HW2

1.

題目給的公式是$\mathrm{P}(\mu > \upsilon + \epsilon) \le \mathrm{e}^{-2\epsilon^2 N}$。在題目中，μ就是真實的$\mu_{\mathrm{m}}$，υ就是取樣的機率，就是$\frac{\mathrm{c_m}}{\mathrm{N_m}}$。$N$就是取樣的數量，就是$\mathrm{N_m}$。

令ϵ是$\sqrt{\frac{\ln t - \frac{1}{2}\ln \delta}{\mathrm{N_m}}}$，其中δ是給定的，則帶入公式可得$\mathrm{P}\left(\mu_{\mathrm{m}} > \frac{\mathrm{c_m}}{\mathrm{N_m}} + \sqrt{\frac{\ln t - \frac{1}{2}\ln \delta}{\mathrm{N_m}}}\right) \le \mathrm{e}^{-2\frac{\ln t - \frac{1}{2}\ln \delta}{\mathrm{N_m}}\mathrm{N_m}} = \mathrm{e}^{-2\left(\ln t - \frac{1}{2}\ln \delta\right)} = \mathrm{e}^{-2\ln t + \ln \delta} = \delta \mathrm{t}^{-2}$。即是題目要求的結果。

2.

m 有 M 種，所以$\mathrm{P}\left(\mu_{\mathrm{m}} > \frac{\mathrm{c_m}}{\mathrm{N_m}} + \sqrt{\frac{\ln t + \ln M - \frac{1}{2}\ln \delta}{\mathrm{N_m}}}\right)$有 M 種。T 有無限個，根據 p1 的結果，對於一個 m 和一個 t，$\mathrm{P}\left(\mu_{\mathrm{m}} > \frac{\mathrm{c_m}}{\mathrm{N_m}} + \sqrt{\frac{\ln t + \ln M - \frac{1}{2}\ln \delta}{\mathrm{N_m}}}\right) \le \delta \mathrm{t}^{-2}$，所以$\mathrm{P}\left(\bigcup_{m=1}^{M}(\mu_{\mathrm{m}} > \frac{\mathrm{c_m}}{\mathrm{N_m}} + \sqrt{\frac{\ln t + \ln M - \frac{1}{2}\ln \delta}{\mathrm{N_m}}})\right) \le \mathrm{MP}\left(\mu_{\mathrm{m}} > \frac{\mathrm{c_m}}{\mathrm{N_m}} + \sqrt{\frac{\ln t + \ln M - \frac{1}{2}\ln \delta}{\mathrm{N_m}}}\right) \le$

$\mathrm{M}e^{-2\frac{\ln t+\ln M-\frac{1}{2}\ln\delta}{\mathrm{N_m}}\mathrm{N_m}}=\mathrm{M}e^{-2\left(\ln t+\ln M-\frac{1}{2}\ln\delta\right)}=\mathrm{M}e^{-2\ln t-2+\ln M+\ln\delta}=\delta t^{-2}M^{-1}$。這代表對所有 m，至少有一個符合$\mu_{\mathrm{m}}>\frac{\mathrm{c_m}}{\mathrm{N_m}}+\sqrt{\frac{\ln t+\ln M-\frac{1}{2}\ln\delta}{\mathrm{N_m}}}$的機率小於$\delta t^{-2}M^{-1}$。接下來，對所有 t>M，至少有一個符合$\mu_{\mathrm{m}}>\frac{\mathrm{c_m}}{\mathrm{N_m}}+\sqrt{\frac{\ln t+\ln M-\frac{1}{2}\ln\delta}{\mathrm{N_m}}}$的機率小於$\sum_{t=1}^{\infty}\delta t^{-2}M^{-1}=\frac{\delta\pi^2}{6\mathrm{M}}<\delta$。換句話說，對所有 t>M，對所有 m，都符合$\mu_{\mathrm{m}}\leq\frac{\mathrm{c_m}}{\mathrm{N_m}}+\sqrt{\frac{\ln t+\ln M-\frac{1}{2}\ln\delta}{\mathrm{N_m}}}$的機率至少是 1-δ。

