Solution 1: VC Dimension

Consider a binary classification learning problem with feature space $\mathcal{X} = \mathbb{R}^p$ and label space $\mathcal{Y} = \{-1, 1\}$.

(a) Assume that $p = 1$, i.e., $\mathcal{X} = \mathbb{R}$. Let

$$\mathcal{H} = \{h_r : \mathcal{X} \to \mathcal{Y} \,|\, r \in \mathbb{R}\}$$

be the hypothesis space of left-open interval classifiers on the reals, where $h_r(x) = 1$ for $x \in (-\infty, r]$ and $= -1$ otherwise. What is $VC_p(\mathcal{H})$?

Solution:

Let $x_1 \in \mathbb{R}$ be an arbitrary point. Then, $h_{x_1}(x_1) = +1$ and $h_{x_1-1}(x_1) = -1$. Thus, $\mathcal{H}$ shatters $\{x_1\}$ and we infer that $VC_1(\mathcal{H}) \geq 1$.

Now, let $x_2 \in \mathbb{R}$ be another arbitrary point such that (w.l.o.g.[1]) $x_1 < x_2$. Note that $h_r(x_2) = +1$ implies $h_r(x_1) = +1$. Thus, there is no $h_r \in \mathcal{H}$ such that $(h_r(x_1), h_r(x_2))^\top = (-1, 1)^\top$ (or $(h_r(x_2), h_r(x_1))^\top = (1, -1)^\top$) holds. We infer that $VC_1(\mathcal{H}) < 2$, as two points cannot be shattered by $\mathcal{H}$. With this, we conclude that $VC_1(\mathcal{H}) = 1$.

(b) Let

$$\tilde{\mathcal{H}} = \{\tilde{h}_l : \mathcal{X} \to \mathcal{Y} \,|\, l \in \mathbb{R}\}$$

be the hypothesis space of right-open interval classifiers on the reals, where $\tilde{h}_l(x) = 1$ for $x \in [l, \infty)$ and $= -1$ otherwise. What is $VC_p(\mathcal{H} \cup \tilde{\mathcal{H}})$?

Solution: Let $x_1, x_2 \in \mathbb{R}$ be some arbitrary points such that (w.l.o.g.[1]) $x_1 < x_2$. Note that $\tilde{h}_l(x_1) = +1$ implies $\tilde{h}_l(x_2) = +1$. We can generate every possible assignment $(y_1, y_2)^\top \in \mathcal{Y}^2$ for x_1, x_2 :

$$\begin{aligned}
(-1,-1)^\top &= \left(h_{x_1-1}(x_1), h_{x_1-1}(x_2)\right)^\top, \\
(-1,1)^\top &= \left(\tilde{h}_{\frac{x_1+x_2}{2}}(x_1), \tilde{h}_{\frac{x_1+x_2}{2}}(x_2)\right)^\top, \\
(1,-1)^\top &= \left(h_{x_1}(x_1), h_{x_1}(x_2)\right)^\top, \\
(1,1)^\top &= \left(h_{x_2}(x_1), h_{x_2}(x_2)\right)^\top.
\end{aligned}$$

Thus, $\mathcal{H} \cup \mathcal{H}'$ shatters $\{x_1, x_2\}$ and we infer that $VC_1(\mathcal{H} \cup \mathcal{H}') \geq 2$.

Now, let $x_3 \in \mathbb{R}$ be another arbitrary point such that (w.l.o.g.[1]) $x_2 < x_3$. There is no $h \in \mathcal{H} \cup \mathcal{H}'$ such that $(h(x_1), h(x_2), h(x_3))^\top = (1, -1, 1)^\top$ holds. Indeed, h is either a

- left-open classifier, i.e. $h = h_r$ for some $r \in \mathbb{R}$, so that $h(x_3) = +1$ implies $h(x_2) = +1$,
- right-open classifier, i.e. $h = \tilde{h}_l$ for some $l \in \mathbb{R}$, so that $h(x_1) = +1$ implies $h(x_2) = +1$.

Therefore, we infer that $VC_1(\mathcal{H} \cup \mathcal{H}') < 3$, as three points cannot be shattered by $\mathcal{H} \cup \mathcal{H}'$. With this, we conclude that $VC_1(\mathcal{H} \cup \mathcal{H}') = 2$.

