v. Consider $|y_2| \leq |y_1| \leq |y_0|$. Using Reductio Ad Absurdum, I am going to prove that the labelling (0, 0, 1) can't be realised. Suppose that there exists a classifier $h \in \mathcal{H}$ that can achieve the (0, 0, 1) labelling for the set of 3 points. If h(A) is equal to 0, it means that A is outside a zero centered axis-aligned rectangle. However, since $|x_0| \leq |x_1| \leq |x_2|$ and $|y_1| \leq |y_2|$, the only way to construct this zero centered axis-aligned rectangle would be to set the first coordinate of corners points with a value a, $|a| < |x_0|$. This means that if A is outside the rectangle, B and C are also outside the rectangle $\Rightarrow$ h(B) and h(C) would be 0. Contradiction (h(C) and h(B) were supposed to be 1).

i. Consider $|y_2| \leq |y_0| \leq |y_1|$. Using Reductio Ad Absurdum, I am going to prove that the labelling (0, 1, 0) can't be realised. Suppose that there exists a classifier $h \in \mathcal{H}$ that can achieve the (0, 1, 0) labelling for the set of 3 points. If h(B) is equal to 1, it means that B is inside a zero centered axis-aligned rectangle. However, since $|x_0| \leq |x_1|$ and $|y_0| \leq |y_1|$, it means that every rectangle that contains B also contains A $\Rightarrow$ h(A) would be 1. Contradiction (h(A) was supposed to be 0).

To sum up, I have proven above that for any set of points of size 3, there is a labelling that can't be realized, meaning that $\mathcal{H}$ cannot shatter a set of 3 points. Thus, $VCdim(\mathcal{H}) < 3$. **(2)**

Using **(1)** and **(2)**, I proved that VCdim($\mathcal{H}$) = 2.

4 Exercise 4

For the PAC-learnable exercises I have used the seminar exercises as model of problem solving.

Let S be the training set $S = \{(x_1, y_1), ..., (x_n, y_n)\}$, where $x_i \in R^2$ and $y_i = h^*(x_i)$ knowing that, under the realizability assumption, h^* is the labelling function that labels the training data.

I have to come up with an algorithm A that constructs the smallest right triangle with AB/AC = α that assigns the label 1 for each point inside the triangle and label 0 for each point outside the triangle.

Knowing that AB and AC are parallel to the axis Ox and Oy and AB/AC = $\alpha, \alpha > 0$, I can come up with the next observations:

1. The coordinates for the points of the triangle are: $A(a_1, b_1)$, $B(a_2, b_1)$ and $C(a_1, b_2)$ (due to the fact that AB parallel to Ox and AB parallel to Oy), $a_1, b_1, a_2, b_2 \in$ R. Without losing the generality, I consider the point A to be the leftest between A and B, $a_1 \leq a_2$.

2. b2 - b1 = (a2 – a1)/alpha (due to the fact that AB/AC = α).

The algorithm A that I propose has the following steps:

1. Choose the point X(x0,y0) from the training set that has the smallest value of the first coordinate and, also, the point Y(x1, y1) from the training set that has the smallest value of the second coordinate. In this way, we construct the first point A(a1, b1) = A(x0, y1).

2. In order to generate C, I have to find the value of b_2 and in order to generate B, I have to find the value of a_1. My idea is to look at the rightmost and highest point in the plane in order to generate the catheti AB and AC with respect to the coefficient α.

If there is no pair $(m, 1) \in S$, I can choose A = B = C = Z(z,z), where $z = |x|+|y|$ knowing that $x = max\{|x1|, |x2|, \ldots, |xn|\}$ and y = $max\{|y1|, |y2|, \ldots |yn|\}$ in order to place Z outside the training set S.

Looking at the figure placed below, h_S is the indicator function of the tightest right triangle enclosing all points that have positive labels. Thus, by construction, A is ERM meaning that h* doesn't make any errors on the training set.

