Probabilistic graphical models, Assignment 3

Siddharth Bhat (20161105)

March 21st, 2020

6.8, Q1:

Monotonicity of VC dimension

Let $\mathcal{H}' \subseteq \mathcal{H}$. Show that $\text{VCdim}(\mathcal{H}') \leq \text{VCdim}(\mathcal{H})$.

Answer

Recall that the definition of VCdimis is that $\text{VCdim}(\mathcal{H})$ is the maximal size of a set $C \subseteq \mathcal{X}$ which can be *shattered* by $\mathcal{H}$.

Expanding the definition of shattering, we get that the $\text{VCdim}(\mathcal{H})$ is the maximal size of *any* set $C \subseteq X$ such that $\mathcal{H}$ restricted to C is the set of all functions from C to $\{0, 1\}$.

Now, If $C \subseteq \mathcal{X}$ is shattered by $\mathcal{H}' \subseteq \mathcal{H}$, then this means that:

$$|\{f|_C : f \in H'\}| = 2^{|C|}$$

Since $\mathcal{H}' \subseteq \mathcal{H}$, we can replace $\mathcal{H}'$ with $\mathcal{H}$ in the above formula to arrive at:

$$|\{f|_C : f \in H\}| = 2^{|C|}$$

So, clearly, $\text{VCdim}(\mathcal{H}') \leq \text{VCdim}(\mathcal{H})$. However, there might be a set that is *larger* than C that can be shattered by $\mathcal{H}$. This lets us get the strict equality $\text{VCdim}(\mathcal{H}) < \text{VCdim}(\mathcal{H})$ in certain cases — that is, we *cannot* assert that $\text{VCdim}(H) \leq \text{VCdim}(H')$. For example, if we choose $\mathcal{H}' = \emptyset$ where $\mathcal{H}$ is a hypothesis class with $\text{VCdim}(\mathcal{H}) = 1$. Then $\text{VCdim}(\emptyset) = 0 < 1 = \text{VCdim}(\mathcal{H})$.

6.8, Q2:

Given a finite domain $\mathcal{X}$, and a finite number $k \leq |\mathcal{X}|$s, find and prove the VC dimension of:

A. Functions that assign 1 to exactly k elements of $\mathcal{X}$

$\mathcal{H} \equiv \left\{ h \in \{0, 1\}^X : \ |\{x : h(x) = 1\}| = k \right\}$.

Solution

VC dimension is 0. For size $|\mathcal{X}| = n$,

if $k = 0$, then we will be able to classify **only** the empty set, and no non-empty set. In this case, the VC dimension is 0. $\square$.

If $k \neq 0$, Then let $S \subseteq \mathcal{X}$ such that S is shattered by $\mathcal{H}$. However, note that we must be able to express the function $f : S \to \{0, 1\}; f(\_) = 0$ using the hypothesis $h_S \in \mathcal{H}$. However, h_S must have $k \neq 0$ entries of 1, which f does not have. Therefore, h_S could not have shattered S. $\square$.

B. Functions that assign 1 to at most k elements of $\mathcal{X}$

Solution

VC dimension is k.

6.8, Q3:

Let $\mathcal{X}$ be the boolean hypercube $\{0, 1\}^n$. We define parity to be:

$$h_I : \mathcal{X} \to \{0, 1\}; h_I((x_1, x_2, \dots, x_n)) \equiv \sum_{i \in I} (x_i) \mod 2.$$

What is the VC dimensions of the set of all parity functions? That is,

$$\mathcal{H}_{parity,n} \equiv \{h_I : I \subseteq \{1, 2, \dots n\}\}$$

Solution

Once again, unwrapping the definition, our hypothesis class can compute the sum modulo 2 of *all of the subsets of* $\vec{x} \in \mathcal{X}$. We need to use this to find the *largest* set $C \subseteq \mathcal{X} \equiv \{0, 1\}^n$ such that $|\mathcal{H}_C| = 2^{|C|}$.

