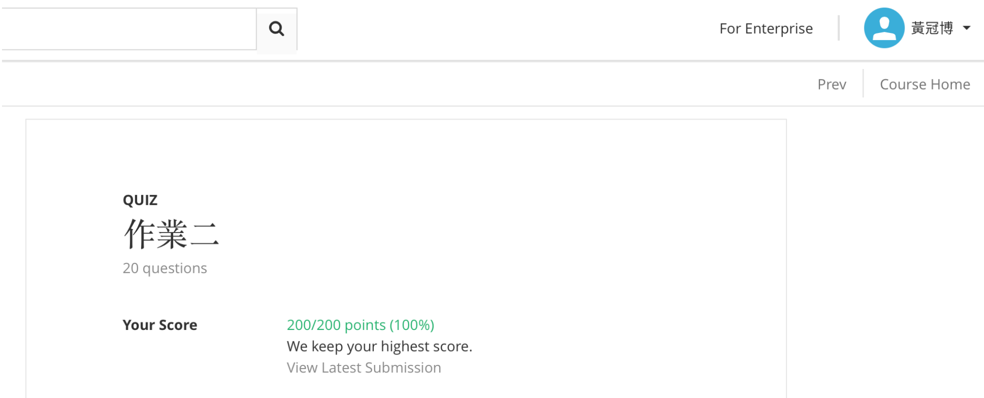
Machine Learning Foundations
Homework 2

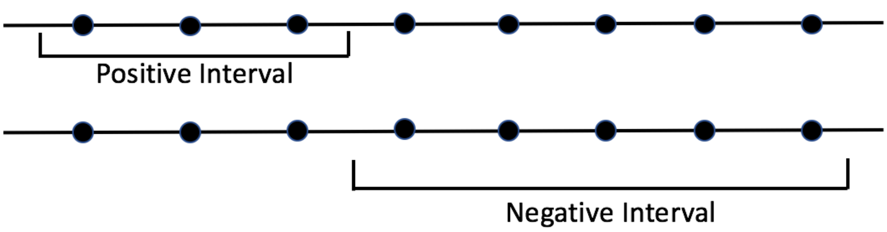
b05902121 黃冠博

December 11, 2017

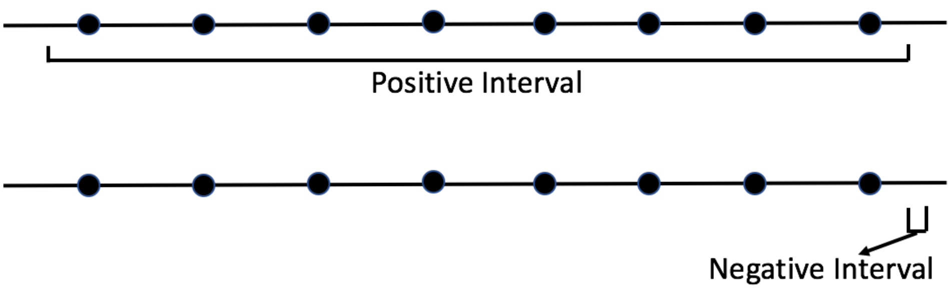
1.



2. For the "positive interval problem", we line up the N dots and there are $N+1$ spaces between them. Pick any two spaces and a positive interval forms. Hence, there are C_2^{N+1} combinations which equal to $\frac{N^2}{2}+\frac{N}{2}$. However, we didn't consider the situation if we pick the same space which results to every dot to be X. Therefore, for the "positive interval problem", $m_H(N)=\frac{N^2}{2}+\frac{N}{2}+1$. Similarly, for the "negative interval problem", $m_H(N)=\frac{N^2}{2}+\frac{N}{2}+1$, too. Nevertheless, these two problems have overlap situations. Consider the situation below, they are actually the same. We can switch the positive and negative intervals. Hence, there are $2(N-1)$ of them.



The situation below is actually the same, too. We can switch the positive and negative intervals. There are 2 of them.