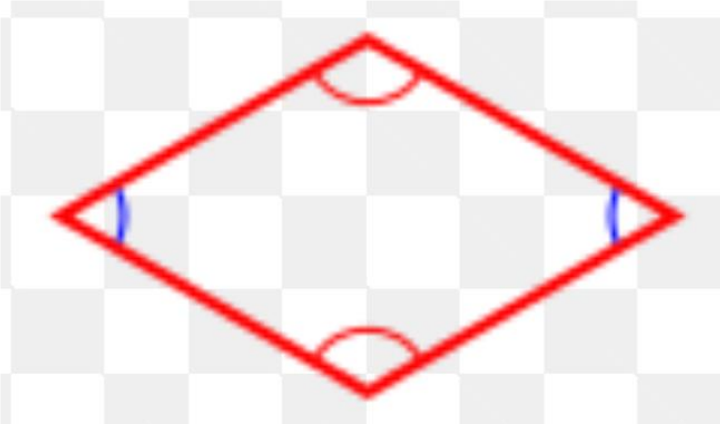
3.

總共有 4^5=1024 種可能性。然後考慮以下限制:A 跟 B 不能同時出現、C 跟 D 不能同時出現。所以 5 個彩票可以歸類為 4 種:AC、AD、BC、BD。每一個類別都是隨便組都符合條件。所以答案是(32+32+32+32+32)/1024=5/32。

4.

只有 BD 這個種類可以讓所有 2 都是綠色。所以是32/1024=1/32。

5

考慮這種菱形。這 4 個頂點不管哪些是-1，都可以找到一個四邊跟 x,y 軸平行的長方形把-1 的點包起來。相連的頂點是-1 就非常簡單，而如果是上下-1 左右 1 的話就在中間拉個瘦長的長方形，如果是上下 1 左右-1 的話就在中間拉個扁平的長方形。這 4 個點可以被 shattered，所以 VC dimension 必須不小於 4。

6

數學歸納法。根據講義，當 M=1 時，$d_{vc}=2$。令當 M=k 時，$d_{vc}=2k$。當 M=k+1 時，一般性假設把多的點加在最右邊某處好了，畢竟就算加在中間，也可以把最右邊的點去掉，變回 k 個點，再加上最右邊的點變回來。如果最右邊的點的 label 是 s，那就把a_{k+1}跟b_{k+1}加在最右邊的點的左右 2 側。根據實數的稠密性，必定存在 2 個點把最右邊的點包起來而不影響左邊其他的點跟b_k。如果最右邊的點的 label 是-s，那就把a_{k+1}跟b_{k+1}加在最右邊的點的右邊。一樣可以達成目標。這裡證明了當 M=k+1 時，$d_{vc}\geq 2(k+1)$。

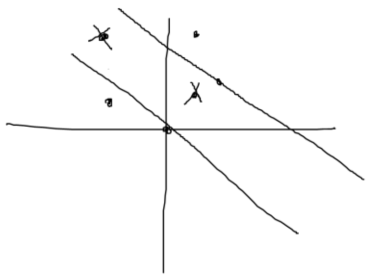
同時，如果存在 k+2 個點可以在 M=k+1 時可以被 shatter，一樣看最右邊的點，如果最右邊的點的 label 是 s，代表最右邊的點被a_{k+1}跟b_{k+1}所包住。這時同時去除a_{k+1}跟b_{k+1}跟最右邊的點，這樣不會影響到左邊其他的點跟b_k，會導致 k+1 個點可以在 M=k 時可以被 shatter，矛盾。如果最右邊的點的 label 是-s，代表最右邊的點在a_{k+1}跟b_{k+1}的範圍之外。這時同時去除a_{k+1}跟b_{k+1}跟最右邊的點，這樣不會影響到左邊其他的點跟b_k，會導致 k+1 個點可以在 M=k 時可以被 shatter，矛盾。因此當 M=k+1 時，$d_{vc} = 2k + 2$。根據數學歸納法，我們可以說，當 M=M 時$d_{vc} = 2M$。

7.

