CSI 5325 Assignment 2

Sadia Nasrin Tisha

1st March 2022

1 Online exercises

2 Problem 1.12

a) Given that,
$E_{in}(h) = \sum_{n=1}^{N}(h - y_n)^2$
$h_{mean} = \frac{1}{N}\sum_{n=1}^{N} y_n$
To minimize the in-sample sum of square deviations, let calculate the derivative of $E_{in}(h)$ and let it equal to 0.
$E_{in}(h) = \sum_{n=1}^{N}(h - y_n)^2$
$\Rightarrow \frac{\partial E_{in}(h)}{\partial h} = 2\sum_{n=1}^{N}(h - y_n)$
$\Rightarrow 0 = 2\sum_{n=1}^{N}(h - y_n)$
$\Rightarrow \sum_{n=1}^{N} y_n = Nh$
$\Rightarrow h = \frac{1}{N}\sum_{n=1}^{N} y_n$
Here, $h = \frac{1}{N}\sum_{n=1}^{N} y_n$ denotes that $h_{mean} = \frac{1}{N}\sum_{n=1}^{N} y_n$.

b) Given that,
$E_{in}(h) = \sum_{n=1}^{N} |h - y_n|$
To minimize the in-sample sum of absolute deviations, let calculate the derivative of $E_{in}(h)$ and let it equal to 0.

$E_{in}(h) = \sum_{n=1}^{N} |h - y_n|$

$\Rightarrow \frac{\partial E_{in}(h)}{\partial h} = \frac{\partial}{\partial h}(\sum_{n=1}^{N} |h - y_n|)$

Let, $\sum_{n=1}^{N}(h - y_n) = u$, so that means $f(x)$ can be written as
$f(x) = |u| = \sqrt{u^2}$
$f'(x) = \frac{1}{2}\frac{2u}{\sqrt{u^2}}\frac{\partial u}{\partial x}$
$\Rightarrow f'(x) = u\frac{u'}{|u|}$
$\Rightarrow f'(x) = \frac{u}{\sqrt{u^2}}$
$\Rightarrow f'(x) = \frac{u}{|u|}$
So that,
$\frac{\partial E_{in}(h)}{\partial h} = \sum_{n=1}^{N} \frac{h-y_n}{|h-y_n|}$

$\Rightarrow 0 = \sum_{n=1}^{N} \frac{h-y_n}{|h-y_n|}$
$\Rightarrow \sum_{n=1}^{N} sign(h - y_n) = 0$
As h_{med} denotes half of the data points are at most h_{med} and half the data points are at least h_{med}. So $\sum_{n=1}^{N} sign(h - y_n) = 0$, denotes the number of at most(positive) terms are equal to the number of at least(negative) terms. So the estimation will be in sample median h_{med}.

c) From a,b we get that,
$h_{mean} = \frac{1}{N} \sum_{n=1}^{N} y_n$
And, $sign \sum_{n=1}^{N} h - y_n = 0$
If y_N is perturbed to $y_N + \epsilon$, where $\epsilon \to \infty$. h_{mean} will increase as y_N contributes to the calculation. So we get $h_{mean} \to \infty$.On the other hand, h_{med} will remain unchanged as in the calculation, h_{med} only requires $h_{med} < y_N$.

3 Problem 2.24

Here,Given that,
Target function, $f(x) = x^2$
Dataset, $\mathbf{D} = (x_1, x_1^2), (x_2, x_2^2)$
$h(x) = ax + b$

Here, $E_{in}(g) = \sum_{i=1}^{2} [f(x_i) - h(x_i)]^2 = \sum_{i=1}^{2} [x_i^2 - (ax_i + b)]^2$

By calculate the derivative of $E_{in}(g)$ with respect to ab,
$\frac{\partial E_{in}(g)}{\partial a} = -2 \sum_{i=1}^{2} x_i(x_i^2 - ax_i - b) = 0$

$\frac{\partial E_{in}(g)}{\partial b} = -2 \sum_{i=1}^{2} (x_i^2 - ax_i - b) = 0$

