Homework 4
CSCI4100

Han Hai
Rin:661534083
haih2@rpi.edu

September 30 2018

1. **[300 POINTS] Exercise 2.4** Consider the input space X=1 XR^d including x0=1. Show that the VC dimension of perceptron (with d+1 parameters counting w_0) is exactly d+1 by showing that it is at least d+1 and at most d+1, as follows.

   (a) To show that $d_{vc} \geq d+1$, find d+1 points in X that the perceptron can shatter.
   $$\begin{bmatrix} 1 & 0 & 0 & 0 & ... & 0 \\ 1 & 1 & 0 & 0 & ... & 0 \\ 1 & 0 & 1 & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 1 & 0 & 0 & 0 & ... & 1 \end{bmatrix}$$
   The above matrix is nonsingular, because its determinant is 1
   We can shatter this dataset, because for any dicatomy on the dataset, we can use perceptron $y = wx$ and $w = x^{-1}y$ to which for sure will have $y = sign(w^T x)$, so it will correctly classify the data, thus we show that $d_{vc} \geq d+1$

   (b) to show that $d_{vc} \leq d+1$ show that no set of d+2 points in X can be shattered by the perceptron.

   For any d+2 points $x_1, x_2 ... x_{d+2}$, there are more points than dimensions
   Therefore we must have $x_j = \sum_{i \neq j} a_i x_i$ (linearly independent)
   , where a_i are nonzeros. The following dicatonmy, x_i with nonzero ai, $y_i = sign(a_i)$ and x_j gets $y_j = -1$ cannot be implemented. Because if we multiply the $x_j w = \sum_{i \neq j} a_i w x_i$ by w as weight of perceptron $y_i = sign(w^T x_i) = sign(ai)$ then $aiw^T x_i > 0$. therefore it's impossible to get $x_j w$= y=-1. Therefore we cannot shatter any set of d+2 and $d_{vc} \leq d+1$. Conclusion:d_{vc} for perceptron is exactly $d+1$

2. **[300 points] Problem 2.3** Compute the maximum number of dichotomies, $mH(N)$, for these learning models, and consequently compute d_{vc} the VC dimension.

   (a) Positive or negative ray: H contains the function which are +1 on [(a,∞)(for some a) toegether with those that are +1 on (-∞,a]
   for a chosen a, there are 2(N+1) situations for +1,-1 on either side minus all +1 and all -1 two situations, 2(N+1)-2=2N.
   Therefore $mH(N) = 2N$ and largest value of N where dicatonmy equals 2^N is 2, therefore, $d_{vc}(H) = 2$

   (b) Positive or negative interval
   Comparing to positive interval, when n=3, 1 new dicatonmy is added (+1,-1,+1), when n=4, 3 are added (+1,-1,-1,+1), (+1,-1,+1,+1) and (+1,+1,-1,-1), for n=5 , 6 are added so it's the original

dicatomy plus$(n-1)C2 = \frac{1}{2}N^2 - \frac{3}{2}N + 1$
$\frac{1}{2}N + \frac{1}{2}N^2 + 1 + \frac{1}{2}N^2 - \frac{3}{2}N + 1 =$
mH(N)=$N^2 - N + 2$
$mH(3) = 8 = 2^N$ $mH(4) = 14 < 2^N$. Therefore, $d_{Vc} = 3$

(c) Two concentric spheres in R^d H contains the function which are +1 for $a \leq \sqrt{x_1^2 + ...x_d^2} \leq b$
We can map points to lower dimension $(x_1, x_2, x_3, x_4...x_d) \rightarrow \sqrt{x_1^2 + ...x_d^2}$. Then it becomes a positive interval problem, which we calculated previously, is $\frac{1}{2}N^2 + \frac{1}{2}N + 1$ and the $d_{vc} = 2$

3. **[200 points] Problem 2.8** Which of the following are possible growth functions $m_H N$ for some hypothesis set?

   As we've proved in Theorem 2.4. We either have a polynomial bound or 2^N for the bound, so $1 + N; 1 + n + \frac{N(N-1)}{2}; 2^N$; are all possible growth functions, while $2^{ceil(\sqrt{N})}$ and $2^{ceil(N/2)}$ are not possible because they are exponential
   $1 + N + \frac{N(N-1)(N-2)}{6}$ is also not possible even thought it's polynomial, because according to theorem, if $m(k) \leq 2^k$, then for all n, $m(N) \leq N^{k-1} + 1$ for this one if k=2 and it is deduced to linear bound which contradict with the fact that this function is cub.

4. **[100 points] Problem 2.10** Show that $m_H(2N) \leq m_H(N)^2$ and hence obtain a generalization bound which only involves $mH(N)$
   for $m_H(2N)$ we can seperate the points into 2 parts, $m_H(N1)+$ $m_H(N2)$ the maximum of combined will be the product of two therefore $m_H(2N) \leq m_H(N_1) * m_H(N_2)$
   and since $m_H(N_1) \leq m_H(N)$ and $m_H(N_2) \leq m_H(N)$, we get $m_H(2N) \leq m_H(N)^2$
   thus we can get bound $E_{out} \leq E_{in}(g) + \sqrt{\frac{8}{N} ln \frac{4 m_H(N)^2}{\delta}}$

5. **[100 points]Problem 2.12** For an H with $d_{vc} = 10$ what sample size do you need to have a 95 precent confidence that your generalization error is at most 0.05?
   $N \geq \frac{8}{\epsilon^2} ln(\frac{4((2N)^{d_{vc}} + 1)}{\delta})$
   plus in numbers N=1000 to be initial guess
   $N \geq \frac{8}{0.05^2} ln(\frac{4((2N)^{10} + 1)}{0.05})$
   RHS is 257251, iterate the process it converges at 452957
   Therefore we need a sample size of 452957