Machine Learning Foundations/hw2

連少甫/r07922107

December 4, 2018

1.

2.

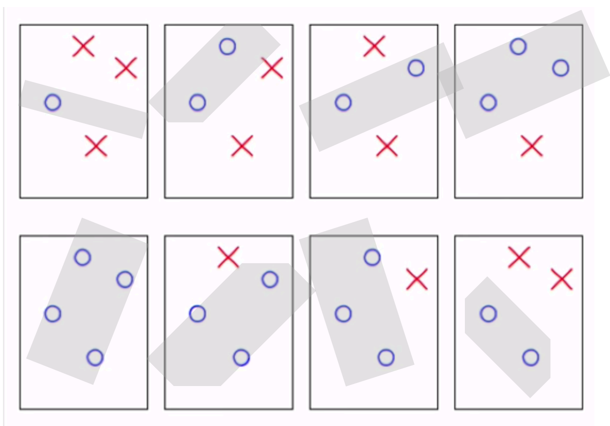
當θ等於0時，點在法向量為w的平面其Y值為+1其餘為-1。

假設一平面W有法向量w,$w = (w_0, w_1, w_2)$

$|w_0 + w_1x_1 + w_2x_2| \le \theta \rightarrow |cos\theta'|x||w|| \le \theta,$

當 x_a滿足$|x_a| \le \frac{\theta cos\theta'}{|w|}$,則$|w_0x_{a1} + w_1x_{a2} + w_2x_{a3}| \le \theta$成立

根據上述討論$|w_0 + w_1x_1 + w_2x_2| \le \theta$考量了W、X向量的長度及夾角，當$\theta$等於某常數時，我可以看作將該平面“增厚”，且落於該區域的點其值為+1其餘為-1。為了證明$d_{vc}(H) \ge 4$，只需要證明給定某四個點，都存在一個H能夠shatter，如下圖所示。

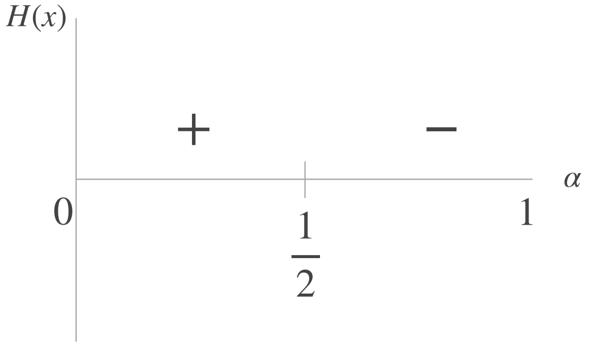


3.

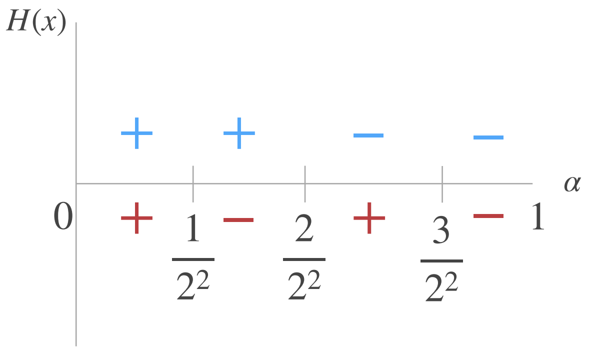
假設有N個點存在於線上時，我們令第i個點的位置為$2^i, i = 1,2,.....N$。

當只考慮第一點$x_1 = 2$時α介於$(0,\frac{1}{2})$時,$H(x_1) = +1$,反之當a介於$(\frac{1}{2},1)$時$H(x_1) = -1$

如下圖所示



此時加入第二點$x_2 = 4$，當x_2介於$(0,\frac{1}{2^2})$ $(\frac{2}{2^2},\frac{3}{2^2})$時$H(x_2) = +1$，其餘$H(x_2) = -1$，如下圖所示，紅色的正負符號代表$H(x_2)$的正負值，藍色代表$H(x_1)$的正負值。



因此可以透過控制α的值改變每一個點的正負，將其推廣到N個點即可證明任一點都可以被Shatter，其2^N種dichotomies組合對應的α值如下:

$\frac{k}{2^N} < \alpha < \frac{k+1}{2^N},\ \ k = 0,1,2,3,4,...,2^N - 1$

因此$d_{vc} = \infty$。

4.