We can interpret elements $(h \in H)$ as a vector $h_I \in \{0, 1\}^n$, where h_I is a vector with 1's at each index $i \in I$, and 0 at other indexes. That is:

$$h_I \in \{0, 1\}^n \qquad h_I[i] \equiv \mathbb{1}[i \in I] = \begin{cases} 1 & i \in I \\ 0 & \text{otherwise} \end{cases}$$

We can reinterpret the function $h_I(x)$ as $h^T x$ where we have a vector space over the galois field GF_2, where $\oplus$ denotes XOR (recall that addition mod 2 is XOR).

$$h_I(x) = \bigoplus_{i \in I} x_i = \bigoplus_{i=1}^{n} \mathbb{1}[i \in I] x_i = \bigoplus_{i=1}^{n} h_I[i] x[i] = h_I^T x$$

Now, we can reinterpret the question of finding the VC dimension as finding the largest collection of vectors $C \subseteq \mathcal{X} = \{0, 1\}^n$ such that the function C_{act} has full image, where the function C_{act} is:

$$
\begin{aligned}
&C_{act} : \mathcal{H} \to \{0,1\}^{|C|} \\
&C_{act} : \{0,1\}^n \to \{0,1\}^{|}C| \\
&C_{act}(h) \equiv (h(c_0), h(c_1), h(c_2), \ldots h(c_n)) \\
&\quad = (h^T c_0, h^T c_1, \ldots, h^T c_n) \\
&\quad = h^T(c_0, c_1, \ldots c_n)
\end{aligned}
$$

If we regard $(c_0, c_1, c_2, \ldots c_n) \subseteq \mathbb{R}^{n \times |C|}$ as a matrix, then we can see that C_{act} is a *linear function*.

Now, if the set C shatters $\mathcal{H}$, then:

1 The function C_{act} will produce every element in $\{0,1\}^{|C|}$

2 the function C_{act} will have full image.

3 This is only possible when the dimension of the domain is less than or equal to the dimension of the range.

4 the largest set that can be shattered is the largest matrix $C \subseteq R^{n \times |C|}$ such that the function C_{act} has full range.

5 Thus, $|C| \leq n$ for C_{act} to have full range.

6 We can achieve $|C| = n$ by picking $C = I_{n \times n}$. In other words, the element c_i will be the ith row of the identity matrix. That is $c_i[j] = \mathbb{1}[i = j] = \begin{cases} 1 & \text{i = j} \\ 0 & \text{otherwise} \end{cases}$. Clearly, this C is shattered since the function C_{act} is the identity function which will produce every single output in $\{0,1\}^{|C|} = \{0,1\}^n = \{0,1\}^{|\mathcal{H}|}$.

7 $|C| = n$ is the largest possible, since C_{act} is a function, and the size of its image is at most the size of the domain. Since the domain $\mathcal{H}$ has 2^n elements, the image too can have at most 2^n elements, which it does when $|C| = n$, since $|\{0,1\}^{|C|}| = 2^{|C|}$.

7.5 $|C| = n$ is the largest possible, since C_{act} is linear. For a linear function to be surjective, we need $Dim(domain) Dim(range)$. Hence, $Dim(Domain) = Dim(\mathcal{H}) = n \geq Dim(range) = Dim(\{0,1\}^{|C|}) = c$. That is, $n \geq |C|$.

Hence, we conclude that VCdim$(\mathcal{H}) = n$.

6.8, Q5:

Let $\mathcal{H}^d$ be the class of axis-aligned bounding boxes in $\mathbb{R}^d$. Show that the VC dimensions of $\mathcal{H}^d$ is $2d$.

Solution

Formally, we have

$$h_{\vec{l},\vec{r}}(\vec{p}) \equiv \begin{cases} 1 & l[i] \leq p[i] \leq r[i] \text{ for all } i \in \{1, 2, \ldots, n\} \\ 0 & \text{otherwise} \end{cases} \qquad \mathcal{H}^d \equiv \left\{ h_{\vec{l},\vec{r}} : \mathbb{R}^d \to \{0,1\} \mid \forall \vec{l}, \vec{r} \in \mathbb{R}^d \right\}$$

We claim that the set of points:

$$\begin{aligned} S &\equiv S^+ \cup S^- \\ S^+ &\equiv \{p[i] = 2; p[j \neq i] = 0 : i \in [d], p \in \mathbb{R}^d\} \\ S^- &\equiv \{p[i] = -2; p[j \neq i] = 0 : i \in [d], p \in \mathbb{R}^d\} \end{aligned}$$

shatters the hypothesis class $\mathcal{H}^d$.