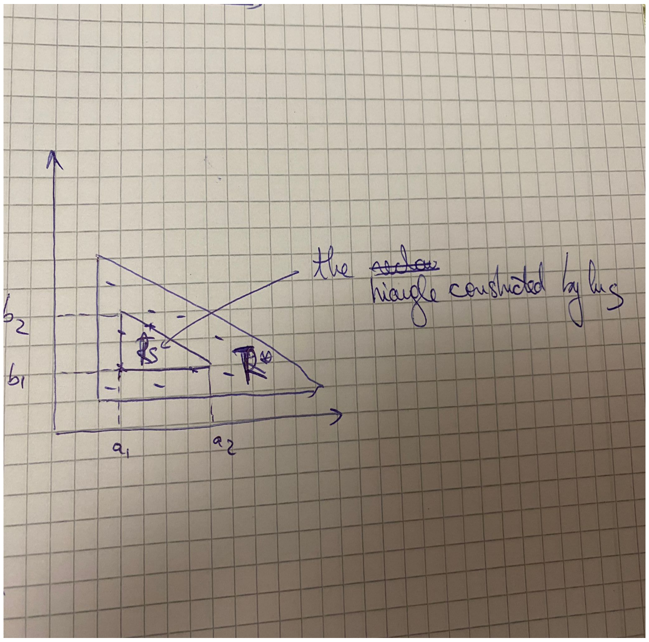

Figure 2. Tightest right triangle enclosing the points with label 1

We make the observation that h* makes errors in region $R^* - R_h$, assigning the label 0 to the values that should get a positive label. However, all the values that are in R_h and outside R^* are labelled correctly (points in the R_h have positive label while points outside R^* have label 0).

Let $\epsilon > 0, \delta > 0$ and D a distribution in R. We want to find m $\geq$m$_H(\epsilon, \delta)$ such that P $= P(L_{D,h^*}(h_S) \leq \epsilon) \geq 1 - \delta$.

1. $D([R^*]) \leq \epsilon \rightarrow L_{D,h^*}(h_S) = P(x \in$ elements that are in R^* but not in $R_h) \leq P(x \in$elements that are in $R^*) \rightarrow L_{D,h^*}(h_S)) \leq \epsilon \rightarrow P = 1$.

2. $D([R^*]) > \epsilon \rightarrow$ Define R_1, R_2 and R_3 as below such that $D(R_1) = D(R_2) = D(R_3) = \frac{\epsilon}{3}$.

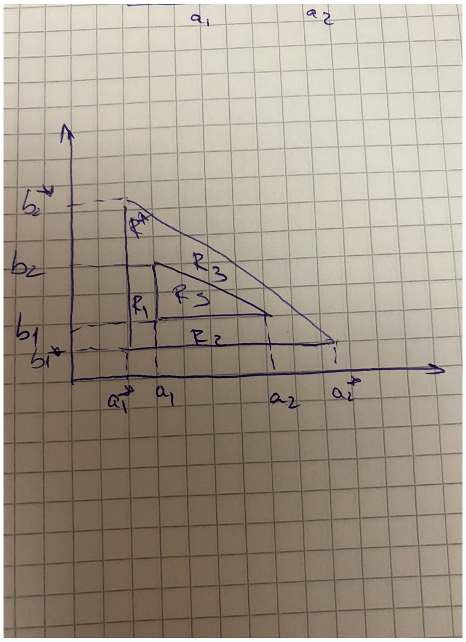

Figure 1. R1, R2, R3 zones

If R_1, R_2, R_3 intersects with R_S, then $\mathrm{L}_{D,h^*}(h_S) = P(x \in$ elements that are in R^* but not in $R_h) \leq P(x \in$elements that are in $\mathrm{R}_1, R_2, R_3) \leq P(x \in R_1) + P(x \in R_2) + P(x \in R_3) = \epsilon \to L_{D,h^*}(h_S)) \leq \epsilon \to P = 1$.

