CSCI 5521: Introduction to Machine Learning (Spring 2019) Homework 0

Christopher White

February 6, 2019

Problem 1.

1)

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{w}}||\mathbf{X}\,\mathbf{w}-\mathbf{y}||^2=0$$
$$||\mathbf{X}\,\mathbf{w}-\mathbf{y}||^2=(\mathbf{X}\,\mathbf{w}-\mathbf{y})'\,(\mathbf{X}\,\mathbf{w}-\mathbf{y})\text{ where }'\text{ is conjugate transpose}$$
$$||\mathbf{X}\,\mathbf{w}-\mathbf{y}||^2=(\mathbf{w}'\,\mathbf{X}'-\mathbf{y}')(\mathbf{X}\,\mathbf{w}-\mathbf{y})$$
$$||\mathbf{X}\,\mathbf{w}-\mathbf{y}||^2=\mathbf{w}'\,\mathbf{X}'\,\mathbf{X}\,\mathbf{w}-\mathbf{y}'\mathbf{X}\,\mathbf{w}-\mathbf{w}'\mathbf{X}'\mathbf{y}+\mathbf{y}'\,\mathbf{y}$$
$$\mathbf{y}'\,\mathbf{X}\,\mathbf{w}=\mathbf{w}'\,\mathbf{X}'\,\mathbf{y}\text{ because each side is a 1x1 matrix}$$
$$||\mathbf{X}\,\mathbf{w}-\mathbf{y}||^2=\mathbf{w}'\,\mathbf{X}'\,\mathbf{X}\,\mathbf{w}-2\,\mathbf{w}\,\mathbf{X}'\;\;\mathbf{y}+\mathbf{y}'\,\mathbf{y}$$
$$-2\;\,\mathbf{X}'\,\mathbf{y}+\frac{\mathrm{d}}{\mathrm{d}\mathbf{w}}\mathbf{w}'\,\mathbf{X}'\,\mathbf{X}\,\mathbf{w}=0$$

It can be shown that $\frac{\mathrm{d}}{\mathrm{d}\mathbf{w}}\mathbf{w}'\,\mathbf{X}'\,\mathbf{X}\,\mathbf{w}^{'}=2\,\mathbf{X}'\,\mathbf{X}\,\mathbf{w}$ from here. Therefore,

$$-2\mathbf{X}'\,\mathbf{y}+2\,\mathbf{X}'\,\mathbf{X}\,\mathbf{w}=0$$
$$\mathbf{X}'\,\mathbf{X}\,\mathbf{w}=\mathbf{X}'\,\mathbf{y}$$

So,

$$\mathbf{w}=(\mathbf{X}'\,\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y}$$

2)

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{w}}||\mathbf{X}\mathbf{w}-\mathbf{y}||^2+\frac{\mathrm{d}}{\mathrm{d}\mathbf{w}}\lambda||\mathbf{w}||^2=0$$
$$2\,\mathbf{X}'\mathbf{X}\mathbf{w}-2\mathbf{X}'\mathbf{y}+\frac{\mathrm{d}}{\mathrm{d}\mathbf{w}}\lambda||\mathbf{w}||^2=0$$

The $\mathbf{w}$is a vector so,

$$\frac{\mathrm{d}}{\mathrm{d}w_i}||\mathbf{w}||^2=\sum_{j=1}^{m}\frac{\mathrm{d}}{\mathrm{d}w_i}w_j^2=\frac{\mathrm{d}}{\mathrm{d}w_i}w_i^2=2w_i$$
$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{w}}||\mathbf{w}||^2=2\mathbf{w}$$

Therefore,

$$
\begin{aligned}
&2\mathbf{X}'\mathbf{X}\mathbf{w} - 2\mathbf{X}'\mathbf{y} + 2\lambda\mathbf{w} = 0 \\
&(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})\mathbf{w} = \mathbf{X}'\mathbf{y} \\
&\mathbf{w} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}
\end{aligned}
$$

Problem 2.

1)

$$
\begin{aligned}
&\Pr(H,H,T,T,H) = p * p * (1-p) * (1-p) * p = p^3 - 2p^4 + p^5 \\
&\ln(\Pr(H,H,T,T,H)) = 3\ln(p) + 2\ln(1-p)
\end{aligned}
$$

2)

a) $\Pr\left(\{H,H,T,T,H\}, p = \frac{1}{2}\right) = \Pr\left(\{H,H,T,T,H\} | p = \frac{1}{2}\right) * \Pr\left(p = \frac{1}{2}\right) = \frac{1}{2^6}$

b)

$$
\begin{aligned}
&\Pr\left(\{H,H,T,T,H\}, p = \frac{2}{3}\right) = \Pr\left(\{H,H,T,T,H\} | p = \frac{2}{3}\right) * \Pr\left(p = \frac{2}{3}\right) \\
&= \left(\frac{2}{3}\right)^3 * \left(\frac{1}{3}\right)^2 * \frac{1}{2} = \frac{8}{486}
\end{aligned}
$$