$H_1 \cap H_2$是H_2的子集合，因此$dichotomies\ H_1 \cap H_2(x_1, x_2, ..., x_N)$同樣是$dichotomies\ H_2(x_1, x_2, ..., x_N)$的子集合， 所以$m_{H_1 \cap H_2}(N) \leq m_{H_2}(N)$，

因此$d_{vc}(H_1 \cap H_2) \leq d_{vc}(H_2)$，

5.

$m_{H_1 \cap H_2}(N) = 2(N-1) + 2 = 2N$

$N = 3,\ m_{H_1 \cap H_2}(3) = 2(N-1) + 2 = 2N = 6 \neq 2^3$

$d_{vc}(N) = 2$

6.

當$s = 1, \theta > 0$時，區間$[0,\theta]$的錯誤率為0.8其餘為0.2。

$E_{out} = (0.2(2 - |\theta|) + 0.8|\theta|)/2 = 0.2 + 0.3|\theta|$

當$s = 1, \theta < 0$時，區間$[\theta,0]$的錯誤率為0.8其餘為0.2。

$E_{out} = (0.2(2 - |\theta|) + 0.8|\theta|)/2 = 0.2 + 0.3|\theta|$

當$s = -1, \theta > 0$時，區間$[0,\theta]$的錯誤率為0.2其餘為0.8。

$E_{out} = (0.2(|\theta|) + 0.8(2 - |\theta|))/2 = 0.8 - 0.6|\theta|$

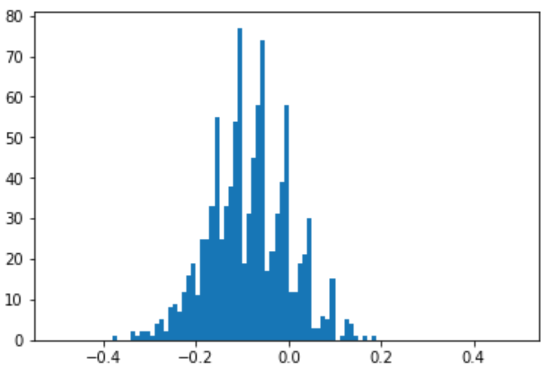
當$s = -1, \theta < 0$時，區間$[\theta,0]$的錯誤率為0.2其餘為0.8。

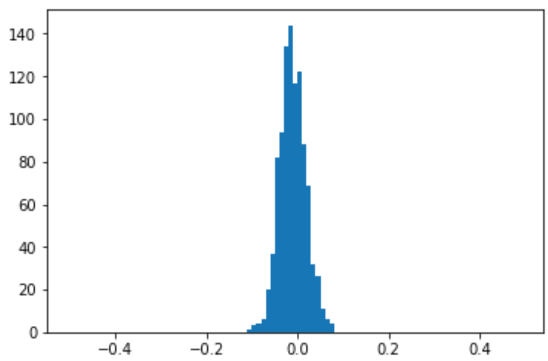
$E_{out} = (0.2(|\theta|) + 0.8(2 - |\theta|))/2 = 0.8 - 0.6|\theta|$

可以歸納出$E_{out} = 0.5 + 0.3s(|\theta| - 1)$

7.

$E_{in} - E_{out}$的histogram(1000 times)如下左圖所示，平均$E_{in} = 0.171$、$E_{out} = 0.254$。照理來說E_{in}和E_{out}應該要很接近，因此可能是因為樣本數過少導致相差有一點的距離，因此我們將data數量增加至200，其結果如下右圖所示，可以發現當樣本數增加時，E_{in}確實可以代表E_{out}的狀況。





Bonus:

證：$B(N,K) \le \sum_{i=0}^{k-1} \binom{N}{i}$

當$N = K$時, $B(N,K) = 2^N - 1 = \sum_{i=0}^{k} \binom{N}{i} - 1 = \sum_{i=0}^{k-1} \binom{N}{i} + \binom{N}{k} - 1 = \sum_{i=0}^{k-1} \binom{N}{i}$

當$N < K$時, $B(N,K) = 2^N = \sum_{i=0}^{N} \binom{N}{i} \le \sum_{i=0}^{K-1} \binom{N}{i}$

當$N > K$時, 已知 $B(N+1,K) \le B(N,K) + B(N,K-1)$

假設$B(N,K) \le \sum_{i=0}^{k-1} \binom{N}{i}$成立，則$B(N+1,K) \le B(N,K) + B(N,K-1) \le \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=0}^{k-2} \binom{N}{i}$

因定理$\binom{N}{r} = \binom{N-1}{r} + \binom{N-1}{r-1}$可知$\sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=0}^{k-2} \binom{N}{i} = \sum_{i=0}^{k-1} \binom{N+1}{i}$