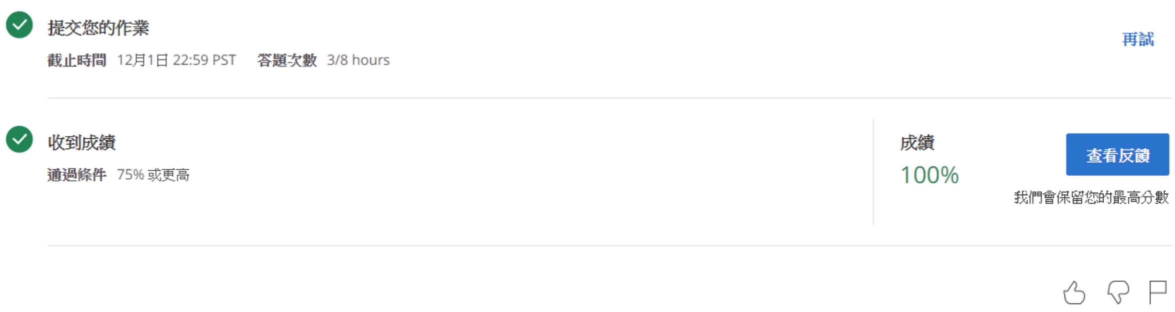
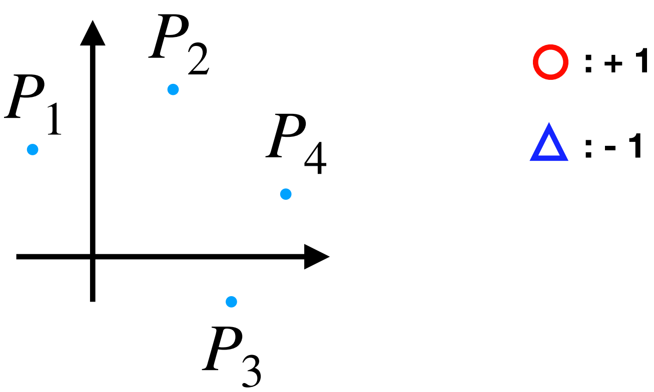
1.

