Amy Pitts
Assignment Number 2
Instructor: Dr. Pablo Rivas

Math 440L: Machine Learning
Due: February 30, 2019

Exercise List

Chapter 2,3

Exercise 2.1. *In Equestion (2.1), set* $\delta = 0.03$ *and let*

$$\epsilon(M, N, \delta) = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}.$$

a) *For* $M = 1$*, how many examples do we need to make* $\epsilon \leq 0.05$*?*

b) *For* $M = 100$*, how many examples do we need to make* $\epsilon \leq 0.05$*?*

c) *For* $M = 10,000$*, how many examples do we need to make* $\epsilon \leq 0.05$*?*

Solution. **a)** We want to obtain $\sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} \leq 0.05$. Using the information provided we have $\sqrt{\frac{1}{2N} \ln \frac{2}{0.03}} \leq 0.05$. Solving for N we get

$$\begin{aligned}
\sqrt{\frac{1}{2N} \ln \frac{2}{0.03}} &\leq 0.05 \\
\frac{1}{2N} \ln \frac{2}{0.03} &\leq 0.0025 \\
\frac{1}{2N}(\ln 2 - \ln 0.03) &\leq 0.0025 \\
\frac{1}{2N} &\leq \frac{0.0025}{(\ln 2 - \ln 0.03)} \\
2N &\geq \frac{\ln 2 - \ln 0.03}{0.0025} \\
N &\geq \frac{\ln 2 - \ln 0.03}{0.005} \\
N &\geq 839.941.
\end{aligned}$$

Thus, $N = 840$ when $M = 1$.

b) We want to obtain $\sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}} \leq 0.05$. Using the information provided we have $\sqrt{\frac{1}{2N}\ln\frac{200}{0.03}} \leq 0.05$. Solving for N with the same steps above we obtain

$$\sqrt{\frac{1}{2N}\ln\frac{200}{0.03}} \leq 0.05$$
$$N \geq \frac{\ln 200 - \ln 0.03}{0.005}$$
$$N \geq 1760.98.$$

Thus, $N = 1761$ when $M = 100$.

b) We want to obtain $\sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}} \leq 0.05$. Using the information provided we have $\sqrt{\frac{1}{20N}\ln\frac{20000}{0.03}} \leq 0.05$. Solving for N with the same steps above we obtain

$$\sqrt{\frac{1}{2N}\ln\frac{20000}{0.03}} \leq 0.05$$
$$N \geq \frac{\ln 20000 - \ln 0.03}{0.005}$$
$$N \geq 2682.01.$$

Thus, $N = 2683$ when $M = 10,000$.

Exercise 2.11. *Suppose $m_H(N) = N + 1$, so $d_{VC} = 1$. You have 100 training examples. Use the generalization bound to give a bouond for E_{out} with confidence 90% . Repeat for $N = 10,000$.*
Hint: you should use equation (2.12) and remember that a confidence of 90% means a $\delta = 0.1$.

Solution. $E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N}\ln\frac{4m_H(2N)}{\delta}}$ plugging in information we have

$$\begin{aligned}
E_{out}(g) &\leq E_{in}(g) + \sqrt{\frac{8}{100}\ln\frac{4m_H(200)}{0.1}} \\
&= E_{in}(g) + \sqrt{\frac{8}{100}\ln\frac{4(200+1)}{0.1}} \\
&= E_{in}(g) + \sqrt{\frac{8}{100}\ln\frac{804}{0.1}} \\
&= E_{in}(g) + 0.8481596
\end{aligned}$$

When $N = 10000$ we have

$$\begin{aligned}
E_{out}(g) &\leq E_{in}(g) + \sqrt{\frac{8}{10000}\ln\frac{4m_H(20000)}{0.1}} \\
&= E_{in}(g) + \sqrt{\frac{8}{10000}\ln\frac{4(20000+1)}{0.1}} \\
&= E_{in}(g) + \sqrt{\frac{8}{10000}\ln\frac{80004}{0.1}} \\
&= E_{in}(g) + 0.1042782
\end{aligned}$$

Exercise 2.12. *For an H with $d_{VC} = 10$, what sample size do you need (as prescribed by the generalization bound) to have a 95% confidence that your generalization error is at most 0.05?*
Hint: Use Equation (2.13)

