1.

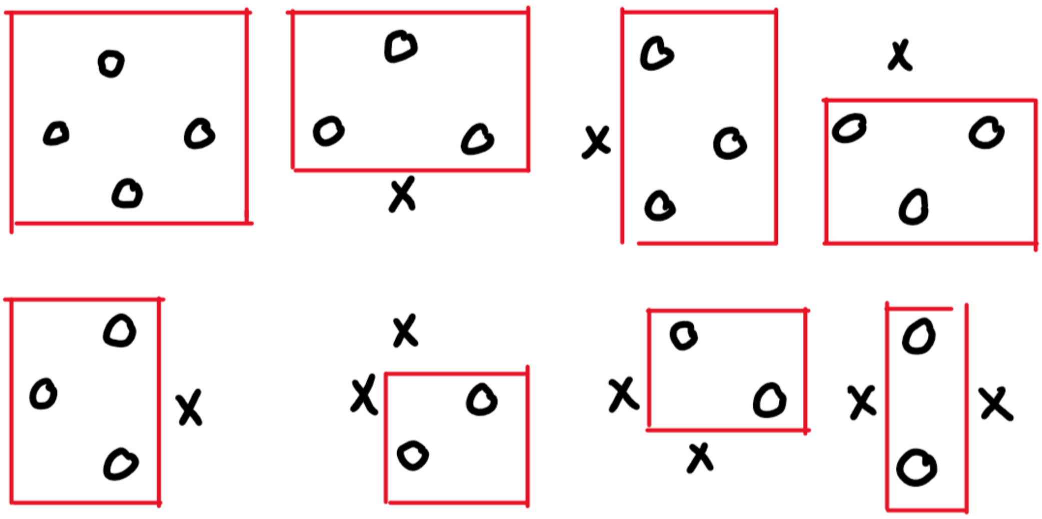
作業二 測驗	已通過	12月16日 14:59 CST	50%	**100%**

2.

相當於證明$m_H(4) = 2^4 = 16$

下圖為其中 8 個狀況，將 O,X 顛倒後，以相同框法可得相反的另外 8 個

得證



3.

$d_{vc}(\mathrm{H}) = \infty$

對於 N 個點，取 $\mathbf{x_i = 10^i}$, for i = 1…N

假設這 N 個 example 對應的 label 分別為 y_i ($y_i \epsilon \{-1,+1\}$,for i = 1…N)

則取 $\boldsymbol{\alpha = 0.z_1z_2\ldots z_N}$

即可滿足 $h_\alpha(x_i) = y_i$ (for i = 1…N)

其中 y 與 z 的關係為:

當 y_i=+1 時，z_i 為 0

當 y_i= -1 時，z_i 為 2

例如: y_1=-1, y_2=+1, y_3=+1,y_4=-1 則取 α=0.2002

因此永遠可以找到一組 2^N 個不同 α 值的 hypothesis 組成一個 hypothesis set

shatter N 個點

Poof:

首先觀察 $h_\alpha(x)=\mathrm{sign}(|\alpha x \bmod 4 - 2| - 1)$可以發現

當 $h_\alpha(x)$=+1 時:

αx mod 4≤1 或 αx mod 4≥3

當 $h_\alpha(x)$=-1 時:

1 < αx mod 4 < 3

對於 N 個點，取 $\mathbf{x_i = 10^i}$, for i = 1…N

討論 $h_\alpha(x)=y$ 是否都成立

①x_1:

a. $y_1=+1$,則 $\alpha= 0.0$

   $\alpha*x_1 \bmod 4 = 0*10^1 \bmod 4 = 0$

   故 $h_\alpha(x_1)=+1= y_1$

b. $y_1=-1$,則 $\alpha= 0.2$

   $\alpha*x_1 \bmod 4 = 0.2*10^1 \bmod 4 = 2$

   故 $h_\alpha(x_1)=-1= y_1$

②x_i, i=2…N:

a. $y_i=+1$,則 $\alpha= 0.z_1z_2…z_{i-1}0z_{i+1}…z_N$

   $\alpha*x_i \bmod 4 = 0.z_1z_2…z_{i-1}0z_{i+1}…z_N *10^i \bmod 4$

   $= z_1z_2…z_{i-1}0.z_{i+1}…z_N \bmod 4$

   $=z_{i-1}0.z_{i+1}…z_N \bmod 4$ (因為 100 mod 4 = 0,所以百位數以上不影響)