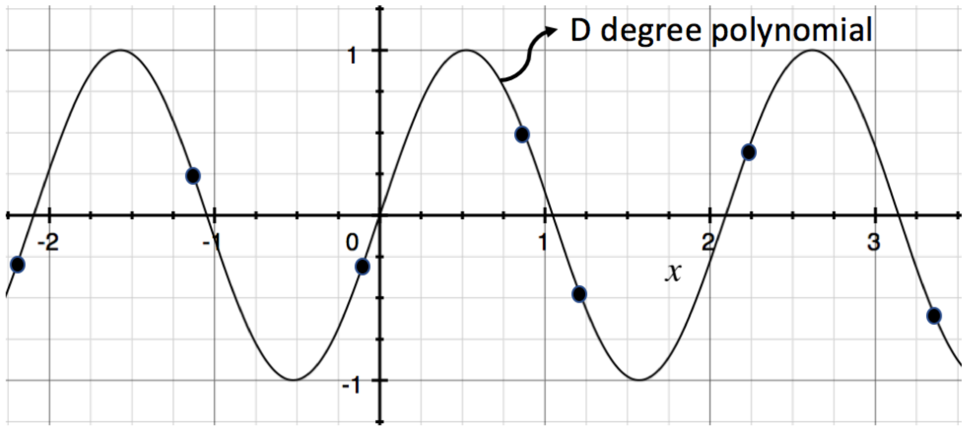
Hence, $m_H(N)=(\frac{N^2}{2}+\frac{N}{2}+1)\times 2-2(N-1)-2=N^2-N+2$

3. The answer for this problem is $d_{vc}(H) = D + 1$.

To proof this, we have to say that $d_{vc} \leq D + 1$ and $d_{vc} \geq D + 1$.

For $d_{vc} \leq D + 1$:

Assume $d_{vc} = D + 2$, we pick $D + 2$ dots like the figure below.



By Blzano's Theorem(勘根定理), we know that there are $D + 1$ roots for this D degree polynomial, which is a contradiction because a polynomial with degree D can only have D roots. Hence, $d_{vc} \leq D + 1$.

For $d_{vc} \geq D + 1$:

Assume there are $x_1, x_2, \cdots, x_{D+1}$, and now given $D + 1$ pairs of (x, y) with $y = \sum_{i=0}^{D} c_i x^i$ as shown below.

$$y_1 = \sum_{i=0}^{D} c_i x_1^i$$

$$y_2 = \sum_{i=0}^{D} c_i x_2^i$$

$$\vdots$$

$$y_{D+1} = \sum_{i=0}^{D} c_i x_{D+1}^i$$

By Lagrange polynomial, given $D + 1$ pairs of (x, y), we can always find a polynomial with degree D that satisfies it. $N = D + 1$ can be shattered. Hence, $d_{vc} \geq D + 1$.

4. We will do a simple analysis to the hypothesis $h_\alpha(x) = sign(|\alpha x \mod 4 - 2| - 1)$.
If we want $h_\alpha(x) = 1$:

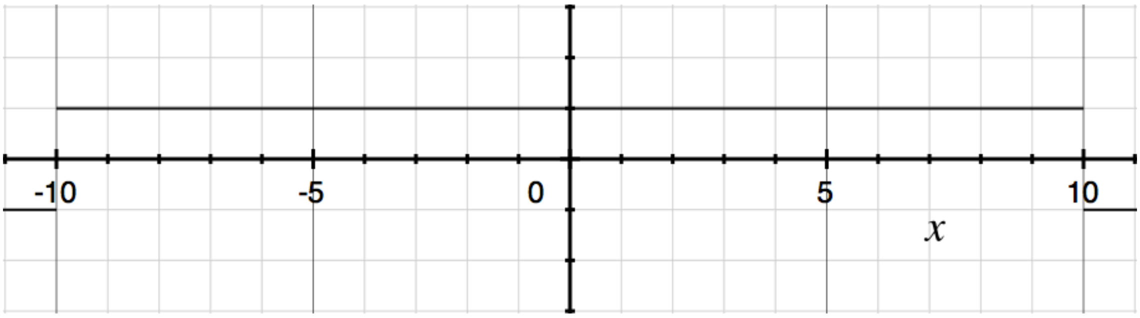
$$\Rightarrow |\alpha x \mod 4 - 2| - 1 \geq 0 \;\Rightarrow |\alpha x \mod 4 - 2| \geq 1 \;\Rightarrow (\alpha x \mod 4) \geq 3 \;or\; (\alpha x \mod 4) \leq 1$$