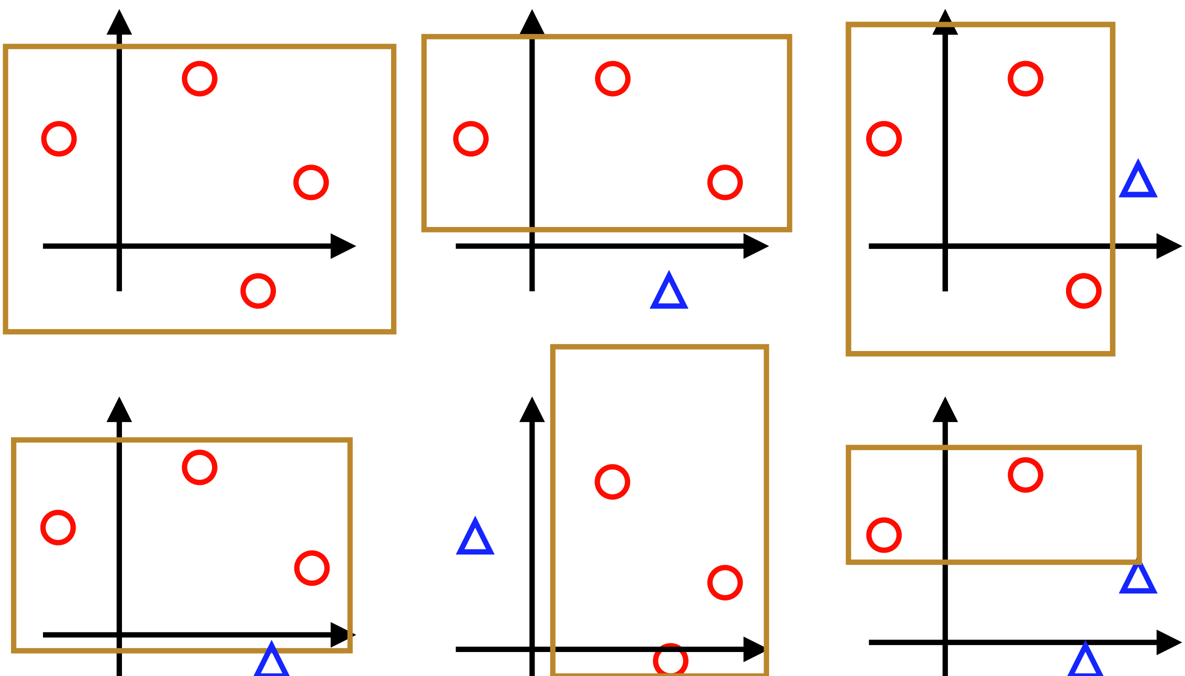


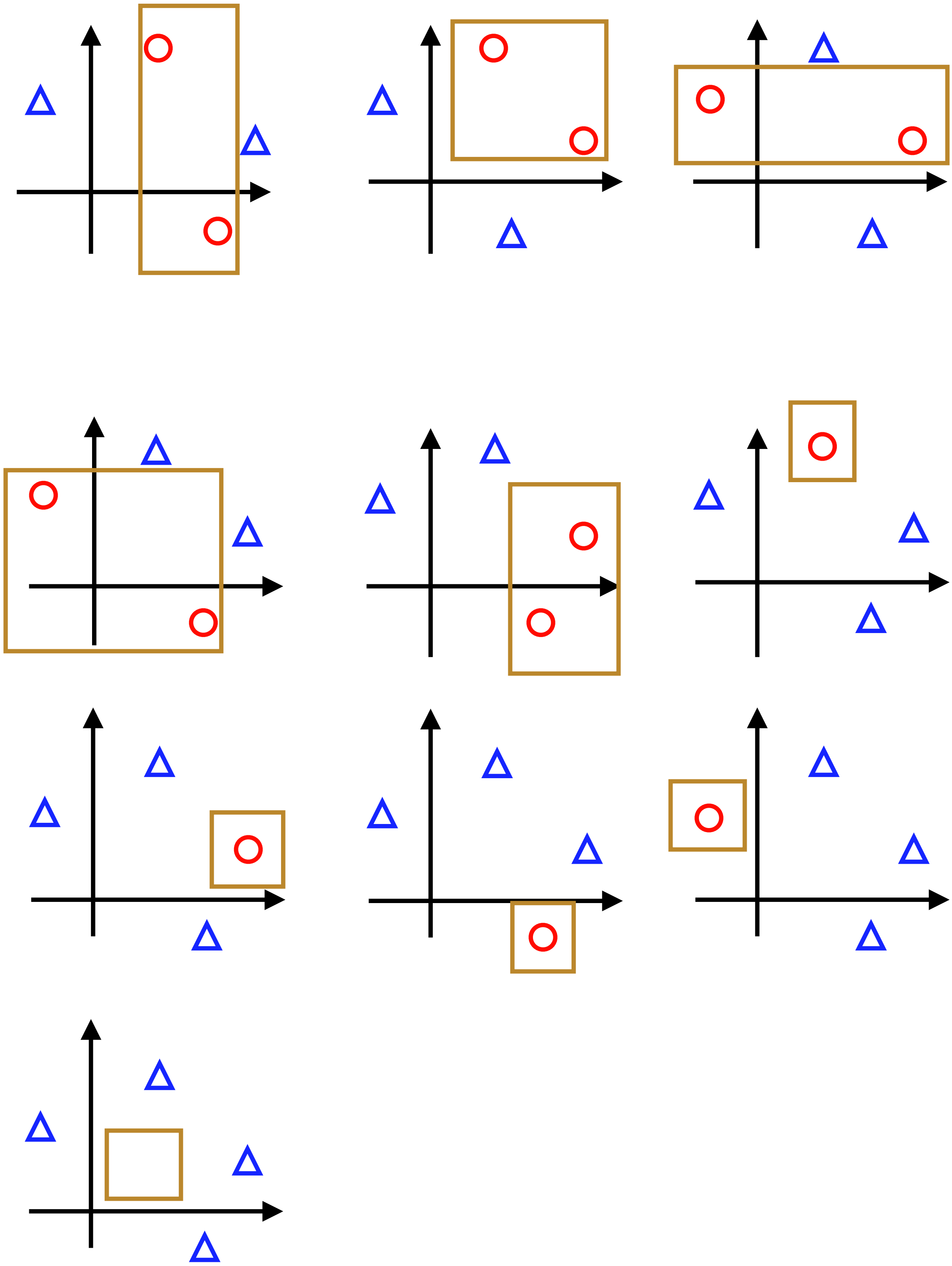
2. Show that the VC-Dimension of the hypothesis set is no less than 4.

