Intelligent Learning and Analysis Systems: Machine Learning — Exercise 8

Yauhen Selivonchyk
Łukasz Segiet
Duc Duy Pham
Dominik Schindler

January 16, 2015

1 VC-dimension of some concept classes over $\mathbb{R}^2$.

Problem a)

$VC_{dim}(Circles) = 3$

Proof:
It is clear that any 2 points can be shattered. For this, just choose any cirlce, s.t. on point is within the circle and the other one outside of it, or a circle that contains both or none of them
For 3 non-colinear points, we can also always find a circle, s.t. 2 of then are inside the circle, leaving one out, and vice verca. One simple method is to first choose a circle, s.t. all 3 points lie on the border of the circle. To leave one point outside the circle, just move the circle towards the other two points in perpendicular direction to the line going through these two points. To keep this one point inside, move the circle in opposite direction. The fact that there also always exist circles that contain either none or all of the points is trivial.
For 4 points we need to consider 3 cases:

- The 4 points are colinear:
  It is clear that we can not find any cirlces, that only contain the first and the third point on the line and exclude the second and the forth.
- The 4 points form a triangle (convex hull):
  In this case 3 points are the corner points of the triangle. Here it is also impossible to find a cirlce that contains the corner points and excludes the non corner point.
- The 4 points form a quadrilateral (convex hull):
  For each point consider the opposing point with greatest distance. This way we achieve the two diagonals of the quadrilateral. It is not possible

to find a circle that contains the two points forming the longer diagonal while excluding the other two.

So we have shown that we can not find any subset of $\mathbb{R}^2$ with magnitude 4, that can be shattered by the concept class of circles.
Thus we get $VC_{dim}(Circles) = 3$. □

Problem b)

$VC_{dim}(Triangles) = 7$

Proof:
Results directly from Problem c.). □

Problem c)

Let $\mathcal{P}^(k)$ be the class of convex polygons consisting of k vertices.
$VC_{dim}(\mathcal{P}^{(k)}) = 2k + 1$

Proof (sketch):
Place d points on a circle and label them arbitrarily $+$ or $-$. We want to find or construct a polygon that contains all points labeled with $+$ und excludes all points labeled with $-$.
Let P be the set of positively labeled points and N the set of negatively labeled points.
Construct the polygon as follows:
If $|P| \leq |N|$, create a convex polygon with $|P|$ vertices, s.t. the points in P are its vertices. This way every point labeled $+$ is in the polygon and every other point outside of it.
If $|P| > |N|$, create a convex polygon with |N| sides, s.t. the sides of the polygon are tangent to the circle in the points of N. This way the circle is within the constructed polycon, so every point labeled $+$ is in the polygon and every point labeled $-$ exactly on one side of the polygon. Now move each side by a very small amount $\epsilon > 0$ towards the center of the circle, s.t. now every point labeled $+$ is still in the polygon but every point labeled $-$ is outside of the polygon.
If we have $2k + 1$ points, the constructed polygon consists of at most k points. Since we labeled the points arbitrarily, we have found a subset of size $2k + 1$ that is shattered by $\mathcal{P}^{(k)}$.
So we can conclude $VC_{dim}(\mathcal{P}^{(k)}) \geq 2k + 1$ We now still need to show that $VC_{dim}(\mathcal{P}^{(k)}) \neq 2k + 2$. This can be done by firstly showing that placing the points an a circle maximizes the number of possible subsets of this particular subset that can be *realized* by $\mathcal{P}^{(k)}$. Secondly we would just need to add another point to the circle arbitrarily and find a subset of these $2k + 2$ points that can not be *realized* by $\mathcal{P}^{(k)}$. □

2 Some basic properties of the VC-dimension.

Problem a)

Let $\mathcal{C}$ be a concept class with $|\mathcal{C}| < \infty$.
Show that $VC_{dim}(\mathcal{C}) \leq \log_2 |\mathcal{C}|$.