Solution. Using the equation $N \geq \frac{8}{\epsilon^2} \ln\left(\frac{4(2N)^{d_{vc}}+1}{\delta}\right)$ and plugging in we have

$$\begin{aligned} N &\geq \frac{8}{\epsilon^2} \ln\left(\frac{4(2N)^{d_{vc}}+1}{\delta}\right) \\ &= \frac{8}{0.05^2} \ln\left(\frac{4(2N)^{10}+1}{0.05}\right) \\ &= 3200 \ln\left(81920N^{10}+20\right) \end{aligned}$$

We want to find the N which makes the equation true.

```
import numpy as np


n = 2000
Sn = 3200 * np.log(81920*n**float(10)+20)
t=True
while(t):
    if(round(Sn,8) == round(n,8)):
        t=False
        print(Sn)
    else:
        n = Sn
        Sn = 3200 * np.log(81920*n**float(10)+20)
```

Using the code above the optimal number for N is 452957 rounding up when comparing 8 decimal places.

Exercise 3.1. *Consider the double semi-circle "toy" learning task below.*

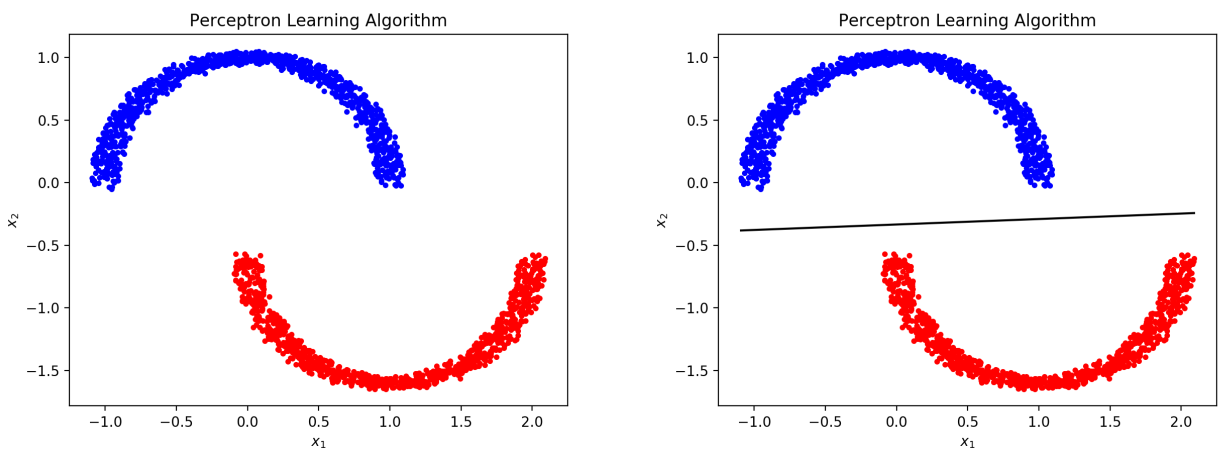
There are two semi-circles of width thk with inner radius rad, separated by sep as shown (red is -1 and blue is +1). The center of the top semi-circle is aligned with the middle of the edge of the bottom semi-circle. This task is linearly separable when sep ≥ 0*, and not so for sep* < 0*. Set rad=10 thk=5 and sep=5. Then, generate 2,000 examples uniformly, which means you will have approximately 1,000 examples for each class*

a) *Run the PLA starting from* $w = 0$ *until it converges. Plot the data and the final hypothesis.*

b) *Repeat part (a) using the linear regression (for classification) to obtain* w*. Explain your observations.*

Use the dataset generator downloadable on Ilearn.

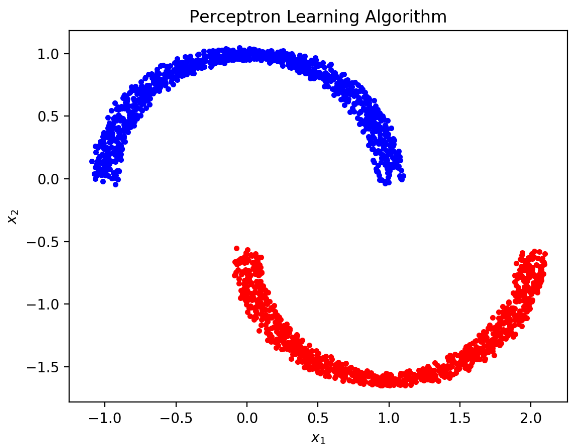
Solution. **a)** The PLA took 5 iteration before it found the ideal line.