因為 perceptron 通過原點，所以我們可以把每個 data 做 normalize，縮成距離原點為 1。這樣所有 data 就會在一個 unit circle 上面。如果說有 M 個點好了，則考慮把 perceptron 的線旋轉半圈。在旋轉半圈的過程中，最多會在 M 個不同角度碰到不同的點，。如果同時旋轉過程中同時碰到 2 個點，則這些 data 的 dichotomy 最多會變化 M 次。這時所有 data 的 label 都跟初始狀況相反。然後再旋轉回來，這些 data 的 dichotomy 又會最多會變化 M 次。再度變回來。因為旋轉了一圈，所以所有可能的 dichotomy 都出現過。而 dichotomy 變化了 2M 次，所以 growth function 是 2M。

8.

　　這題我不太確定題目的意思。我這裡就設定說對於每個點，都可以各自挑一個 hypothesis 當作判斷基準。所以如 4 個點，設成



就可以處理對角線同號其他異號的 case。5 個點的 case，不管每個點的 label 是 1 還是-1，都可以用 2 條線把這 5 個點的 5 邊形根據 label 做分割，達成一個區域就 1 種 label。調一下線的角度，就可以做到讓線通過(0,0)跟(1,1)。所以 5 個點也可以被 shatter。但是當 6 個點時，把這個 6 邊形通過每個點的凸包畫出來，是個凸 6 邊形，如果這 6 邊形每相鄰的點都是不同 label，總共 3 個 1，3 個-1，這種情況就無法用 2 條線把 6 個點做分割。因為 2 條線最多 4 個區塊，所以一定有相鄰的 2 個點在同個區塊中。但相鄰的點是不同 label，所以無法用題目給的 H 去生成這種 dichotomy。因此 6 個點無法被 H 所 shatter。因此$\mathrm{d}_{\mathrm{vc}} = 5$。

9.

考慮說當$|\theta| = 1$時，$h_{s,\theta}$會全部猜-1 或 1。但 label 中 1 跟-1 各占一半，所以$\mathrm{E_{out}}(h_{s,\theta}) = 0.5$。考慮$\theta = 0$，則當 s=1 時，正確率有 0.9。當 s=-1 時，錯誤率有 0.9。所以這時$\mathrm{E_{out}}(h_{s,\theta}) = 0.5 - 0.4\mathrm{s}$。考慮說當$0 < \theta < 1$且 s=1 時，對於$\mathrm{x} \leq 0$的點，其 label 有 90%正確，10%錯誤。對於$0 < \mathrm{x} \leq \theta$，其 label 有 90%錯誤，10%正確。對於$\theta < \mathrm{x} \leq 1$，其 label 有 90%正確，10%錯誤。這時$\mathrm{E_{out}}(h_{s,\theta}) = 0.5 * 0.1 + \theta * 0.9 * 0.5 + (1-\theta) * 0.1 * 0.5 = 0.1 + 0.4\theta$。考慮說當$0 < \theta < 1$且 s=-1 時，s=1 的那些正確錯誤全部顛倒，這時$\mathrm{E_{out}}(h_{s,\theta}) = 0.5 * 0.9 + \theta * 0.1 * 0.5 + (1-\theta) * 0.9 * 0.5 = 0.9 - 0.4\theta$。