3)

Assume $p \geq 0$

$$
\frac{\mathrm{d}}{\mathrm{d}p}\ln(\Pr(H,H,T,T,H)) = \frac{\mathrm{d}}{\mathrm{d}p}(3\ln(p) + 2\ln(1-p)) = 3\left(\frac{1}{p}\right) - 2\left(\frac{1}{1-p}\right)
$$

The first derivative can be set to zero,

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}p}\ln(\Pr(H,H,T,T,H)) = 0 \\
&\frac{3}{p} - \frac{2}{1-p} = \frac{3 - 3p - 2p}{p(1-p)} = \frac{3 - 5p}{p(1-p)} = 0 \\
&3 - 5p = 0 \\
&p = \frac{3}{5}
\end{aligned}
$$

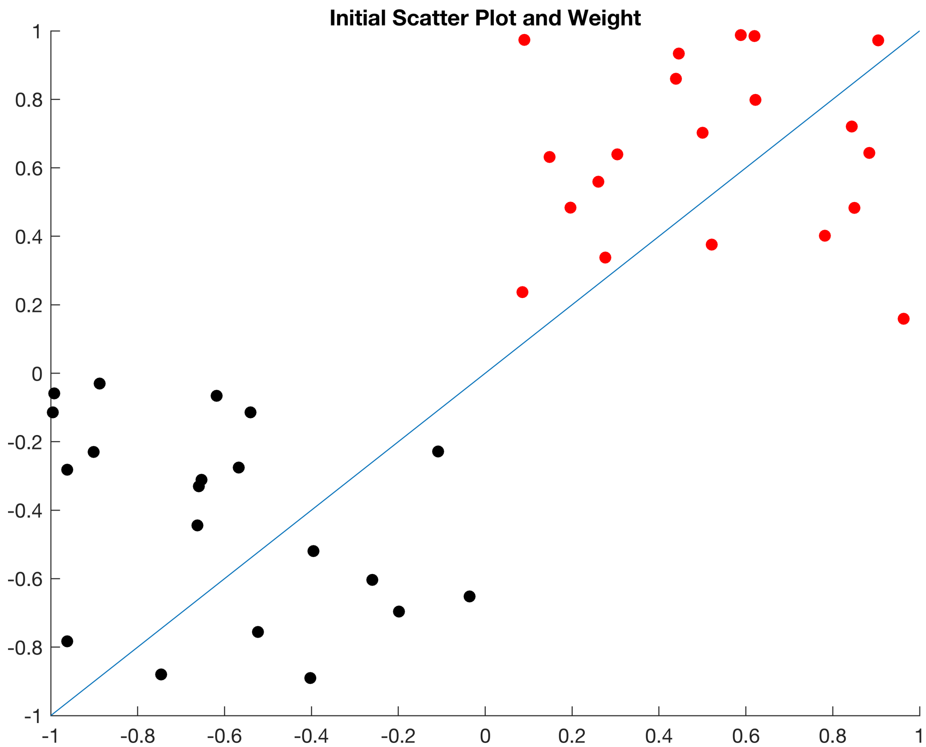
Thus p=3/5 is the maximum likelihood. So, the probability of H,H,T,T,H is:

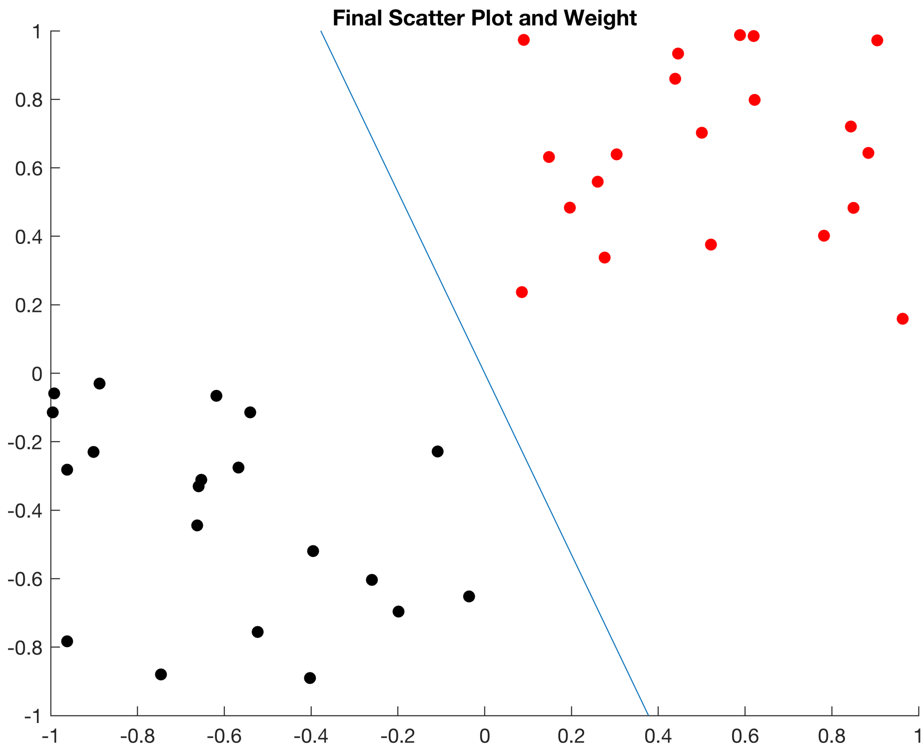
$$
\Pr(H,H,T,T,H) = \left(\frac{3}{5}\right)^3 * \left(\frac{2}{5}\right)^2 = \frac{108}{3125} \approx 0.03456
$$