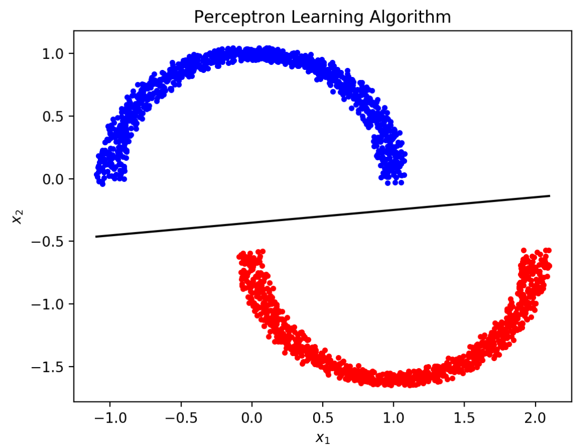


b) When starting the PLA off with linear regression it took zero iterations to find a line that works. Curious to see if what would happen I ran this function several more times and every time it took it 0 iterations to find the correct line. Therefore, starting with linear regression can decrease run time because it helps the pla converge faster. The code used to replace the vector of zeros is listed below.

```
#linear regression

Xs = np.linalg.pinv(X.T.dot(X)).dot(X.T)

wlr = Xs.dot(y)
```

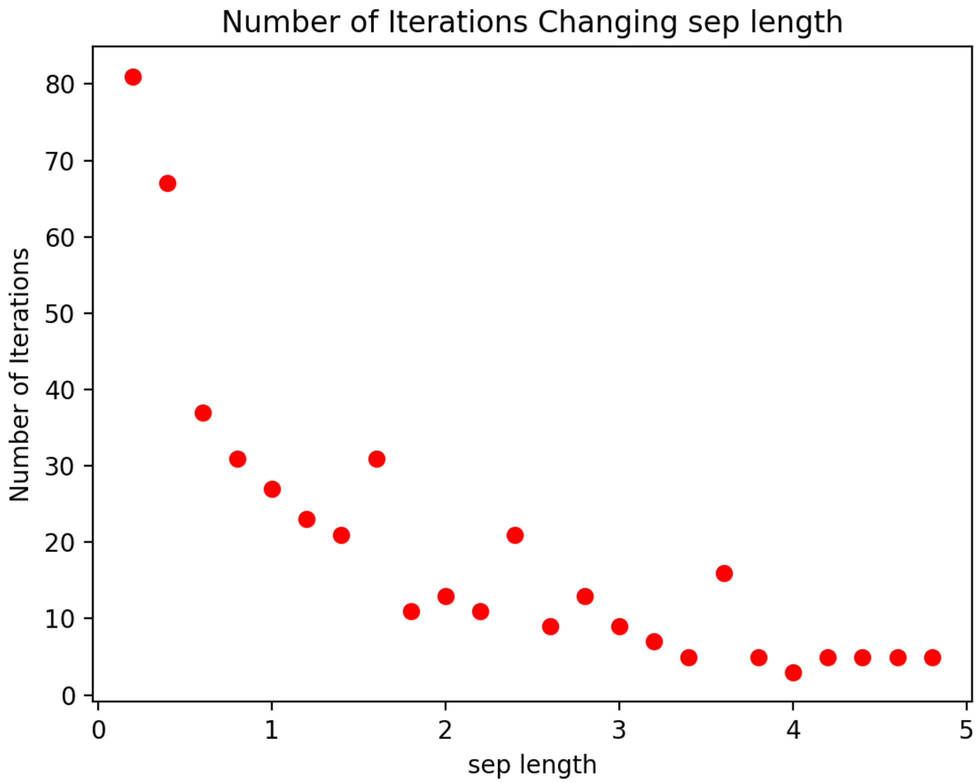
```
# initialize the weigths to linear regression

w = wlr
```





Exercise 3.2. ***Extra Credit*** *For the double-semi-circle task in Problem 3.1, vary sep in the range* $\{0.2, 0.4, \ldots, 5\}$. *Generate 2,000 examples and run the PLA starting with* $w = 0$. *Record the number of iterations PLA takes to converge.*
Plot sep versus the number of iterations taken for PLA to converge. Explain your observations
Hint: Problem 1.3

Solution. The bigger the sep number the less iterations it takes to find the optimal line to split the data. There seems to be a general trend from left to right of the number of iterations goes down as the sep length increase.


Number of Iterations Changing sep length
Number of Iterations
sep length
0
10
20
30
40
50
60
70
80
0
1
2
3
4
5