Proof : show that we could shatter $point(i) = P_i = (x_i, y_i), i = 1,...,4$, where $x_1 < x_2 < x_3 < x_4$, and $y_1 \neq y_2 \neq y_3 \neq y_4$

Ex :







Since we could find 2^4 = 16 different dichotomies
=> We could shatter 4 inputs
=> break point ≥ 5
=> dvc(positive rectangle) ≥ 4

QED

3.

考慮$h_\alpha(x) = -1$的情況 $\iff |\alpha x \; mod \; 4 - 2| - 1 < 0$

$\iff 1 < \alpha x \mod 4 < 3$

令$x = (x_1, x_2, \ldots, x_k)$, 其中$x_i = 4^i$

$\alpha = (\frac{1}{4})^1 a_1 + (\frac{1}{4})^2 a_2 + \ldots + (\frac{1}{4})^i a_i + \ldots + (\frac{1}{4})^k a_k$

$$\rightarrow \alpha x_i = [(\frac{1}{4})^1 a_1 + (\frac{1}{4})^2 a_2 + \ldots + (\frac{1}{4})^i a_i + \ldots + (\frac{1}{4})^k a_k]4^i$$
$$= (\frac{1}{4})^{1-i} a_1 + \ldots + a_i + (\frac{1}{4})a_{i+1} \ldots + (\frac{1}{4})^{k-i} a_k$$

其中，$(\frac{1}{4})a_{i+1} \ldots + (\frac{1}{4})^{i-k} a_k \leq 2 * (\frac{\frac{1}{4}}{1-\frac{1}{4}}) = \frac{2}{3}$，when $k \rightarrow \infty$

(assume $a_{i+1}, \ldots, a_k = 2 \; here, \; since \; we \; could \; choose \; \alpha$)

$\rightarrow \alpha x_i \; mod 4 \leq a_i + \frac{2}{3}$

$\rightarrow let \; a_i = \; 0, \; \alpha x_i \; mod 4 \leq \frac{2}{3} \; \rightarrow \; |\alpha x_i \; mod 4 - 2| - 1 > 0 \; \rightarrow h_\alpha(x_i) = 1$

or $\quad 2, \; \alpha x_i \; mod 4 \leq 2 + \frac{2}{3} \; \rightarrow \; |\alpha x_i \; mod 4 - 2| - 1 < 0 \; \rightarrow h_\alpha(x_i) = -1$

∴透過此α之設計，可以shatter任意數量的inputx

$\therefore dvc(H) = \infty$

4. Show that $if (H_1 \cap H_2) \neq empty\ , then\ dvc(H_1 \cap H_2) \leq dvc(H_1)$.

Proof :

已知$(H_1 \cap H_2) \subseteq H_1$

case 1 : $dvc(H_1 \cap H_2)\ is\ finite$

$assume\ dvc(H_1) = n_1,\ dvc(H_1 \cap H_2) = n_2,\ claim\ n_1 \geq n_2$

$\therefore\ \exists h_1, \ldots h_{2^{n2}} \in (H_1 \cap H_2),\ could\ shatter\ some\ of\ n_2\ inputs$

