CMPE 257 Machine Learning Spring 2019
HW#2

Name: Akshata Deo
Sjsu id: 012565761

---

1. In equation,

$E_{out}(g) <= E_{in}(g) + ((1/2N) \ln(2M/d))^{1/2}$

given, d=0.03

and e = $((1/2N) \ln(2M/d))^{1/2}$

we can find out N with the following equation-

N >= (1/(2*(pow(e,2))) * math.log((2*M)/0.03))

(a). For M=1, and e<= 0.05

N >= 839.9

(b). For M=100, and e<= 0.05

N >= 1760.98

(c). For M=10000, and e<= 0.05

N >= 2682.01

2.

(a). h(x) belongs to H.

First, $h_1(x)$ = +1 if x>=a and h(x) = -1 if x <a.

Second, $h_2(x)$ = +1 if x<=a and h(x) = -1 if x >a.

$m_H(N) = 2N$

$d_{VC} = 2$

Explanation: This can be proved by induction-

- When N=1, h(x) would be either-
  $h_1(x)$ = +1 or -1
  $h_2(x)$ = -1 or +1
  So, there are 2 dichotomies possible in this case.
  So, $m_H(1) = 2$
- When N=2, h(x) would be-
  $h_1(x)$ = {-1,-1}, {-1,+1}, {+1,+1}
  $h_2(x)$ = {+1,+1}, {+1,-1}, {-1,-1}
  So, there are 4 dichotomies possible in this case.
  So, $m_H(2) = 4$
- When N=3, h(x) would be-
  $h_1(x)$ = {-1,-1,-1}, {-1,-1,+1}, {-1,+1,+1}, {+1,+1,+1}
  $h_2(x)$ = {+1,+1,+1}, {+1,+1,-1}, {+1,-1,-1}, {-1,-1,-1}
  So, there are only 6 dichotomies possible in this case.
  So, $m_H(3) = 6$

Here we can observe that d_{vc}=2 and k=3 and $m_H(N) = 2N$

(b). h(x) belongs to H.

First, h(x) = +1 if a<= x<=b and h(x) = -1 if x <a or x>b

Second, h(x) = -1 if a<= x<=b and h(x) = +1 if x <a or x>b

$m_H(N) = N^2 – N + 2$
$d_{VC} = 3$
Explanation: This can be proved by induction-

- When N=1, h(x) would be either-
  $h_1(x)$ = +1 or -1
  $h_2(x)$ = -1 or +1
  So, there are 2 dichotomies possible in this case.
  So, $m_H(1) = 2$
- When N=2, h(x) would be-
  $h_1(x)$ = {-1,-1}, {-1,+1}, {+1,+1}, {+1,-1}
  $h_2(x)$ = {+1,+1}, {+1,-1}, {-1,-1}, {-1,+1}
  So, there are 4 dichotomies possible in this case.
  So, $m_H(2) = 4$
- When N=3, h(x) would be-
  $h_1(x)$ = {-1,-1,-1}, {-1,-1,+1}, {-1,+1,+1}, {+1,+1,+1}, {+1,+1,-1}, {+1,-1,-1}, {-1,-1,-1}
  $h_2(x)$ = {+1,+1,+1}, {+1,+1,-1}, {+1,-1,-1}, {-1,-1,-1}, {-1,-1,-1}, {-1,-1,+1}, {-1,+1,+1}
  So, there are only 8 dichotomies possible in this case.
  So, $m_H(3) = 8$
- When N=4, h(x) would be-
  $h_1(x)$ = {-1,-1,-1,-1}, {-1,-1,-1,+1}, {-1,-1,+1,+1}, {-1,+1,+1,+1}, {+1,+1,+1,+1}, {+1,+1,+1,-1}, {+1,+1,-1,-1}, {+1,-1,-1,-1}, {-1,-1,-1,-1}, {-1,+1,+1,-1}, {-1,+1,-1,-1}, {-1,-1,+1,-1}
  $h_2(x)$ = {+1,+1,+1,+1}, {+1,+1,+1,-1}, {+1,+1,-1,-1}, {+1,-1,-1,-1}, {-1,-1,-1,-1}, {-1,-1,-1,+1}, {-1,-1,+1,+1}, {-1,+1,+1,+1}, {+1,+1,+1,+1}, {+1,-1,-1,+1}, {+1,-1,+1,+1}, {+1,+1,-1,+1}
  So, there are only 14 dichotomies possible in this case.
  So, $m_H(4) = 14$