Proof:
Let X be the instance space and let $VC_{dim}(\mathcal{C}) = d$ with $d \in \mathbb{N}$.
This means that the largest subset S of X that is shattered by $\mathcal{C}$ is of size d, i.e. $|S| = d$.
This again means that any(!!!) subset of S can be *realized* by some concept in $\mathcal{C}$.
Since $|S|$ is bounded by d, the number of its subsets is bounded by $|2^S| = 2^{|S|} = 2^d$.
Because for each subset of S there must exist a concept in $\mathcal{C}$ that *realizes* it, we can conclude $|\mathcal{C}| \geq |2^S| = 2^d$,
thus $\log_2 |\mathcal{C}| \geq d$. □

Problem b)

Let $\mathcal{C}_1, \mathcal{C}_2 \subseteq 2^X$ be concept classes over some instance space X.
Show that if $\mathcal{C}_1 \subseteq \mathcal{C}_2$ then $VC_{dim}(\mathcal{C}_1) \leq VC_{dim}(\mathcal{C}_2)$.

Proof:
Let $VC_{dim}(\mathcal{C}_1) = d_1$ with $d_1 \in \mathbb{N}$.
Let S_1 be the largest subset of X that is shattered by $\mathcal{C}_1$.
This means that any subset of S_1 can be *realized* by some concept in $\mathcal{C}_1$.
Since $\mathcal{C}_1 \subseteq \mathcal{C}_2$, S_1 is also shattered by $\mathcal{C}_2$.
So $VC_{dim}(\mathcal{C}_2)$ is obviously bounded by $|S_1| = VC_{dim}(\mathcal{C}_1)$,
thus $VC_{dim}(\mathcal{C}_1) \leq VC_{dim}(\mathcal{C}_2)$. □

Problem c)

Let $\mathcal{C} \subseteq 2^X$ be a concept class over X and let $\bar{\mathcal{C}} := \{X \backslash c : c \in \mathcal{C}\}$. Is it true that $VC_{dim}(\mathcal{C}) = VC_{dim}(\bar{\mathcal{C}})$? Argue why or why not.

Claim:
$VC_{dim}(\mathcal{C}) = VC_{dim}(\bar{\mathcal{C}})$

Proof:
Let S be a subset of X that is shattered by $\mathcal{C}$. Let $T \subseteq S$ be any arbitrary subset of S. Since S is shattered by $\mathcal{C}$, we can find a concept in $\mathcal{C}$ that realizes T, i.e. $T \in \Pi_{\mathcal{C}}(S)$.
Can we also find a concept in $\mathcal{C}$ that realizes T?
To answer this question consider the subset $S \backslash T \subseteq S$ of S. Since S is shattered by $\mathcal{C}$, we can also find a concept $c \in \mathcal{C}$ that realizes $S \backslash T$. Let $\bar{c}$ be defined as

$\bar{c} := X \backslash c$, which lies in $\bar{\mathcal{C}}$. By definition $\bar{c}$ realizes $X \backslash (S \backslash T)$ w.r.t X. Since $S \subseteq X$, we get $T = S \backslash (S \backslash T) \subseteq (X \backslash (S \backslash T))$. So $\bar{c}$ also realizes T w.r.t. S, so $T \in \Pi_{\bar{\mathcal{C}}}(S)$. Since we have chosen $T \in \Pi_{\mathcal{C}}(S)$ arbitrarililiy, we can conclude $\Pi_{\mathcal{C}}(S) \subseteq \Pi_{\bar{\mathcal{C}}}(S)$. Thus, any subset of X that is shattered by $\mathcal{C}$ is also shattered by $\bar{\mathcal{C}}$.
Apply the same thoughts to $\bar{\mathcal{C}}$ and we get, that any subset of X that is shattered by $\bar{\mathcal{C}}$ is also shattered by $\bar{\bar{\mathcal{C}}} = \mathcal{C}$.
This leaves us with $VC_{dim}(\mathcal{C}) = VC_{dim}(\bar{\mathcal{C}})$. □

3 Application of the BEHW-Theorem.

Claim The concept class $\mathcal{C} = \left\{x \in \mathbb{R}^2 \mid a \leq x_1 \leq b,\, c \leq x_2 \leq d\right\}$ consisting of all axis parallel (aligned) rectangles is polynomially PAC-learnable.

Proof By corollary 1 of slide 12 of lecture 2014-12-19, a concept class $\mathcal{C}$ is polynomially PAC-learnable if

1. the VC-dimension of $\mathcal{C}$ is bounded by polynomial of the parameters of $\mathcal{C}$, and
2. the problem of finding a hypothesis consistent with a sample S can be solved in time polynomial in $|S|$ and the parameters of $\mathcal{C}$.