If none of R_1, R_2, R_3 intersect with R_S, then $P(L_{D,h^*}(h_S) > \epsilon) \leq 3 * (1 - \frac{\epsilon}{3})^m \leq 3 * e^{\frac{\epsilon}{3}} < \delta \to m > \frac{3}{\epsilon} * \log \frac{3}{\delta}$. $\to \mathcal{H}$ PAC-learnable

5 Exercise 5

a & b. Let S be the training set $S = \{(x_1, y_1), ..., (x_n, y_n)\}$, where $x_i \in R$ and $y_i = h^*(x_i)$ knowing that, under the realizability assumption, h^* is the labelling function that labels the training data.

Before constructing the learning algorithm, I want to have a look at the labelling of each hypothesis class, if x_i are sorted without the lose of generality:

1. $\mathcal{H}_1 \to (0, 0, 0, ..., 0, 1, 1, ..., 1)$
2. $\mathcal{H}_2 \to (1, 1, 1, ..., 1, 0, 0, ..., 0)$
3. $\mathcal{H}_3 \to (0, 0, 0, ..., 0, 1, 1, ..., 1, 0, 0, ..., 0)$

Looking above, we can construct the learning algorithm A that gets the training set S and output h_s as the tightest interval containing all the positive examples, because whatever classifier I choose from each of three hypothesis classes, it is clear that each of the labelling contains at most one interval that contains positive elements. I have to mention that I have used the demonstration of PAC-learnability of the intervals hypothesis class.

$h_s = h_{a,b}$, where $h_{x,y} \in \mathcal{H}_{intervals}$, a = minimum x_i, b = maximum x_i and $R_S = [a, b]$

If there is no pair $(m, 1) \in S$, I can choose a = b = maximum x_i * 2 because (maximum x_i * 2, 0) $\in S$.

Therefore, h_S is the indicator function of the tightest interval enclosing all values that have positive labels. Thus, by construction, A is ERM meaning that h_S doesn't make any errors on the training set.

We make the observation that h* makes errors in region $R^* - R_h$, assigning the label 0 to the values that should get a positive label. However, all the values that are in R_h and outside R^* are labelled correctly (points in the R_h have positive label while points outside R^* have label 0).

Let $\epsilon > 0, \delta > 0$ and D a distribution in R. We want to find m $\geq$m$_H(\epsilon, \delta)$ such that P = $P(L_{D,h^*}(h_S) > \epsilon) < \delta$.

1. $D([a^*, b^*]) \leq \epsilon \rightarrow$P = 0.

2. $D([a^*, b^*]) \geq \epsilon \rightarrow R_1 = [a_1^*, a_1]$ and $R_2 = [a_2^*, a_2]$ such that $D(R_1) = D(R_2) = \frac{\epsilon}{2}$.

If none of R_1, R_2 intersect with R_S, then P = 0.

If one of them intersects with R_S, then $P \leq 2 * (1 - \frac{\epsilon}{2})^m \leq 2 * e^{\frac{\epsilon}{2}} < \delta \rightarrow m > \frac{2}{\epsilon} * \log \frac{2}{\delta}. \rightarrow \mathcal{H}$ PAC-learnable.

c. In order to demonstrate that the VCdimension of VCdim($\mathcal{H}$) = 2, according to *Lecture 6*, I have to show that:

1. *There exists a set A of size 2 that is shattered by $\mathcal{H}$.*

Let A be A = $\{a_0 = 0.2, a_1 = 1.2\}$.

a. If the label of a_0 has to be 0 and label of a_1 has to be 0, I can choose the classifier h_{θ_1} with $\theta_1 = 3$ because $0.2 < 3$ and $1.2 < 3$, thus the labelling for each of them is 0.

b. If the label of a_0 has to be 0 and label of a_1 has to be 1, I can choose the classifier h_{θ_1} with $\theta_1 = 1$ because $0.2 < 1$ and $1.2 > 1$, thus the labelling for $a_0 = 0.2$ is 0 and the labelling for $a_1 = 1.2$ is 1.