$\because\ (H_1 \cap H_2) \subseteq H_1$

$\therefore\ h_1, \ldots h_{n_2} \in H_1$

$\therefore\ h_1, \ldots h_{n2} \in H_1,\ could\ shatter\ some\ of\ n_2\ inputs$

$\therefore\ dvc(H_1) = n_1 \geq n_2 = dvc(H_1 \cap H_2)$

case 2 : $dvc(H_1 \cap H_2)\ is\ infinite$

$\forall\ n\ input, \exists\ h_1, \ldots, h_{2^n}\ could\ shatter\ them$

$\because\ (H_1 \cap H_2) \subseteq H_1$

$\therefore\ h_1, \ldots h_{2^n} \in H_1$

$\therefore\ h_1, \ldots h_{n2} \in H_1,\ could\ shatter\ some\ of\ n_2\ inputs$

$\therefore\ H_1\ could\ shatter\ all\ number\ of\ inputs$

5. $m_{H_1 \cup H_2}(N) = 2N,\ dcv(H_1 \cup H_2) = 2$

One of solutions :

$\because\ H_1\ is\ positive\ ray\ and\ H_2\ is\ negtive\ ray$

$\therefore\ H_1 \cup H_2\ is\ one\ dimension\ perceptron$

$\therefore\ m_{H_1 \cup H_2}(N) = 2N$

$\therefore\ dvc(H_1 \cup H_2) = d + 1 = 2$

6.

Derivation :

According to question 1 form Coursera, we know the formula which could calculate the probability of error that hypothesis makes in approximating the noisy target y

$$E_{out} = \lambda\mu + (1-\lambda)(1-\mu) \quad - \quad (1)$$

,where $(1-\mu)$ is the probability that hypothesis predict the same value as target function do, and $(1-\lambda)$ is the noise rate of input.

So, the problem now is how could we use represent λ and μ by θ and s

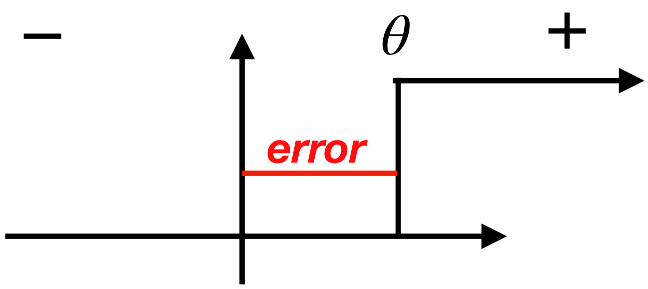
Now we try to give θ and s some mathematical meanings

θ determines the threshold of this hypothesis, and

s determines positive(+1) or negative(-1) ray it is

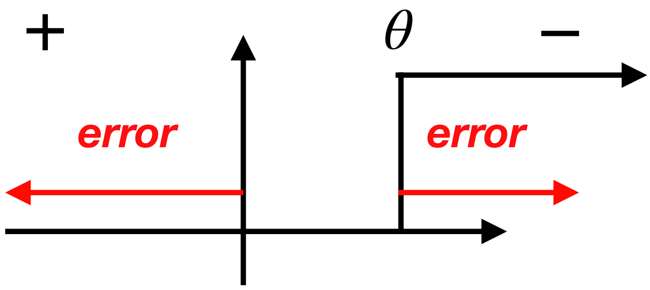
Discuss λ and μ from 4 case:

