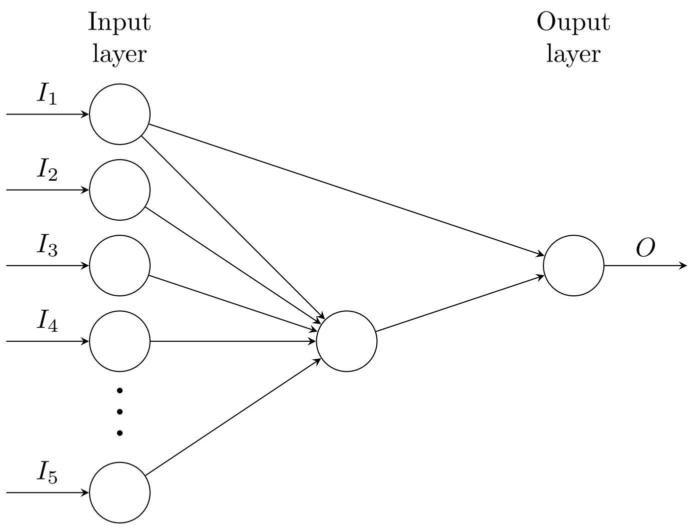
IML Exam

David Ponarovsky

July 2021

1 IML Exam

Q1 ANN VC dim The following net contains n neurons in the input layer and a pair of neurons in the middle layer. One of the neurons in the middle layer is connected to all of the input neurons, while the other is connected only to the first input neuron. Let the activation function be $\sigma_i(x) = sign(x + b_i)$ what is the VC dim of the network?



Q1 Solution the form of hypothesis class is

$$h(\mathbf{x}) = sign\left(w_1^{(2)}x_1 + w_2^{(2)}sign\left(\langle \mathbf{w^{(1)}}, \mathbf{x}\rangle + b_1\right) + b_2\right)$$

where $\mathbf{w^{(1)}}$ and b_1 are the parameters of the middle layer and the $w^{(2}$ is the weight of the synapses which connect the first neuron and the middle neuron to the output neuron. b_2 is the parameter of output layer activation function. Clearly the middle layer is just an hyper plane, therefore it shatters a set of $n+1$ points, denote that set by S. let choose $w_2^{(2)} = \frac{1}{2}$, $|b_2| \leq \frac{1}{4}$ and $w_1^{(2)}$ such that $|w_1^{(2)}| \leq \frac{1}{4 \cdot \max_{x \in S} |x_1|}$, on the one hand for every $x \in S$ we get that $h(x) = sign\left(\langle \mathbf{w^{(1)}}, \mathbf{x}\rangle + b_1\right)$, on the other, we could choose points $\mathbf{x}', \mathbf{y}'$, such that $|\mathbf{x}'_1| = 4 \cdot \max_{x \in S} |x_1|$ and $\mathbf{y}'_1 = 100\frac{b}{w_1^{(2)}} - \mathbf{x}'_1$. let's denote by $z_1, z_1 \in \{0, 1\}$ the output of the middle layer over $\mathbf{x}', \mathbf{y}'$ and by $z_3, z_4 \in \{0, 1\}$

the excepted classification of h, than we could reduce the problem to find a solution for the given equation system:

$$\begin{bmatrix} \mathbf{x'}_1 & 1 \\ -\mathbf{x'}_1 & 100 \end{bmatrix} \begin{bmatrix} w_1^{(2)} \\ b2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \cdot (-1)^{z_1} + \frac{1}{3} \cdot (-1)^{z_3} \\ \frac{1}{2} \cdot (-1)^{z_2} + \frac{1}{3} \cdot (-1)^{z_4} \end{bmatrix} \Rightarrow \begin{bmatrix} w_1^{(2)} \\ b2 \end{bmatrix} = \frac{1}{101\mathbf{x'}_1} \begin{bmatrix} 100 & -1 \\ \mathbf{x'}_1 & \mathbf{x'}_1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \cdot (-1)^{z_1} + \frac{1}{3} \cdot (-1)^{z_3} \\ \frac{1}{2} \cdot (-1)^{z_2} + \frac{1}{3} \cdot (-1)^{z_4} \end{bmatrix}$$

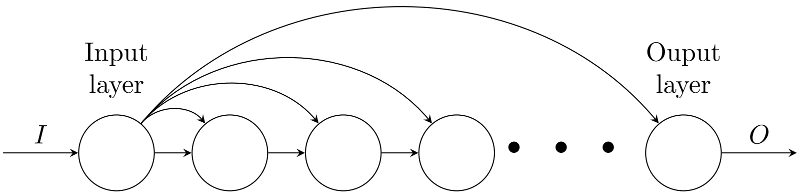
Furthermore, that system has a solution for all the assignments of z_1, z_2, z_3, z_4, which satisfy the constraints.

Contradiction existence of shattered set at size $n + 4$ get by the fact that there is an assignment such h doesn't agree with $sign\left(\langle \mathbf{w^{(1)}}, \mathbf{x} \rangle + b_1\right)$ over at least 3 points, then we get the system:

$$\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ b \end{bmatrix} = - \begin{bmatrix} ay_1 \\ by_2 \\ cy_3 \end{bmatrix}$$

where $a, b, c > 0$ and $y_i \in \{-1, 1\}$ and that system doesn't has solutions.