c. If the label of a_0 has to be 1 and label of a_1 has to be 0, I can choose the classifier h_{θ_2} with $\theta_2 = 0.5$ because $0.2 < 0.5$ and $1.2 > 0.5$, thus the labelling for $a_0 = 0.2$ is 1 and the labelling for $a_1 = 1.2$ is 0.

d. If the label of a_0 has to be 1 and label of a_1 has to be 1, I can choose the classifier h_{θ_2} with $\theta_2 = 4$ because $0.2 < 4$ and $1.2 < 4$, thus the labelling for $a_0 = 0.2$ is 1 and the labelling for $a_1 = 1.2$ is 1.

I have proved that for any labelling of the set A of size 2 presented above I have found a classifier from the finite hypothesis class $\mathcal{H}$ to label correctly each case. Thus, VCdim($\mathcal{H}$) $\geq$ 2.(**1**)

2. *Every set A of size 3 is not shattered by $\mathcal{H}$.*

Using RAA (Reductio Ad Absurdum), I am going to prove that the labelling (1,0,1) can't be realised for any set of size 3 and, therefore, any set of size 3 cannot be shattered by $\mathcal{H}$.

Suppose that there is a set $A = \{a_0, a_1, a_2\}$ with $a_0 \leq a_1 \leq a_2$ and there exists a classifier $h \in \mathcal{H}$ that can achieve the (1, 0, 1) labelling: $h(a_0) = 1, h(a_1) = 0, h(a_2) = 1$. Knowing $h(a_0) = 1$, this means that the labelling was obtained using the next classifiers:

1. h_{θ_1} and knowing that $h(a_0) = h_{\theta_1}(a_0) = 1$ and since $a_0 \leq a_1 \leq a_2$, this means that $\theta_1 \leq a_0$. Therefore, $\theta_1 \leq a_1 \leq a_2$ and $h(a_1) = h_{\theta_1}(a_1) = 1$ and $h(a_2) = h_{\theta_1}(a_2) = 1$. The labelling is (1, 1, 1) which represents a contradiction.

2. h_{θ_2} and knowing that $h(a_0) = h_{\theta_2}(a_0) = 1$ and since $a_0 \leq a_1 \leq a_2$, this means $\theta_2 \geq a_0$. However, using this classifier $h(a_1) = h_{\theta_2}(a_1) = 0$, we get that $\theta_2 \leq a_1$ and because of the fact that $a_1 \leq a_2$, $h(a_2) = h_{\theta_2}(a_2)$ would be equal to 0. The labelling is (1, 0, 0) which represents a contradiction.

3. h_{θ_1,θ_2} and knowing that $h(a_0) = h_{\theta_1,\theta_2}(a_0) = 1$ and since $a_0 \leq a_1 \leq a_2$, this means that $\theta_1 \leq a_0 \leq \theta_2$. However, using this classifier $h(a_1) = h_{\theta_1,\theta_2}(a_1) = 0$ and because of the fact that $a_1 \leq a_2$, $h(a_2) = h_{\theta_2}(a_2)$ would be equal to 0. The labelling is (1, 0, 0) which represents a contradiction.

I have proven above that there is no case that produces the labelling (1, 0, 1). There is no set of size 3 that can produce the labelling (1, 0, 1), so there is no set of size 3 that can be shattered by $\mathcal{H}$. Thus, $VCdim(\mathcal{H}) < 3$. **(2)**

Using **(1)** and **(2)**, I have proved that VCdim($\mathcal{H}$) = 2.

6 Exercise 6

Brief explanation of the solution: receiving the hint, in this exercise I am going to convert the 1-decision list classifier into a threshold linear function based on the next observation: being given x, it is important that the feature x_i has to dominate among all the features if i is the smallest index such that $c_i(x) = 1$. In order to do that, we can create a dot product between x_i and powers of 2 in order to allow important variables to dominate the sum of the rest.