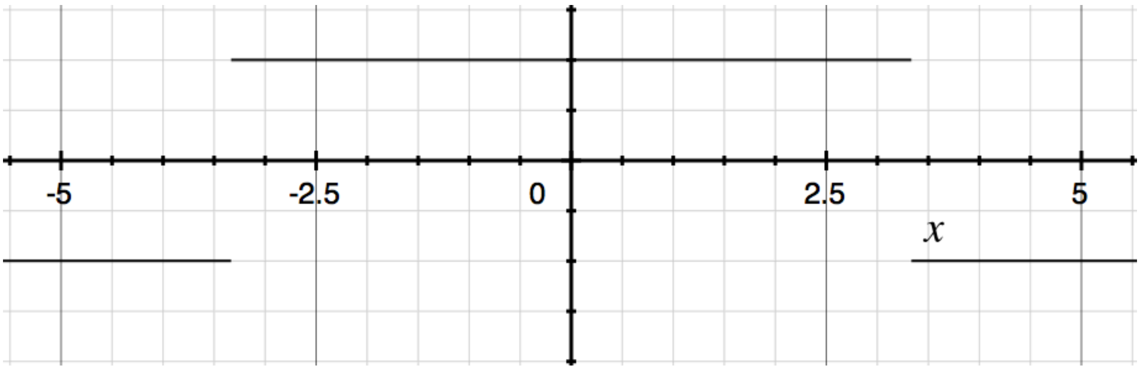
Let x_i be $4^0, 4^1, 4^2, \cdots$, for $i = 1$ to ∞

For the i^{th} point,

$$\frac{1}{2}(\frac{2k-2}{2^{N+1}}) < \alpha_k < \frac{1}{2}(\frac{2k-1}{2^{N+1}})$$

with $k = 1$ to 2^N, α_k can build all $(\pm1, \pm1, \pm1, \cdots)$.
For example, $N = 1$, $0 < \alpha_1 < \frac{1}{8}$ and $\frac{2}{8} < \alpha_2 < \frac{3}{8}$, take $\alpha_1 = 0.1$ and $\alpha_2 = 0.3$. As shown below, $h_{\alpha=0.1}(x_1) = +1$, $h_{\alpha=0.3}(x_1) = -1$.

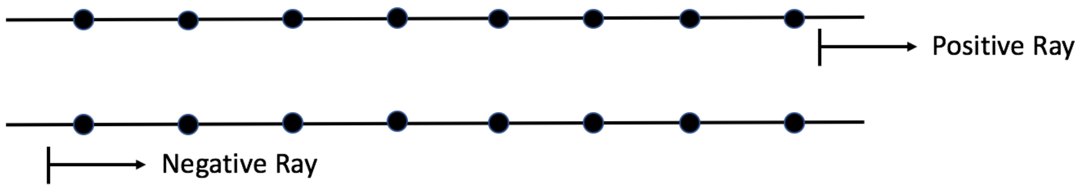




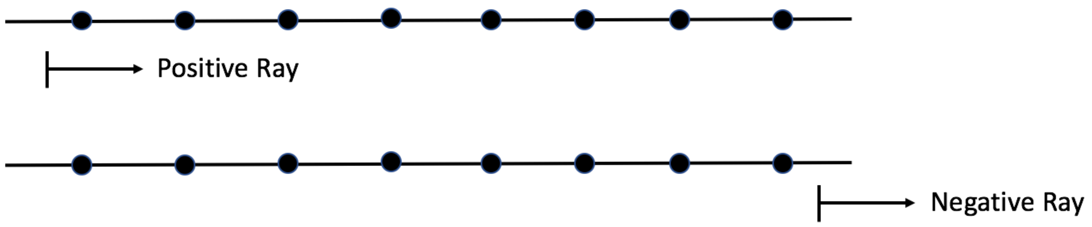
5. Assume that $d_{vc}(H_1) = n$. This means for some set $|S| = n$, hypothesis H_1 shatters it. Since $H_1 \subseteq H_2$, hypothesis H_2 also shatters S. Hence, $d_{vc}(H_1) = n \leq d_{vc}(H_2)$.

