Department of Computer Science
Ashoka University

Introduction to Machine Learning: CS-3410-1

Assignment 2

Name: Dhruman Gupta

---

1 Question 1

Done on jupyter notebook.

2 Question 2

Done on jupyter notebook.

3 Question 3

Done on jupyter notebook.
Question 4 is done next page onwards.

4 Do you even shatter?

4.1 $1\{a < x\}$

The VC dimension of this hypothesis class is 1. Given an $x \in \mathbb{R}$, the hypothesis $h(x) = 1\{a < x\}$ can shatter the set $\{x\}$ by choosing $a < x$ or $a > x$ depending on the label of x.
However, if we have 2 points, such that $x_1 < x_2$, where x_1 is labelled 1 and x_2 is labelled 0, then any choice of a will result in the other point being misclassified. Thus, the VC dimension is 1.

4.2 $1\{a < x < b\}$

The VC dimension of this hypothesis class is 2. Given two points $x_1 < x_2$, the hypothesis $h(x) = 1\{a < x < b\}$ can shatter these two points:

- For labelling $(1, 1)$, we can choose $a < x_1 < x_2 < b$.
- For labelling $(0, 0)$, we can choose $x_1 < a < b < x_2$.
- For labelling $(1, 0)$, we can choose $a < x_1 < b < x_2$.
- For labelling $(0, 1)$, we can choose $x_1 < a < x_2 < b$.

Now, say we have 3 points, $x_1 < x_2 < x_3$. Given the labelling $(1, 0, 1)$, we can see that no choice of a and b will result in all the points being labelled correctly. Thus, the VC dimension is 2.

4.3 $1\{a \sin(x) > 0\}$

The VC dimension of this hypothesis class is 1. Pick any x_1 with $\sin x_1 \neq 0$. To label 1, we can choose $a = \frac{\sin x_1}{|\sin x_1|}$. This will label x_1 as 1. If the label is 0, we can choose $a = -\frac{\sin x_1}{|\sin x_1|}$. This will label x_1 as 0.

Now, consider two points. If the signs of $\sin x_1$ and $\sin x_2$ are the same, then the labelling $(1, 0)$ or $(0, 1)$ cannot be achieved. If the signs are different, then the labelling $(1, 1)$ cannot be achieved. Thus, the VC dimension is 1.

4.4 $1\{\sin(x + a) > 0\}$

The VC dimension of this hypothesis class is 2. Given two points $x_1 < x_2$, the hypothesis $h(x) = 1\{\sin(x + a) > 0\}$ can shatter these two points:

- For labelling $(1, 1)$, we can choose a such that $\sin(x_1 + a) > 0$ and $\sin(x_2 + a) > 0$.
- For labelling $(0, 0)$, we can choose a such that $\sin(x_1 + a) < 0$ and $\sin(x_2 + a) < 0$.
- For labelling $(1, 0)$, we can choose a such that $\sin(x_1 + a) > 0$ and $\sin(x_2 + a) < 0$.
- For labelling $(0, 1)$, we can choose a such that $\sin(x_1 + a) < 0$ and $\sin(x_2 + a) > 0$.

As long as $x_1 - x_2 \mod \pi \equiv 0$ is not true, we can always find an a such that the labelling is achieved. Since for any x_1, x_2 where the above is not true, we can shatter them, the VC dimension is 2.

Consider 3 points, x_1, x_2, x_3. Calculate $\theta_1, \theta_2, \theta_3$ as $\theta_i = x_i \mod 2\pi$. Now, WLOG, assume $\theta_1 \leq \theta_2 \leq \theta_3$. With this ordering (which will be there for every single point), the labelling $(1, 0, 1)$ cannot be achieved. Because if $\sin(a + x_1) > 0$, and $\sin(a + x_3) > 0$, then it must be the case that $\sin(a + x_2) > 0$ too. This comes from the fact that sin is continous, and x_2 is sandwiched between x_1 and x_3. So, $\sin x_2$ cannot be lesser than 0. Thus, the VC dimension is 2.

5 Question 5

In this case, we can infact shatter all 3 options (I almost tried with a compass to verify this).

What this tells me is that the VC dimension is at least 4 - as it can shatter a set of 4 points.