Q2 ANN VC dim The following net comprises n layers; each contains only a single neuron. A synapse connects every adjacent neuron pair, and they are all connected to the first one. Let the activation function be $\sigma_i(x) = sign(x)$ what is the VC dim of the network?



Q2 Solution Proof by induction, suppose that for a chain of $n-1$ neurons the VC dim of the net h_{n-1} is $n-1$, and denote by S_{n-1} the set which is shattered by the class, than, we could choose w_n such that $|w_n \cdot \max_{x \in S_{n-1}} x| \leq \frac{1}{2}$, of course that for every $x \in S_{n-1}$ we get that

$$h_n(x) = sign(w_n \cdot x + h_{n-1}(x)) = h_{n-1}(x)$$

also we could choose x' such that $|x'| \geq |\frac{4}{w_n}|$ and therefore, we could determine the sign of w_n such that $h_{n+1}(S_{n-1} \cup \{x'\})$ will match the given classification. therefore $S_n = S_{n-1} \cup \{x'\}$ is a set of size n which shattered by the class.

The Contradiction of the existence of shattered set at size $n+1$ is given by the reverse reduction; the last neuron could not add more than 1 for the total VC dim, therefore there is must exist a shattered set at size n for the $n-1$ neuron chain. repeating until $n = 1$ leads to a contradiction.

Q3 Periodic function VC dim Consider the Rectangle function over the segment $[0, T]$ as

$$r(x) = \begin{cases} 1 & x \in [-\frac{T}{2}, \frac{T}{2}] \\ -1 & else \end{cases}$$

And define the Periodic Rectangle $r : \mathbb{R} \to \{0,1\}$ be an extension of the rectangle over $\mathbb{R}$ such that $r(x+T) = r(x)$. For each of the following mark whether or not the hypothesis class

$$\mathcal{H} = \left\{ h_w(x) = r(w \cdot x) \,|w \in \mathbb{R} \right\}$$

shatters the following samples.

Q3 Solution First note that h_w is an even function, therefore, each assignment of the form $\{(X_1, Y_1), (X_2, Y_2)\}$ where $X_1 \subset \mathbb{R}^+$ and $X_2 \subset \mathbb{R}^-$ is equivalence to assignment of the form $\{(-X_1, Y_1), (X_2, Y_2)\}$. So first, it's drop all the samples which can't be classified by an even function.

Lemma The VC dim of Periodic function f with Periodic T over the positive axis $\mathbb{R}^+$ is ∞

Proof: Assume that (X, Y) is a sample where X contains only rational positive numbers $\mathbb{Q}^+$.Assume by induction, that there are frequencies w_1, w_2 such that h_{w_1} classify the first $\frac{n}{2}$ points and h_{w_2} classify the other half. we will show that one can find a frequency w' such that $h_{w'}$ classify $\{X, Y\}$ i.e:

$$w'x_i + m_i \cdot T = \begin{cases} w_1 x_i & i \leq k \\ w_2 x_i & else \end{cases}$$

Let add the following constrain $\sum_0^{2k} x_i m_i = 1$ and then we will get the next matrix:

$$\overbrace{\begin{bmatrix} 0 & x_1 & x_2 & x_3 & \cdots \\ x_1 & T & & & \\ x_2 & & T & & \\ x_3 & & & T & \\ \vdots & & & & \ddots \end{bmatrix}}^{A} \cdot \begin{bmatrix} w' \\ m_1 \\ m_2 \\ m_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ w_1 x_1 \\ \vdots \\ w_2 x_k \\ \vdots \end{bmatrix}$$

We could assume without loose a generality that the points sample from the domain $[0, T]$ (otherwise, we could project the points in the beginning), it follows that, A is PSD matrix:

$$\begin{aligned} u^T A u = & \sum_{ij} a_{ij} u_i u_j = \\ & \sum a_{ii} u_i^2 + a_{jj} u_j^2 + a_{ij} u_i u_j = \\ & \sum_{i,j \neq 0} T\left(u_i^2 + u_j^2\right) + \sum_{i=0} 2 a_{ij} u_i u_j \geq 0 \end{aligned} \tag{1}$$

Therefore, the matrix is invertible and $m_i = \frac{p_i}{q_i} \in \mathbb{Q}$ (because $A \in \mathbb{M}(\mathbb{Q})$). lets choose a numbers $M, \alpha \in \mathbb{N}$ such that $\alpha\left(\prod q_i\right) - 1 = M \cdot T$ and assign $w' \leftarrow \alpha\left(\prod q_i\right) w'$ and we will get that for every

x_j which classified by h_{w_1}, it is also classifed by $h_{w'}$:

$$\begin{aligned} h_{w'}\left(x_j\right) = sign\left(f\left(\alpha \prod q_i w' x_j\right)\right) = \\ sign\left(f\left(\alpha \prod q_i \left(w_1 x_j + m_j T\right)\right)\right) = sign\left(f\left(\left(\alpha \prod q_i\right) w_1 x_j\right)\right) = \\ sign\left(f\left(\left(\alpha \prod q_i - 1\right) w_1 x_j + w_1 x_j\right)\right) = sign\left(f\left(w_1 x_j\right)\right) = \\ h_{w_1}\left(x_j\right) = y_j \end{aligned} \tag{2}$$

At the same way, it can be shown that $h_{w'}(x_j) = h_{w_2}(x_j)$ for every x_j which classified by h_2. for the completeness of the proof, Its left to show that if a class of continues function shatter a set at size n over the rational numbers then it's also shatters a set at size n over the real numbers.