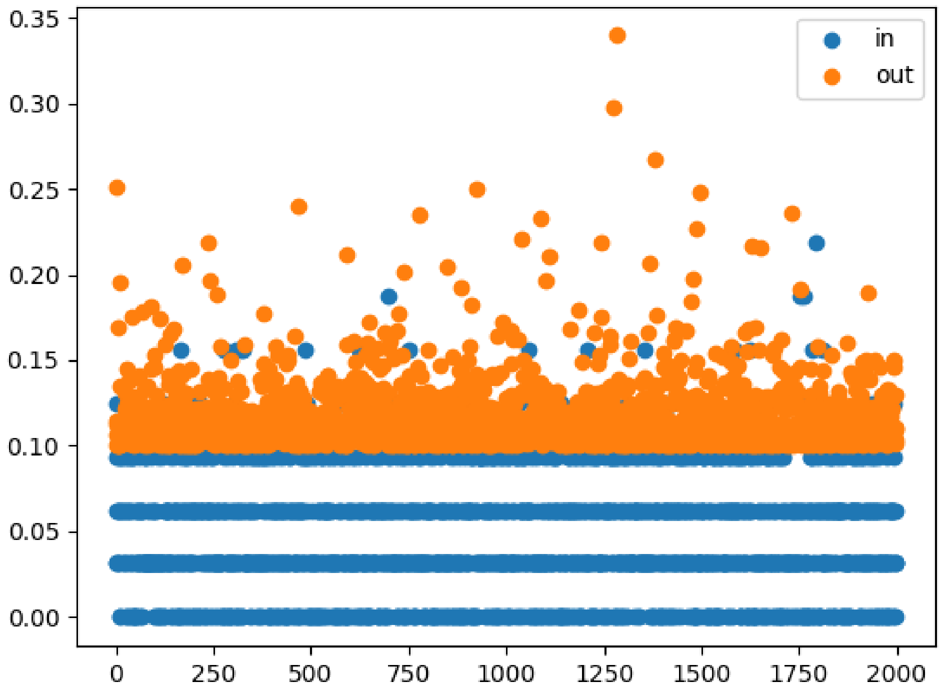
考慮說當$-1 < \theta < 0$且 s=1 時，對於$\mathrm{x} > 0$的點，其 label 有 90%正確，10%錯誤。對於$\theta \leq \mathrm{x} \leq 0$，其 label 有 90%錯誤，10%正確。對於$-1 \leq \mathrm{x} < \theta$，其 label 有 90%正確，10%錯誤。這時$\mathrm{E_{out}}(h_{s,\theta}) = 0.5 * 0.1 + \theta * 0.9 * 0.5 + (1-\theta) * 0.1 * 0.5 = 0.1 + 0.4\theta$。考慮說當$-1 < \theta < 0$且 s=-1 時，s=1 的那些正確錯誤全部顛倒，這時$\mathrm{E_{out}}(h_{s,\theta}) = 0.5 * 0.9 + \theta * 0.1 * 0.5 + (1-\theta) * 0.9 * 0.5 = 0.9 - 0.4\theta$。

以上考慮了所有可能的 case，而符合所有的 case 的公式是

$\mathrm{E_{out}}(h_{s,\theta}) = 0.5 - 0.4s + 0.4s|\theta|$。所以這個就是答案。

10.

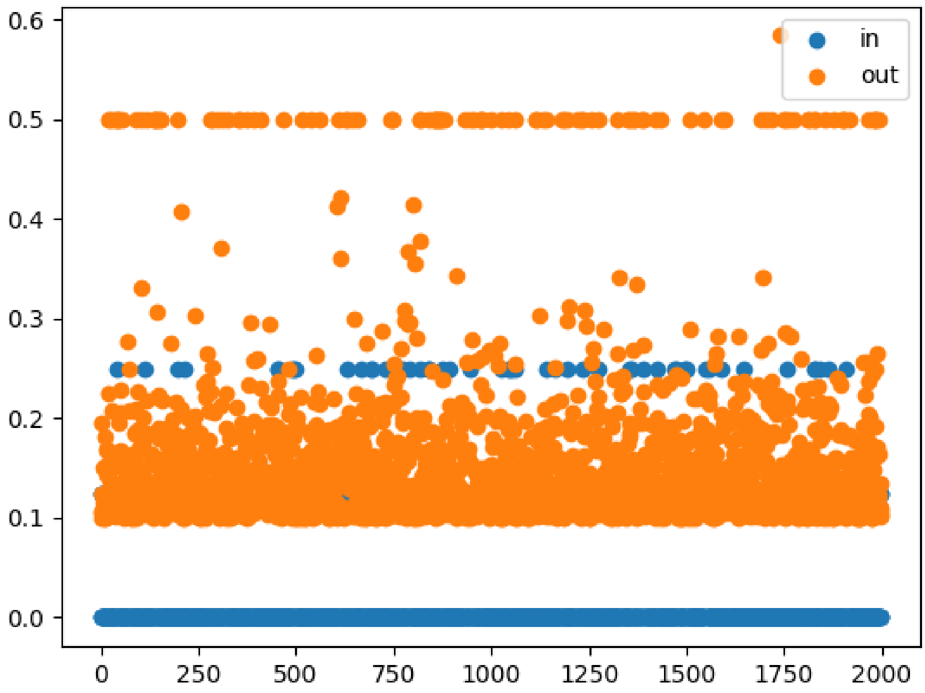
這是我的圖:



Eout-Ein 的 median 是 0.11。

11.