case 1 : $\theta > 0,\ s = 1$



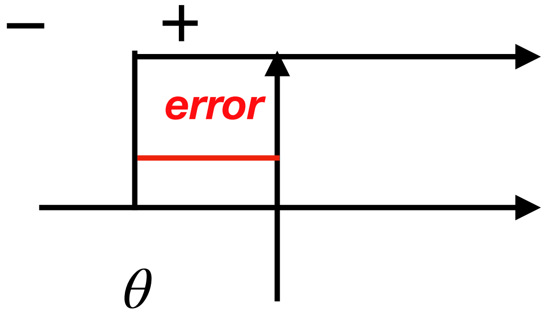
$error = \theta/2$

case 2 : $\theta > 0,\ s = -1$



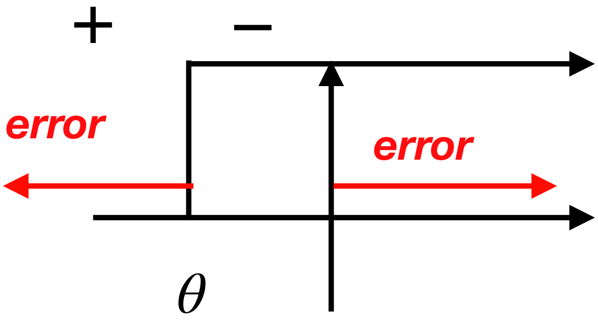
$$error = 1 - \theta/2$$

case 3 : $\theta < 0,\ s = 1$



$$error = \theta/2$$

case 4 : $\theta < 0,\ s = -1$



$$error = 1 - \theta/2$$

So, we could get the formula :

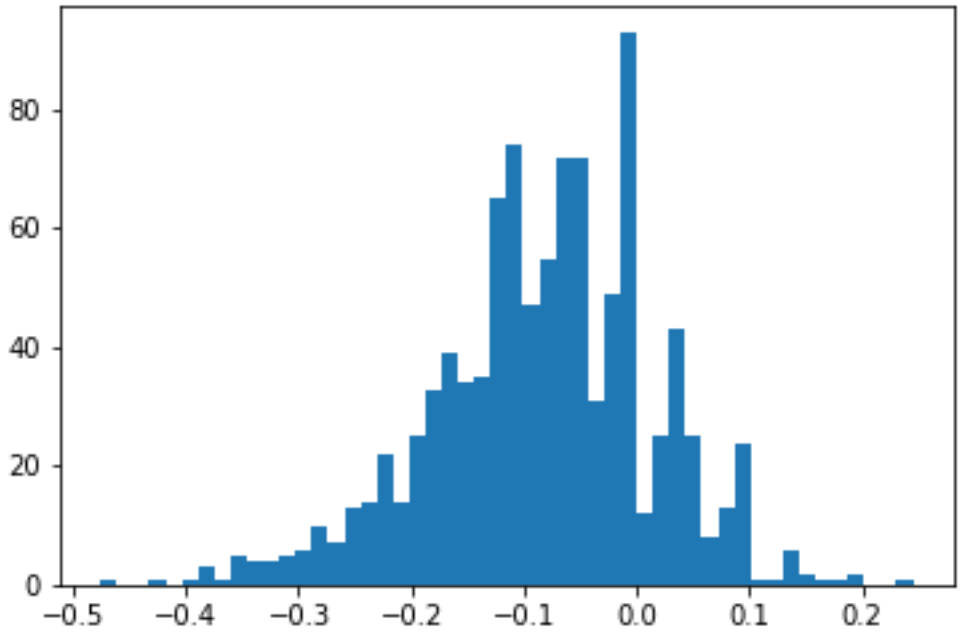
$$\mu = (\frac{1-s}{2}) + (s * \frac{\theta}{2})$$

And $(1-\lambda) = 0.2$

Then we can modify formula(1)

$$\begin{aligned} E_{out} &= (1-\mu)(1-\lambda) + \mu\lambda \\ &= (\frac{1+s}{2} - (s * \frac{\theta}{2})) * 0.2 + (\frac{1-s}{2} + (s * \frac{\theta}{2})) * 0.8 \\ &= \frac{1}{2} + \frac{-0.6 * s}{2} + \frac{0.6 * s * \theta}{2} \\ &= \frac{1}{2} + \frac{0.6 * s}{2}(-1+\theta) = 0.5 + 0.3s(\theta - 1) \end{aligned}$$

7.