   $= 0.z_{i+1}…z_N$ (不管 z_{i-1} 是 0 還是 2,都是此結果)

   故 $h_\alpha(x_i)=+1= y_i$

b. $y_i=-1$,則 $\alpha= 0.z_1z_2…z_{i-1}2z_{i+1}…z_N$

   $\alpha*x_i \bmod 4 = 0.z_1z_2…z_{i-1}2z_{i+1}…z_N *10^i \bmod 4$

   $= z_1z_2…z_{i-1}2.z_{i+1}…z_N \bmod 4$

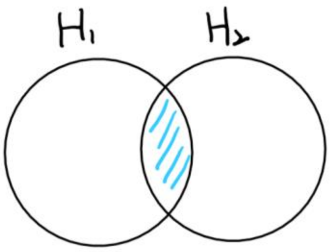
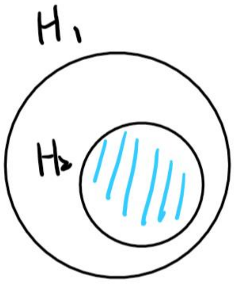
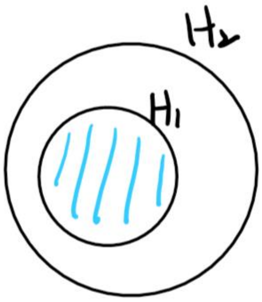
$=z_{i-1}2.z_{i+1}\ldots z_N \bmod 4$ (因為 100 mod 4 = 0,所以百位數以上不影響)

$= 2.z_{i+1}\ldots z_N$ (不管 z_{i-1} 是 0 還是 2,都是此結果)

故 $h_\alpha(x_i)=-1= y_i$

由於 N 可以是任意大於 0 的正整數，又對於任意 N 值都能從 H 中找到一組 hypothesis subset 來處理 2^N 種(因為 binary classification)可能的 label 的組合，故$d_{vc}(\mathrm{H}) = \infty$

根據題意，H1 及 H2 間可能的關係有三種(交集不為空集合):

Case1		此時$(H1 \cap H2) \subset H1$， 故必滿足$d_{vc}(\mathrm{H1} \cap \mathrm{H2}) \leq d_{vc}(H1)$
Case2		此時$(H1 \cap H2) \subset H1$， 故必滿足$d_{vc}(\mathrm{H1} \cap \mathrm{H2}) \leq d_{vc}(H1)$
Case3		此時$(H1 \cap H2) = H1$， 故必滿足$d_{vc}(\mathrm{H1} \cap \mathrm{H2}) \leq d_{vc}(H1)$

在所有可能的情況下都滿足$d_{vc}(\mathrm{H1} \cap \mathrm{H2}) \leq d_{vc}(H1)$, 故得証

5.

$M_{H1}(N) = N + 1$ 且 $M_{H2}(N) = N + 1$

然而當所有點皆為+1 或-1 時，發生重疊的狀況

故 $M_{H1 \cup H2}(N) = (N+1)+(N+1)-2 = 2N$

根據上式:

N=1 時, $M_{H1 \cup H2}(1) = 2*1 = 2 = 2^1$

N=2 時, $M_{H1 \cup H2}(2) = 2*2 = 4 = 2^2$

N=3 時, $M_{H1 \cup H2}(3) = 2*3 = 6 < 2^3$

故 $d_{vc}(H1 \cup H2) = 2$

6.

可以把 $E_{out}(h_{s,\Theta})$以下式表達:

$E_{out}(h_{s,\Theta})$ = λ*μ + (1-λ)*(1-μ)

其中λ及μ定義同 Coursera HW2 的第一題

由題目敘述已知 noise 的比例為 0.2，故可知 1-λ為 0.2

另外本題的 input 分布為 uniform distribution in [−1, 1]

討論所有 s 與Θ的關係(共 4 種 case)

Case1	s = 1, Θ>0	μ=|Θ|/2
Case2	s = 1, Θ<0	μ=|Θ|/2
Case3	s = -1, Θ>0	μ= (2-|Θ|)/2
Case4	s = -1, Θ<0	μ= (2-|Θ|)/2