Here we can observe that d_{vc}=3 and k=4 and $m_H(N) = N^2 – N + 2$

(c). h(x) belongs to H.
$m_H(N) = (N^2 + N + 2)/2$
$d_{VC} = 2$
Explanation: This problem will reduce to Example 2.2 Question 2 [Positive Intervals] if we consider a line with region a<=x<=b as +1 and all other regions as -1 on the line.
Here we can observe that d_{vc}=2 and k=3 and $m_H(N) = (N^2 + N + 2)/2$

---

3. Given, d_{vc}=10, e=0.05
we need 95% confidence i.e. d=0.05
putting all the values in equation-

N >= ((8/pow(0.05,2))*math.log((((4*pow((2*N),10))+ 4)/0.05)))
N >= 452957

---

4. The hypothesis h that maximizes the likelihood $\prod$(n=1 to N) $P(y_n|x_n)$ is –

Now we know, if P is the probability of some event of happening, then (1-P) is the probability of some event not happening. With this concept, we can equivalently minimize a more convenient quantity.
We also put ln because minimizing a function and minimizing its log is same.

-1/N * ln($\prod$(n=1 to N) $P(y_n|x_n)$) = 1/N * $\sum$(n=1 to N) ln (1/$P(y_n|x_n)$) ----- eq 1

Clearly, LHS is a natural decreasing function. And we know that-

$P(y|x) = \Theta\ (yw^Tx)$

Putting above value in the RHS of eq 1, we will get-

$=1/N * \sum(n=1 \text{ to } N) \ln (1/(\Theta\ (y_nw^Tx_n)))$ ----- eq 2

What we are trying to do with above equation is, minimizing it. So, it can be treated as measure error. Hence, above equation can be converted into in-sample error measure for logistic regression by putting some standard values. Now, we know that-

$\Theta(s) = e^s/(1 + e^s)$

Substituting above value in eq 2 we will get-

$$\begin{aligned} Ein(w) &= 1/N * \sum(n=1 \text{ to } N) \ln (1/(e^{ynwTxn} /(1 + e^{ynwTxn}))) \\ &= 1/N * \sum(n=1 \text{ to } N) \ln ((1 + e^{ynwTxn})/(e^{ynwTxn})) \\ &= 1/N * \sum(n=1 \text{ to } N) \ln (1 + e^{-ynwTxn}) \end{aligned}$$

Hence Proved...

5.

$\Delta Ein(w) = \sum_{n=1toN} \ln(1-e^{-ynwTxn})$

d/dw_i of $(E_{in}(w)) = d/dw_i$ of $\sum_{n=1toN} \ln(1-e^{-ynwTxn})$

on diffrentiating above equation

$\Delta Ein(w) = -1/N \sum_{n=1toN} y_nx_n\ \Theta(-y_nw^Tx_n)$

6.

Given

$\Phi_2(x) = (1,x1,x2,(x1)^2,x1x2, (x1)^2)$

a. $(x1-3)^2 + x2 = 1$
   $= (x1)^2-6x1 + 8 + x2$
   $= 8 – 6x1 + x2 + (x1)^2$
   $[8, -6, 1, 1, 0 , 0, 0]^T$

b. $(x1-3)^2 + (x2-4)^2 = 1$
   $= (x1)^2-6x1+9+(x2)^2-8x2+16-1$
   $= 24-6x1-8x2+(x1)^2+(x2)^2$
   $[24, -6, -8, 1, 0 , 1]^T$

c. $2(x1-3)^2 + (x2-4)^2 = 1$
   $= 2(x1)^2 – 12x1 + 18 + (x2)^2 -8x2+16 -1$
   $= 33-12x1-8x2+2(x1)^2+(x2)^2$
   $[33, -12, -8, 2, 0, 1]^T$

---

7. Attached ipynb file

Linear regression provides more realistic separation function.

---

8. Attached Q8.ipynb file
The number of iterations drops drastically as sep approaches 1 and then remain low, which means that when sep >= 1.0, PLA converges very quickly.

9.
a. It will not converge
b. Pocket algorithm is taking too long.
Attached Q9.ipynb file

---

10. (a). Cost of acceptance = P[y = +1|x] * 0 + P[y = -1|x] * Ca
= (1-g(x))*Ca

Cost of rejection = P[y = +1|x] * Cr + P[y = -1|x] * 0
= g(x)*Cr

(b). (1-g(x))*Ca <= g(x)*Cr
Ca <= g(x) * (Ca + Cr)
g(x) >= Ca / (Ca + Cr)

(c). k = 1 / 11 for supermarket
and, k = 1000/1001 for CIA

Sources

• Learning from data.pdf
• http://book.caltech.edu/bookforum/