(1.) $E_{in} - E_{out} < 0$

Explain： 因為我們取hypothesis時是依照E_{in}而不是E_{out}挑選，因此挑出的hypothesis，平均多次之後有很大機率 $E_{in} < E_{out}$

(2.)

由觀察得知，此時大部分的$E_{in} - E_{out}$都落在-0.2到0.04之間(平均值約落在-0.083)，根據上課提到的公式

$$P_D[\,|E_{in}(g) - E_{out}(g)| > \epsilon] \leq 4(2N)^{dc}exp(-\frac{1}{8}\epsilon^2 N)$$

已知$N = 20,\ dvc = 2$

觀察 $\epsilon = 0.05, 4(2N)^{dc}exp(-\frac{1}{8}\epsilon^2 N) = 6360.1$

觀察$\epsilon = 0.12, 4(2N)^{dc}exp(-\frac{1}{8}\epsilon^2 N) = 6173.7$

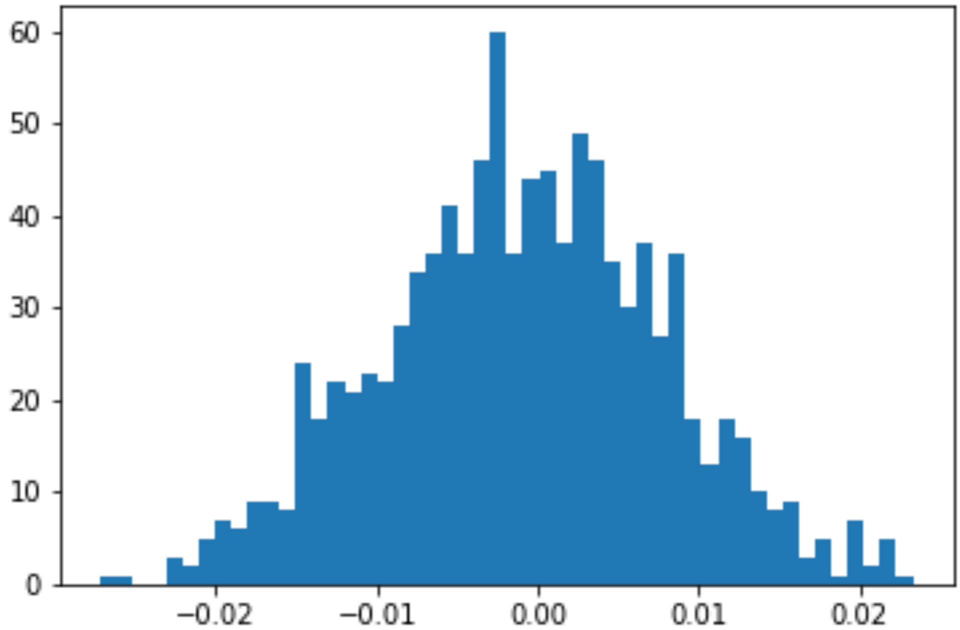
但實際上E_{out}與E_{in}已經相當靠近，發現理論數值與實際數值的差距，但若是根據老師上課所述，將N放大1000倍(相當於需要的input size量減少1000倍)，再利用同樣的公式去計算

$\epsilon = 0.05, 4(2N)^{dc}exp(-\frac{1}{8}\epsilon^2 N) = 1.2355e+07$

$\epsilon = 0.12, 4(2N)^{dc}exp(-\frac{1}{8}\epsilon^2 N) = 1.4845e-06$

即當$\epsilon = 0.05$, 壞事情($|E_{in}(g) - E_{out}(g)| > \epsilon$)發生的機率已經非常大，當$\epsilon = 0.12$, 壞事情($|E_{in}(g) - E_{out}(g)| > \epsilon$)發生的機率已經非常小，雖然和真實數據仍有一定的差距，但相對原本的公式已經合理許多。

8.