Show (1):

To show that $\mathrm{VC}_{\mathrm{dim}}(\mathcal{C}) \geq 4$, consider the points $\{(1,0),(0,1),(-1,0),(0,-1)\}$ on the unit circle. This set can be shattered by $\mathcal{C}$. All four points are realized by $\left\{x \in \mathbb{R}^2 \mid -m \leq x_1 \leq m,\, -m \leq x_2 \leq m\right\}$ and the empty set is realized by $\left\{x \in \mathbb{R}^2 \mid -m \leq x_1 \leq m,\, -m \leq x_2 \leq m\right\}$ for some $m \geq 1$ and some $m < 1$, respectivley. Single points y (and thus also three points) are realized by $\left\{x \in \mathbb{R}^2 \mid y_1 - \epsilon \leq x_1 \leq y_1 + \epsilon,\, y_2 - \epsilon \leq x_2 \leq y_2 + \epsilon\right\}$ for some $\epsilon < 1$. The case of two adjacent points can be realized by unit squares with a vertex in the origin. The case of two points on the same coordinate axis can be realized by rectangles of sufficiently large extent along the corresponding coordinate axis and sufficiently small extent along the perpendicular axis.

Five and more points can not be shattered by $\mathcal{C}$. If the convex hull of five points form a quadrilateral (or a polygon with less than four vertices), it is not possible to realize the set of the quadrilateral vertices without the point in the interior (or the point on an edge if the remaining point is not in the interior) because the rectangles in $\mathcal{C}$ are convex. If the convex hull of five points forms a pentagon, one may try to realize single points by starting with a sufficiently large rectangle to realize all points. Thereafter, the rectangle is scaled from one edge until a vertex of the pentagon is not realized anymore. Since the rectangle only has four edges, it is impossible to do this procedure individually for all five vertices of a pentagon.

Show (2):

Such an algorithm is given by the FIND_S algorithm:

1. Initialize concept $c = \{x \in \mathbb{R}^2 \mid \infty \leq x_1 \leq -\infty, \infty \leq x_2 \leq -\infty\}$.
2. For all positive training examples $x \in S$ do:
   (a) If $x \in c$, do nothing.
   (b) Otherwise, shift (a, b, c, d) of $c = \{x \in \mathbb{R}^2 \mid a \leq x_1 \leq b, c \leq x_2 \leq d\}$ by the least possible amount to (a', b', c', d') such that
   $$x \in \{x \in \mathbb{R}^2 \mid a' \leq x_1 \leq b', c' \leq x_2 \leq d'\} .$$
   Update $c = \{x \in \mathbb{R}^2 \mid a' \leq x_1 \leq b', c' \leq x_2 \leq d'\}$.
3. Return c.

This algorithm is time constant in the parameter of $\mathcal{C}$ and time linear in $|S|$.

□

4 Application of the BEHW-Theorem.

Let consider the function g(x) which is the compact representation of symmetric boolean function:

$$g : \{y\} \to \{0, 1\}, where\ y \in N, 0 \leq y \leq n$$

Clearly,

$$f(\vec{x}) = g(\sum_{i=1}^{n} x_i) = g(y)$$

Any function g(y) can be represented with a table of size equal to n. Table of size n can perfectly classify any set of inputs of size at most n, therefore we can state that VC(g) <= n. And, as follows, VC(f) <= n.

Now lets introduce binary function h(z).

$$\vec{z} \in 0, 1^j,\ where\ j \in N, j < \sqrt{n} + 1$$

Where vector z is a binary representation of number k, which corresponds to a number of entry of table representation of function g(y).

Function h(z) can be represented in a CNF with at most $j \leq (\sqrt{(}n) + 1)^2$ underlying disjunctions. Let us introduce conjunctive function i(t), where every element of vector t corresponds to a disjunction of a CNF representation of h(z).

$$i(\vec{t})|\ t_i \in \{0, 1\}, i \in [0, (\sqrt{n} + 1)^2]$$

It is already shown, that function i(t) is polynomialy PAC learnable. Learning function i(t) we can reconstruct every presiding function up to f(x) in polynomial time, because every conversion was done in polynomial time.