6. The positive-ray hypothesis set H_1 and the negative-ray hypothesis set H_2 do not overlap except for the two situations below:

all negative:

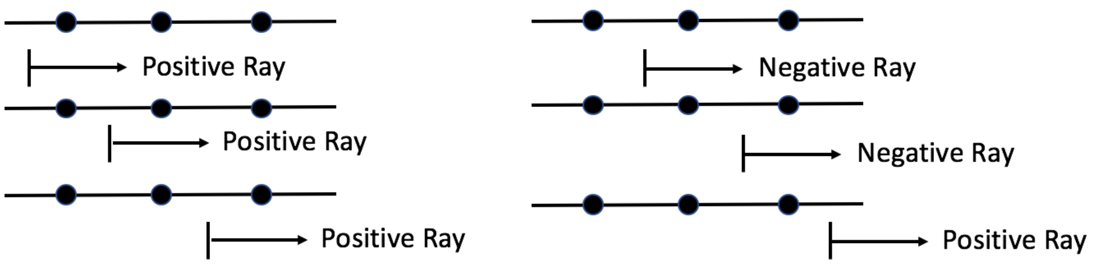


all positive:



Hence, $m_{H_1 \cup H_2}(N) = (N+1) + (N+1) - 2 = 2N$

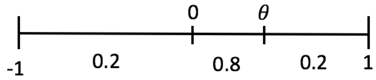
When $N = 3$, as seen below, there are only $6 \neq 2^3$.



Hence, $d_{vc}(H_1 \cup H_2) = 2$

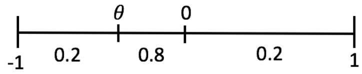
7. For s and θ, there are four scenarios to discuss.

1. s = 1, $\theta > 0$,



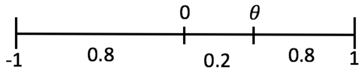
$$E_{out}(h_{s,\theta}) = \frac{1}{2} \cdot 0.2 + \frac{\theta}{2} \cdot 0.8 + \frac{1-\theta}{2} \cdot 0.2 = 0.2 + 0.3\theta$$

2. s = 1, $\theta < 0$,



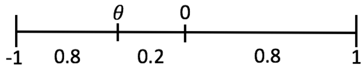
$$E_{out}(h_{s,\theta}) = \frac{\theta-(-1)}{2} \cdot 0.2 + \frac{0-\theta}{2} \cdot 0.8 + \frac{1}{2} \cdot 0.2 = 0.2 - 0.3\theta$$

3. s = -1, $\theta > 0$,



$$E_{out}(h_{s,\theta}) = \frac{1}{2} \cdot 0.8 + \frac{\theta}{2} \cdot 0.2 + \frac{1-\theta}{2} \cdot 0.8 = 0.8 - 0.3\theta$$