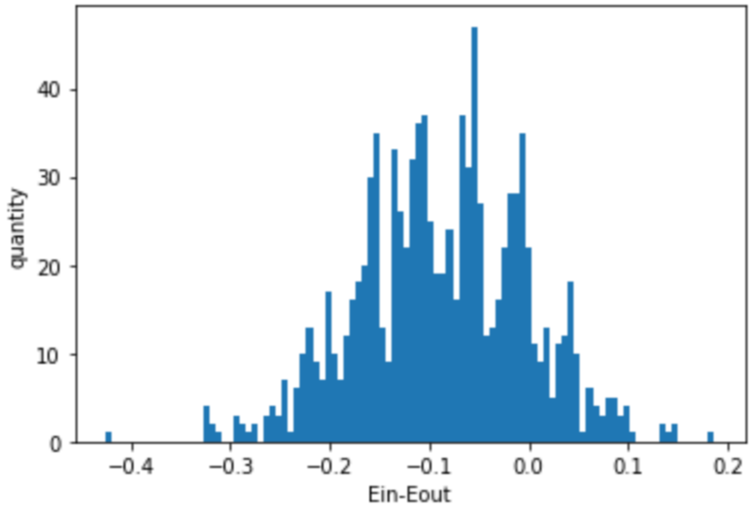
綜合上面 4 個 case 的結果

可得 μ=((s+1)/2)*(|Θ|/2) - ((s-1)/2)*((2-|Θ|)/2) = (2s|Θ|-2s+2)/4

代入 $E_{out}(h_{s,\Theta})$ = λ*μ + (1-λ)*(1-μ)

得 $E_{out}(h_{s,\Theta})$ = 0.3*s*|Θ|-0.3s+0.5

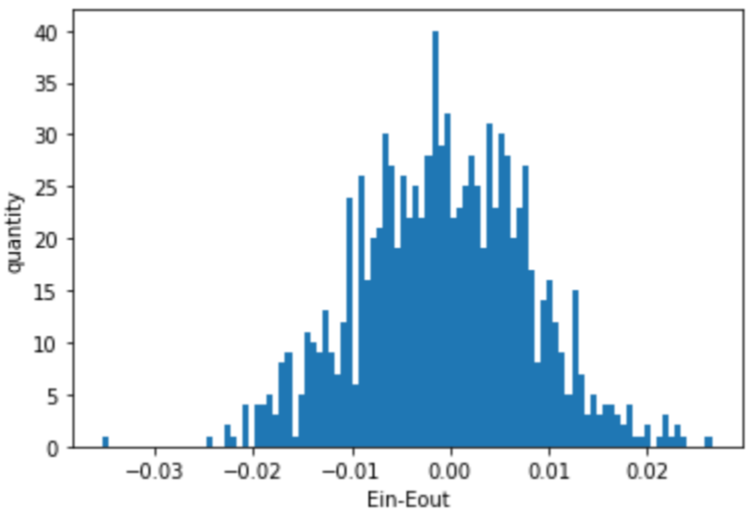
7.



由於只有 20 筆資料的關係，Ein 及 Eout 的分布較為分散，但可以看的出來

Eout 之值通常比較大

8.



本題的資料量遠大於第 7 題，Ein 與 Eout 基本上都落在 0.2 左右，故差值不大

9.

由題目敘述可知 H 中的每個 hypothesis 可以把 R^d 空間分成 2^d 個 hyper-rectangular regions, 並決定這 2^d 個 region 的 label(+1 or -1)

試證①、②

① "simplified decision trees'' hypothesis set 的 VC-dimension **>= 2^d**

Proof: 將 2^d 個點分別至於 2^d 個不同 region 中，則此 2^d 個點可以被 shatter

② "simplified decision trees'' hypothesis set 的 VC-dimension < **2^d+1**

Proof: 將 2^d+1 個點置於 2^d 個不同 region 中，根據鴿籠定理必定會有一個 region 中含有兩個點，當此兩點的 label 不同時(一為+1，一為-1)即無法 shatter

由於"simplified decision trees'' hypothesis set 的 VC-dimension **>= 2^d** 且 "simplified decision trees'' hypothesis set 的 VC-dimension < **2^d+1**，因此 **"simplified decision trees'' hypothesis set 的 VC-dimension 為 2^d**，得証