H^d **shatters** S**:** We first show that H^d restricted to S expresses all functions $S \to \{0,1\}$. Note that the set S has $2d$ points. We will consider the d subspaces, indexed by $i \in \{1, \ldots\}$. We will build some notation for this consideration:

$$\begin{aligned} &S[i, pos] \in S; \quad S[i, pos] \equiv p[i] = 2; p[j \neq i] = 0 : p \in \mathbb{R}^d \\ &S[i, neg] \in S; \quad S[i, neg] \equiv p[-i] = -2; p[j \neq i] = 0 : p \in \mathbb{R}^d \\ &S[i, :] \subseteq S; \quad S[i, :] \equiv \{S[i, pos] \cup S[i, neg]\} \end{aligned}$$

For all classifications $c : S \to \{0,1\}$, we will show how to build a hypothesis $BB_v \in \mathcal{H}^d$ such that $BB_c(s) = c(s) \forall s \in S$. For each subset $S[i, :]$, we will build up a bounding box in $BB_c^i \in \mathcal{H}^d$. We will then show that the convex hull of all bounding boxes $conv(BB_c^1, \ldots, BB_c^n)$ is the BB_c we are looking for: $BB_c = conv(BB_c^1, \ldots, BB_c^n)$.

We first describe the BB_c^i:

$$BB_c^i : \mathcal{H}^d \equiv conv\Big(c(S[i, neg]) * S[i, neg]), c(S[i, pos]) * S[i, pos] \Big)$$

That is, we try to cover the points which have $c(S[i]) = 1$ with a convex hull. If a point is not covered, then we will use $0 \times \vec{p} = \vec{0}$. If a point is indeed covered, then we use $1 \times \vec{p} = \vec{p}$ (the point itself). We then take the convex hull of these. Thus, BB_v^i only contains those points in $S[i, :]$ that need to be covered.

Claim 1: each BB_c^i **classifies** $S[i, :]$ **according to c** immediate from construction.

Claim 2: The convex hull of correct classifiers of BB_v^i **classifies** S**:** Consider some point $S[i, pos]$ (a similar argument will hold for $S[i, neg]$). We know from Claim 1 that BB_c^i correctly classifies $S[i, pos]$. The other classifiers BB_c^j cannot influence what happens in the i dimension, since they only attempt to cover the value 0 along dimension i. It is only BB_c^i that can "expand" the cover in dimension i to cover $S[i, pos]$. Hence, the full convex hull will indeed cover the points of interest.

H^d cannot shatter 2d+1:
We are given a set $S \subseteq X. |S| = 2d+1$ into $S[i, pos]$, $S[i, neg]$, $S[\star]$. We define:

$$S[i, pos] \equiv \arg\max_{p \in S} p[i] \quad \text{(point with max. value in } i \text{ dimension)}$$
$$S[i, neg] \equiv \arg\min_{p \in S} p[i] \quad \text{(point with min. value in } i \text{ dimension)}$$
$$S[\star] \equiv s \in S, s \neq S[i, pos], s \neq S[i, neg] \text{ for all } i \in \{1, \ldots, n\} \quad \text{(leftover point)}$$

Note that $S[\star]$ will always exist, since there are $2d$ points that we get from all the $S[i, pos], S[i, neg]$, while $|S| = 2d+1$. We note that the points $S[i, pos], S[i, neg]$ together form the vertices of a bounding box for S.

Now, let $f_S : S \to \{0, 1\}$ be a classifier such that:

$$f : S \to \{0,1\} f_S(s) \equiv \begin{cases} 1 & s = S[\star] \\ 0 & \text{otherwise} \end{cases}$$

Since our hypothesis class consists of bounding boxes, for any $h \in \mathcal{H}$, if the value of h on the vertices of the bounding box must be the same as the interior. However, $f_S(vertices) = 1$, while $f_S(interior) = 0$. Hence, such an f cannot be realised by any $h \in \mathcal{H}$.

Therefore, no set of size $2d+1$ can be shattered.

6.8, Q9:

Let $\mathcal{H}_{si}$ (si for signed interval) be the class of signed intervals. That is: $\mathcal{H} \equiv \{h_{a,b,s} : a \leq b, s = \pm 1\}$ where

$$h_{a,b,s}(x) \equiv \begin{cases} s & a \leq x \leq b \\ -s & \text{otherwise} \end{cases}$$

Solution

We will first show that VCdim($\mathcal{H}_{si}$) is 3 by exhaustive enumeration. We will then show a slicker method, by proving that if a hypothesis space $\mathcal{H}$ has VCdim($\mathcal{H}$) $= n$, then the VC dimension of the space that is $\mathcal{H}' \equiv \mathcal{H} \times 0, 1$ where the $\{0, 1\}$ controls whether we should negate the output of $h \in \mathcal{H}$ will have VCdim($\mathcal{H}'$) $= \mathcal{H}+1$. Now, clearly the above hypothesis class H_{si} is $\mathcal{H}_{interval} \times 0, 1$. We know that VCdim($H_{interval}$) $= 2$, and hence VCdim(H_{si}) $= 3$.