Problem 3.

1)

```
load data1.mat
w0 = [1;-1];
[w,step] = MyPerceptron(X,y,w0)
```



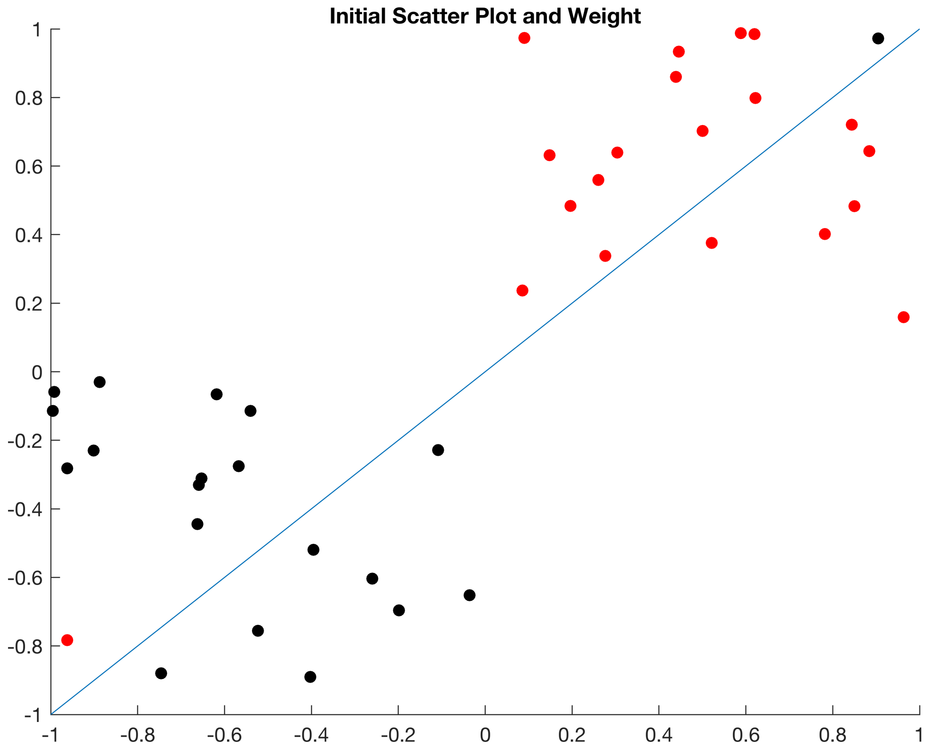


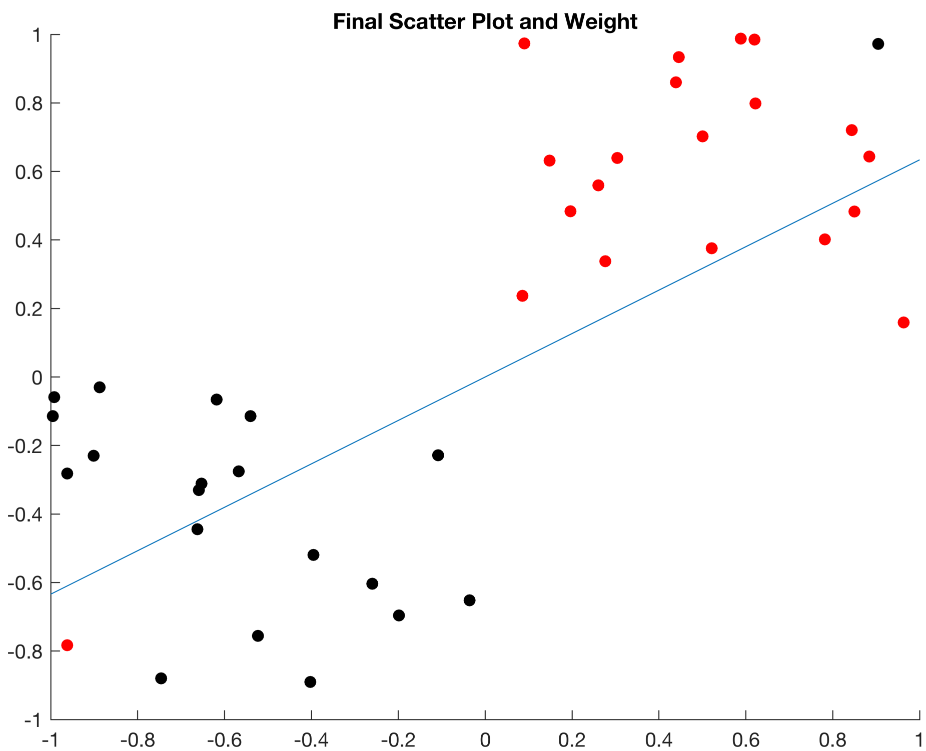

```
w = 2×1
    1.4524
    0.5484
step = 2
```

The data took 2 interations to converge.

2)

```
load data2.mat
w0=[1;-1];
[w,step] = MyPerceptron(X,y,w0)
```






```
w = 2×1
    0.1225
   -0.1932
step = 51
```

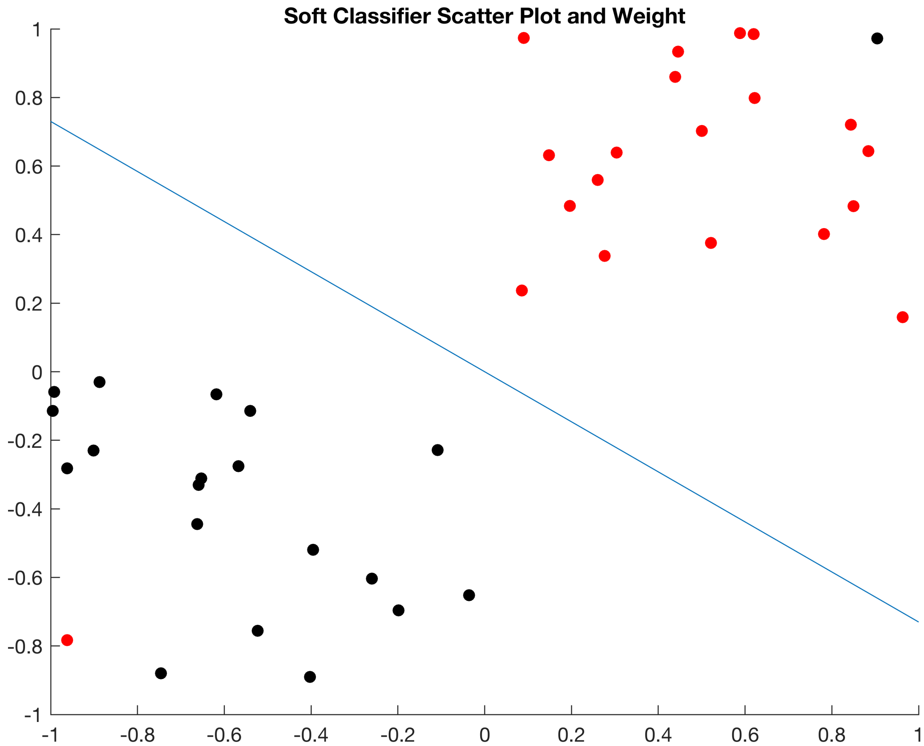
The code was not able to converge after 50 iterations so it terminated. The reason it didn't converge was that there are data points that will always be on the wrong side of the line. And the error is just a discrete count and therefore cannot reach zero in this dataset.

```
load data2.mat
[m,n]=size(X);
f=[zeros(n,1);ones(m,1)]; % transform problem into a standard LP
A1=[X.*repmat(y,1,n),eye(m,m)];
A2=[zeros(m,n),eye(m,m)];
A=-[A1;A2];
b=[-ones(m,1);zeros(m,1)];
x = linprog(f,A,b);% solve LP
```

```
Optimal solution found.
```

```
w=x(1:n);% return varible w

figure;
map = [0 0 0
        1 0 0];
colormap(map);
scatter(X(:,1),X(:,2),[],y,'filled');
xplt = xlim;
yset = ylim;
yplt = -w(1)*xplt/w(2);
line(xplt,yplt);
ylim(yset);
title('Soft Classifier Scatter Plot and Weight');
```


Soft Classifier Scatter Plot and Weight
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1