4. s = -1, $\theta < 0$,

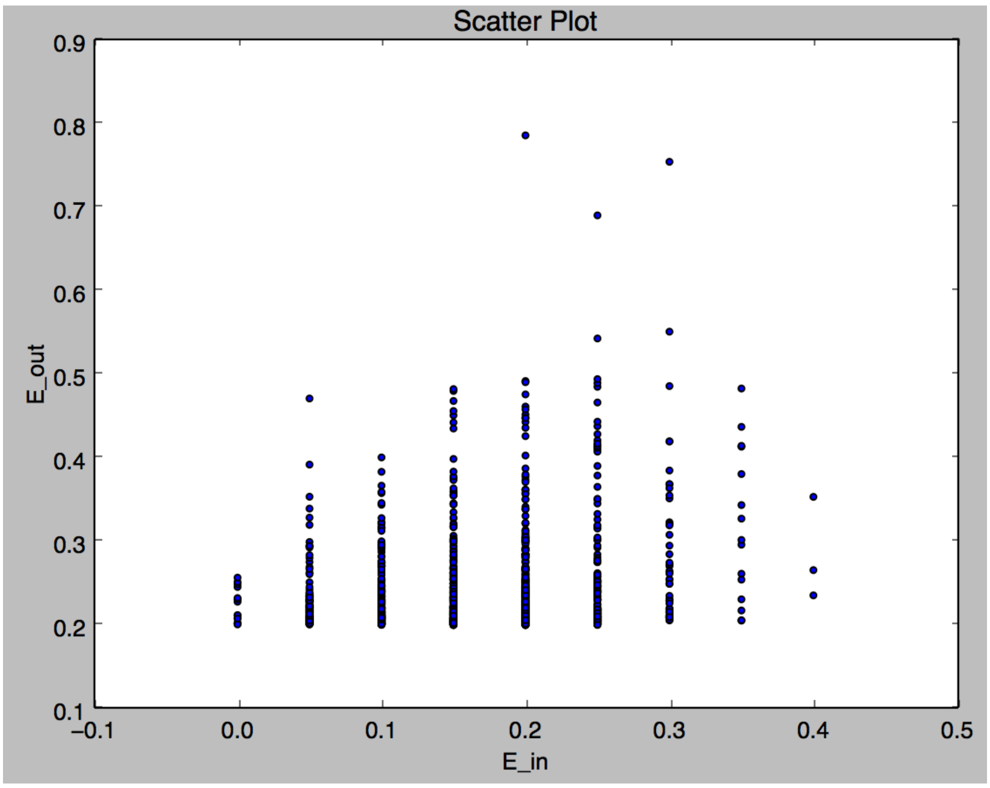


$$E_{out}(h_{s,\theta}) = \frac{\theta-(-1)}{2} \cdot 0.8 + \frac{0-\theta}{2} \cdot 0.2 + \frac{1}{2} \cdot 0.8 = 0.8 + 0.3\theta$$

$\Rightarrow s = 1,\ E_{out}(h_{s,\theta}) = 0.2 + 0.3|\theta|$ and $s = -1,\ E_{out}(h_{s,\theta}) = 0.8 + 0.3|\theta|$

And we can generate a general formula $0.5 + 0.3s(|\theta| - 1)$.

8.



E_{in} mostly falls in $0.05 \sim 0.35$, this means for a hypothesis, the number of data with $h_{s,\theta}(x) \neq y$ is about $1 \sim 7$.
The three points on the top means E_{out} is large. This means their s and θ forms a bad hypothesis.

9. There are some lemmas we need to know first.

Lemma 1 $(1+x)^N = C_0^N + C_1^N x^1 + \cdots + C_N^N x^N$

Lemma 2 $C_{k-1}^{N-1} + C_k^{N-1} = C_k^N$

$$2\sum_{i=0}^{d} C_i^{N-1} = 2(C_0^{N-1}+C_1^{N-1}+\cdots+C_d^{N-1}) = 2[(C_0^{N-1}+C_1^{N-1})+(C_2^{N-1}+C_3^{N-1})+\cdots+(C_{d-1}^{N-1}+C_d^{N-1})] \quad (1)$$

Apply Lemma 2, equation (1) then becomes

$$2[(C_0^{N-1}+C_1^{N-1})+(C_2^{N-1}+C_3^{N-1})+\cdots+(C_{d-1}^{N-1}+C_d^{N-1})] = 2(C_1^N + C_3^N + \cdots + C_d^N) \quad (2)$$

Substitute x with -1 in Lemma 1, and we derive:

$$(1+(-1))^N = 0 = C_0^N - C_1^N + C_2^N \cdots C_N^N$$

Hence, we know that:

$$C_0^N + C_2^N + C_4^N \cdots = C_1^N + C_3^N + C_5^N \cdots$$

Equation (2) then becomes:

$$2(C_1^N + C_3^N + \cdots + C_d^N) = (C_0^N + C_2^N + C_4^N \cdots) + (C_1^N + C_3^N + C_5^N \cdots) = \sum_{i=0}^{d} C_i^N \quad (3)$$

In class, we already know that:

$$B(N,k) = \sum_{i=0}^{k-1} C_i^N \quad (4)$$

Compare equation (3) and (4), we can know that:

$$m_H(N) = 2\sum_{i=0}^{d} C_i^{N-1}$$