Exhaustive enumeration
Let us consider all possibilities for three points $\{1, 3, 5\}$. We will write down for each subset the classifier to be used, thereby showing that this set is shattered. For a subset, we will need to pick a classifier that has value $+1$ on elements $s \in S$, and has value -1 on elements $s' \notin S$.

$$
\begin{aligned}
&\emptyset \mapsto h_{0,0,1} \\
&\{1\} \mapsto h_{0,2,1} \quad \{3\} \mapsto h_{2,4,1} \quad \{5\} \mapsto h_{4,6,1} \\
&\{1,3\} \mapsto h_{0,4,1} \quad \{3,5\} \mapsto h_{2,6,1} \quad \{1,5\} \mapsto h_{2,3,-1} \\
&\{1,3,5\} \mapsto h_{0,6,1}
\end{aligned}
$$

Hence, the set is shattered.

Consider any set of size 4. For concreteness, we pick the set $\{1,3,5,7\}$. Since we will only make use of the *ordering* of the elements, hence our argument will work for any set of size 4 (and higher). We claim that the subset $\{3,7\} \subseteq \{1,3,5,7\}$ cannot be classified by any hypothesis $h \in H$ correctly. That is, no hypothesis $h \in \mathcal{H}$ can be such that $h(1) = 1, h(3) = -1, h(5) = 1, h(7) = -1$.

This is because every function $h_{a,b,s} \in H$ can change its value twice, when hopping from the boundary of being to the left of (a,b) to entering (a,b), and then again exiting (a,b) from the right:

$$
\begin{aligned}
h_{a,b,s}(x < a) = -s \mapsto h_{a,b,s}(a \leq x \leq b) = s \quad \text{change 1} \\
h_{a,b,s}(a \leq x \leq b) = s \mapsto h_{a,b,s}(x \geq b) = -s \quad \text{change 2}
\end{aligned}
$$

However, in the case outlined above, to detect $\{3,7\}$, we would need to change sign three times: once from $1 \mapsto 3$, once again from $3 \mapsto 5$, and finally from $5 \mapsto 7$.

So, sets of size 4 cannot be shattered by $\mathcal{H}$. For even larger sets, we can concentrate what happens on any 4 elements and replicate the same argument.

Hence, $\text{VCdim}(\mathcal{H}) = 3$.

Augmentation

Let us consider a set $\mathcal{X}$, and a hypothesis class $\mathcal{H} \equiv \{f : \mathcal{X} \to \pm 1\}$. Let $\text{VCdim}(\mathcal{H}) = n$, and $|\mathcal{X}| > 2^n$ (if not, then X is already fully classified by $\mathcal{H}$, and there is no point studying how to make $\mathcal{H}$ stronger).

We will now consider an augmented classifier space $\mathcal{H}' \equiv \mathcal{H} \times \{+1, -1\}$, with the action of elements of $(h, sgn) \in \mathcal{H}'$ being defined as:

$$
act : \mathcal{H}' \to (\mathcal{X} \to \pm 1) \quad act(h, sgn)(x) \equiv h(x) \times sgn
$$

we will often abbreviate $(h,s)(x)$ instead of writing $act(h,s)(x)$. We will now show that $\text{VCdim}(\mathcal{H}') = \text{VCdim}(\mathcal{H}) + 1$. this is the best we can hope for, since VCdim increases logarithmically for sizes in $\mathcal{H}$:

$$
\text{VCdim}(\mathcal{H}') \leq \log_2(|\mathcal{H}'|) = \log_2(2 \times |\mathcal{H}|) \leq 1 + \log_2(|\mathcal{H}|) \overset{\text{at best}}{=} 1 + \text{VCdim}(\mathcal{H})
$$

To show that we can shatter subsets $S \subseteq \mathcal{X}$ such that $|S| = n+1$, pick any subset $S \subseteq X$ of size $n+1$.

For each value $v \in \{+1,-1\}^{|S|} = \{+1,-1\}^{n+1}$, we will need to produce a hypothesis $(h, sgn) = h' \in \mathcal{H}'$ such that $(h, sgn)(S) = v$.

TODO