(c) Consider now the feature space $\mathcal{X} = \{0, 1\}^p$ for some $p \in \mathbb{N}$ and let

$$\mathcal{H} = \{h_t : \mathcal{X} \to \mathcal{Y} \,|\, t \in \{0, 1, 2, \ldots, p+1\}\}$$

be the hypothesis space of threshold classifiers on bitstrings, where $h_t(\mathbf{x}) = 1$ for $\sum_{i=1}^p x_i \geq t$ and $= -1$ otherwise. Thus, instances are bitstrings of length p, and h_t classifies an instance as positive if the number of 1s in the bitstring is at least t, e.g., $h_3(0, 1, 1, 0, 0) = -1$ and $h_3(1, 1, 1, 0, 1) = +1$. What is $VC_p(\mathcal{H})$?

Solution: One arbitrary point $\mathbf{x} \in \mathcal{X}$ can be shattered since $h_{d+1} \equiv -1$ and $h_0 \equiv 1$. Therefore, $VC_p(\mathcal{H}) \geq 1$.

[1]Otherwise, relabel the points.

Now define $N_1(\mathbf{x}) = \#\{x_j = 1 \mid \mathbf{x} = (x_1, \ldots, x_n)\}$, which denotes the number of ones in $\mathbf{x} \in \mathcal{X}$. If $N_1(\mathbf{x}) = N_1(\mathbf{x}')$, then $h_t(\mathbf{x}) = h_t(\mathbf{x}'), \forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}, t \in \{0, \ldots, p+1\}$, i.e., the ordering of the zeros resp. ones is not relevant, but only their total number. Thus, the "interesting" candidate points are

$$X_{cand} = \left\{ \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \right\}.$$

But for each $\mathbf{x}, \mathbf{x}' \in X_{cand}$ it holds that $\exists\, t : (h_t(\mathbf{x}), h_t(\mathbf{x}')) = (0, 1)$, then $\forall t' \in \{0, \ldots, p{+}1\}$ $(h_{t'}(\mathbf{x}), h_{t'}(\mathbf{x}')) \neq (1, 0)$. We can show this indirectly, by assuming such a t' exists and distinguish two cases:

(i) $t' < t$, then $h_t(\mathbf{x}') = 1 \implies h_{t'}(\mathbf{x}') = 1$

(ii) $t' \geq t$, then $h_t(\mathbf{x}) = 0 \implies h_{t'}(\mathbf{x}) = 0$

Both implications are contradictions. It can be shown similarly: if for each $\mathbf{x}, \mathbf{x}' \in X_{cand}$ it holds that $\exists\, t : (h_t(\mathbf{x}) = h_t(\mathbf{x}')) = (1, 0)$, then $\forall t' \in \{0, \ldots, p+1\}$ $(h'_t(\mathbf{x}) = h'_t(\mathbf{x}')) \neq (0, 1)$. Hence, $VC_p(\mathcal{H}) < 2$ as two points cannot be shattered by $\mathcal{H}$. In summary, we conclude that $VC_p(\mathcal{H}) = 1$.

(d) Let the feature space be $\mathcal{X} = \mathbb{R}^p$ and let $\mathcal{H}$ be a finite hypothesis space, i.e., $|\mathcal{H}| < \infty$. Show that $VC_p(\mathcal{H}) \leq \log_2(|\mathcal{H}|)$ holds.

Hint: Consider a set of points of size $\log_2(|\mathcal{H}|) + 1$.

Solution: Let $B := \log_2(|\mathcal{H}|) + 1$ and consider B many arbitrary points $\mathbf{x}_1, \ldots, \mathbf{x}_B$. Note that there are 2^B many possible assignments for these points, as each point can be assigned either a $+1$ or a -1. This corresponds to $2^B = 2^{\log_2(|\mathcal{H}|)+1} = 2|\mathcal{H}|$ many possible assignments. In other words, $\mathcal{H}$ should be able to provide all $2|\mathcal{H}|$ many possible assignments in order to shatter the points $\mathbf{x}_1, \ldots, \mathbf{x}_B$.

However, each $h \in \mathcal{H}$ can provide only one assignment $(h(\mathbf{x}_1), \ldots, h(\mathbf{x}_B))^\top \in \mathcal{Y}^B$, which means that **at most** $|\mathcal{H}|$ many different assignments are possible. Thus, $|\mathcal{H}|$ cannot shatter B many points, so that $VC_p(\mathcal{H}) \leq B - 1 = \log_2(|\mathcal{H}|)$.