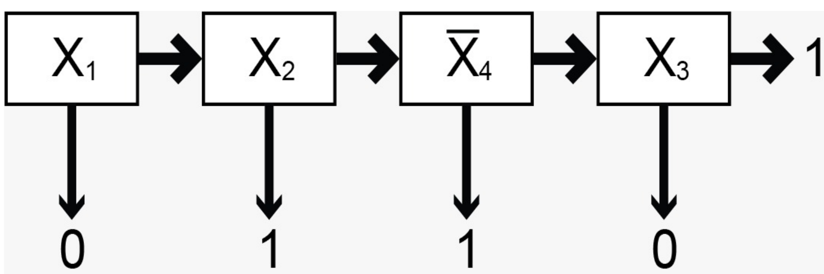

Figure 4. Example of 1-decision list

Solution:

First of all, I modify the ordered sequence L by changing every bit b_i and, also the bit b, that are equal to 0 to -1 (this is going to help me in order to create the classifier based on a dot product succeeded by applying sign function. After that, due to the fact that I want to construct a linear function, I have to compute the weights that are multiplied with the features in the linear function. Each weight should take into consideration the b_i value and, also, I have to make sure that it has to dominate all the values after it in order to be the only active feature if applicable. This idea of a dominating variable can be concretized by multiplying each b_i with a power of 2 that decrease for the next index geometrically by dividing it by 2 and adding the products to create a sum. Also, to the sum we add the b given.

For example, taking into consideration the 1-decision list shown in the Figure 4, after converting L to L=$\{(x_1, -1), (x_2, 1), (\overline{x_4}, 1), (x_3, -1)\}$ by replacing b_1 and b_4 with -1, the computed sum would be: sum $= x_1 * -1 * 2^4 + x_2 * 1 * 2^3 + \overline{x_4} * 1 * 2^2 + x_3 * -1 * 2^1 + 1$. So, it is easily to see that each b_i multiplied with a power of 2 represent a factor that is multiplied with x_i and could serve easily as a weight in a half-space classifier. The only problem is that the number of factors is equal to the length of sequence L and, in order to use it as a array of weight, the weight array should have the length of a input $x \in R^n$. This can be easily constructed because, due to the fact that the decision list takes into account only one boolean variable, $c_i \in \{x_1, x_2, ..., x_n, \overline{x_1}, \overline{x_2}, ..., \overline{x_n}\}$. Therefore, we can modify also the input and the weight in this way:

x = $(x_1, x_2, ..., x_n) \rightarrow x =$(x$_1, x_2, ..., x_n, \overline{x_1}, \overline{x_2}, ..., \overline{x_n})$ and

$$w_i = \begin{cases} w_i, x_i & \text{is literal in the decision list} \\ 0, x_i & \text{is not literal in the decision list} \end{cases}$$

where w_i computed as described above.

I am going to construct a function $g : R^{2n} \rightarrow R$ g(x) = $< w, x > +b$ that describes the sum and a threshold linear function $h : R \rightarrow \{-1, 1\}$ h(x) = sign($g(x)$). It is important to mention that we can switch the places of each x_i, without changing in fact the input x, by putting in the first places the x_i that are 1 in order to activate and dominate the variables in the sum function g(x). Thus, I have proved that a 1-decision list is equal to a threshold function. According to *Lecture 7*, the VCdim($\mathcal{H}$), knowing $\mathcal{H}$ is the set of threshold half-spaces in R^n, is equal to n+1.

I had to show that $a * n + c \leq VCdim$(H) $\leq b * n + d$ with $a, b, c, d \in R$ and knowing that $n + 1 \leq VCdim$(H) $\leq n + 1 \rightarrow$ I have proven that there exists a = 1, b = 1, c = 1 and d = 1 that satisfies the inequality.