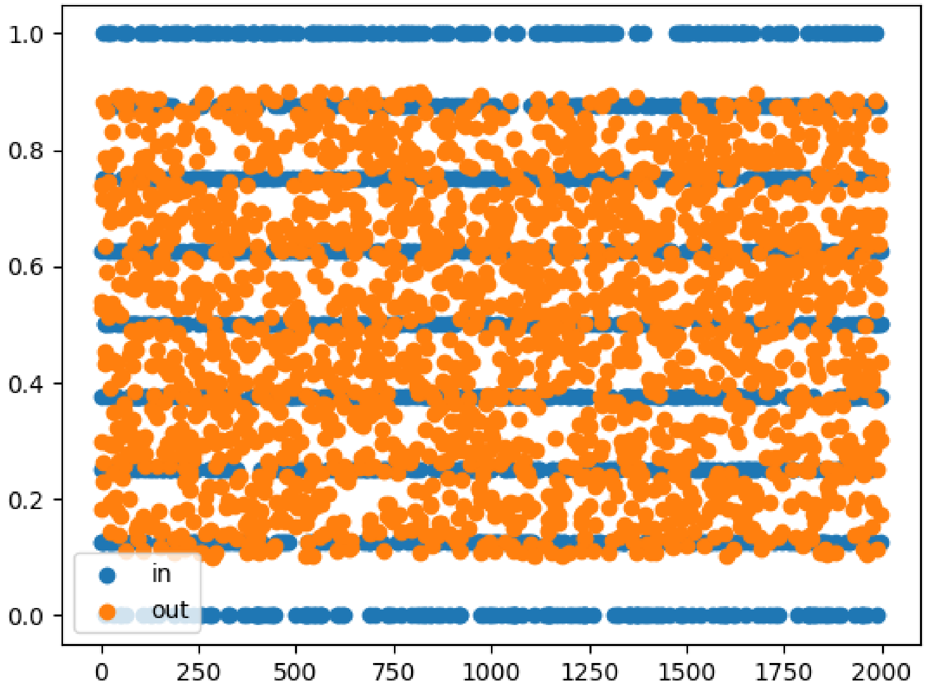
這是我的圖:



Eout-Ein 的 median 是 0.13309575480110164。可以發現 Eout 在 0.5 多了一條線。檢查後發現，因為 8 這個數字太小，所以較為容易產生一種 label 是 8 個全負的。這種情況 theta=-1 且 s=-1 就會正確，所以就會產生很多 theta=-1 但表現很好的狀況。用上課的話來說就是 data size 不夠大，所以產生 bad data。

12.

這是我畫的圖



Eout-Ein 的 median 是 0.0024354740491804305。可以看出來 theta 隨便取，s 隨便取，所以 Eout 跟 Ein 都非常隨機。而 median 會低是因為平均 Eout 跟 Ein 的平均都是在很接近 0 的位置。

13.

根據題目中給出的講義，因為必須要通過 k 個點，維度會降到 N-k，所以可以令通過 k 的點的 dichotomy 的數量是 C(P,N-k)。同樣如果多加一個點則可能會生出更多 dichotomy，所以 C(P+1,N-k)= C(P,N-k)+D，D 是增加的 dichotomy 的數量。因為要通過 k 個點，所以這時考慮的維度降到 N-k-1。這時點的數量還是 P 個，所以 D= C(P,N-k-

1)。因此遞迴式就是 C(P+1,N-k)= C(P,N-k)+ C(P,N-k-1)。

一樣數學歸納法。令$C(P,N-k)=2\sum_{i=0}^{N-k-1}\binom{P-1}{i}$，則

$C(P+1,N-k)=2\sum_{i=0}^{N-k-1}\binom{P-1}{i}+2\sum_{i=0}^{N-k-2}\binom{P-1}{i}=$

$2\sum_{i=0}^{N-k-1}\binom{P-1}{i}+2\sum_{i=0}^{N-k-1}\binom{P-1}{i-1}=2\sum_{i=0}^{N-k-1}\binom{P}{i}$，符合上面原式。把

符號換成本題，答案就變成是$2\sum_{i=0}^{d-k-1}\binom{N-1}{i}$。