(1.) $E_{in} - E_{out} < 0$

Explain： 因為我們取hypothesis時是依照E_{in}而不是E_{out}挑選，因此挑出的hypothesis，平均多次之後有很大機率 $E_{in} < E_{out}$

(2.) 可以發現當 N 變大的時候，在ϵ相同的情況下，壞事情($|E_{in}(g) - E_{out}(g)| > \epsilon$)發生的機率變小，此時大部分的$E_{in} - E_{out}$都落在-0.02到0.018之間(平均值約落在-0.0008)

已知$N = 2000,\ dvc = 2$ ，

觀察$\epsilon = 0.005, 4(2N)^{dc}exp(-\frac{1}{8}\epsilon^2 N) = 6.3601e + 07$

觀察$\epsilon = 0.012, 4(2N)^{dc}exp(-\frac{1}{8}\epsilon^2 N) = 6.1737e + 07$

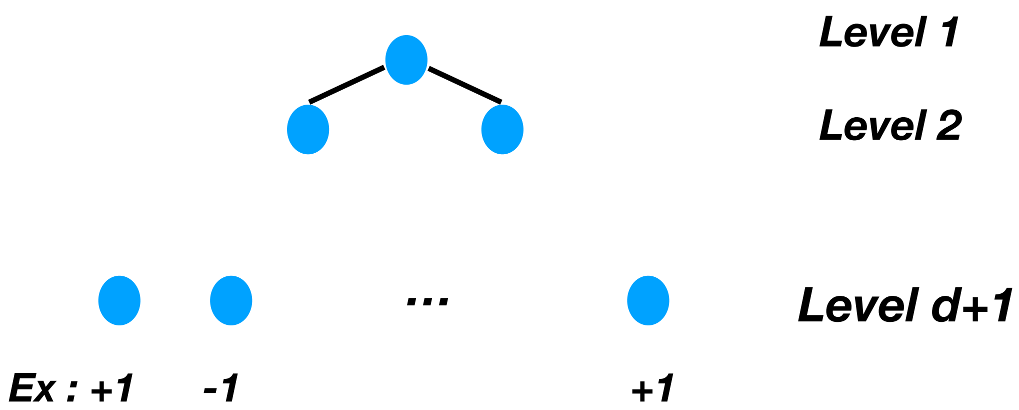
當N放大1000倍之後

$\epsilon = 0.005, 4(2N)^{dc}exp(-\frac{1}{8}\epsilon^2 N) = 1.2355e + 11$

$\epsilon = 0.012, 4(2N)^{dc}exp(-\frac{1}{8}\epsilon^2 N) = 0.0148$

結果合理許多，與Problem 7的結果相比較後，證明N的數量的確會影響壞事情發生的機率。

9. Decision tree looks like this



Since the dimension of S is d, the height of tree is d + 1, and this tree have 2^d leaves. Each leave could be or +1 or -1, so at most we could have 2^{2^d}dichotomies which is not influenced by input number.

對於decision tree的每一個level，我們可以看成在dimension i = t_i上切一刀，令其分成兩個subregion，$x_i \leq t_i$代表decision tree的左子樹，$x_i > t_i$，代表decision tree右子樹

因為實數為countable infinity set，所以對於任意$t_i \in R$,
$\exists t_i^l = t_i - 1 \in R \ and \ t_i^r = t_i + 1 \ \in R$

因此對於任意一個leave我們都可以根據decision tree的rule找到一個point ∈ 該leave代表的subregion

∴ for each subregion s，我們至少可以找到一個點∈s
∴ 我們可以找到2^{2^d}種dichotomies

Then we try to find break point of H, which is the first number of input we could not shatter.
$2^{2^d} < 2^N$
=> $2^d < N$
=> $N = 2^d + 1$

And we could calculate that $dvc(H) = break \ point(H) - 1 = 2^d$