By solving both equation, we get that,
$a = x_1 + x_2$
$b = -x_1 x_2$

That means the final hypothesis, $g^D(x) = ax + b = (x_1 + x_2)x - x_1 x_2$

Given that, the input variable x is uniformly distributed in the interval [-1,1],
Here,

$$E[x] = \frac{1}{2} \int_{-1}^{1} x \, \mathrm{d}x = \tfrac{1}{2} \left[\frac{x^2}{2} \right]_{-1}^{1} = 0$$

$$E[x^2] = \frac{1}{2} \int_{-1}^{1} x^2 \, \mathrm{d}x = \tfrac{1}{2} \left[\frac{x^3}{3} \right]_{-1}^{1} = 1/3$$

$$E[x^3] = \frac{1}{2} \int_{-1}^{1} x^3 \, \mathrm{d}x = \tfrac{1}{2} \left[\frac{x^4}{4} \right]_{-1}^{1} = 0$$

$$E[x^4] = \frac{1}{2} \int_{-1}^{1} x^4 \, \mathrm{d}x = \tfrac{1}{2} \left[\frac{x^5}{5} \right]_{-1}^{1} = 1/5$$

a) The analytic expression for the average function $\bar{g}(x)$,
$\bar{g}(x) = E_D[g^D(x)]$
$\Rightarrow \bar{g}(x) = E_D[(x_1 + x_2)x - x_1x_2]$
$\Rightarrow \bar{g}(x) = E_D[x_1x] + E_D[x_2x] - E_D[x_1x_2]$
$\Rightarrow \bar{g}(x) = E_D[x_1]x + E_D[x_2]x - E_D[x_1]E_D[x_2]$
$\Rightarrow \bar{g}(x) = 0$ [By applying the value]
So, $\bar{g}(x) = 0$

b) For describe an experiment, let take 2000 data sample and 2000 x sample. The Target function, $f(x) = x^2$ and $h(x) = ax + b$.
For calculating the bias, variance and E_{out} for 2000 data sample, calculate the $\bar{g}(x)$ for each sample from the mean of x using the $h(x)$ function. From the $\bar{g}(x)$, calculate bias(bias=$(\bar{g}(x) - h(x))^2$). Then also calculate the variance (variance=$([g^D(x)]^2 - \bar{g}(x)^2)$ of 2000 sample, where $h(x)$ calculate the $g^D(x)$.
Then calculate the E_{out}= $(g^D(x) - f(x))^2$, from $h(x)$ for 2000 data sample.
Finally, calculate the whole calculation for 2000 x samples. And from bias, variance and E_{out} for all sample find the mean bias, variance and E_{out} and from the generated hypothesis of 2000 sample, find the average hypothesis which is $\bar{g}(x)$ and also calculate the deviation from the true function based on target function.
After running the experiment,
Bias=0.1932961127591476
Variance= 0.32820260059550616
E_{out} = 0.5212010382103563

c) From the experiment, E_{out} = 0.5212010382103563.
Bias+ variance= 0.5214987133546538, which is almost same.

Figure 1 shows the plot of $\bar{g}(x)$ and $f(x)$, where target function shows the x^2 curve and the $\bar{g}(x)$ is almost zero.

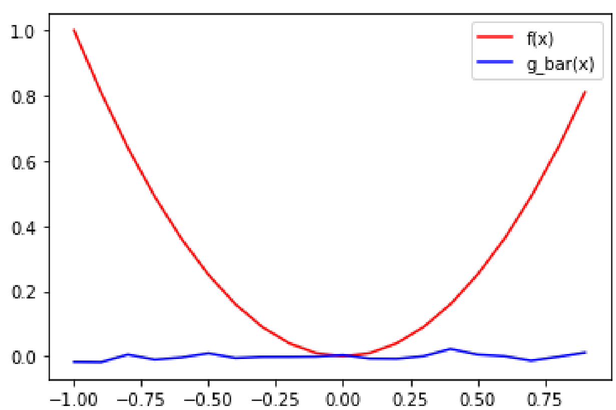

Figure 1: **f(x) and $\bar{g}(x)$ plot**

d) Here, $E_D[E_{out}(g^D)] = Variance + Bias$

Variance= $E_x[E_D[g^D(x)]^2 - \bar{g}(x)^2]$
$= E_x[E_D[(x_1 + x_2)x - x_1x_2]^2 - 0]$
$= E_x[E_D[(x_1 + x_2)^2x^2 - 2(x_1 + x_2)xx_1x_2 + x_1^2x_2^2]]$
$= E_x[x^2E_D[(x_1 + x_2)^2] - 2xE_D[x_1x_2(x_1 + x_2)] + E_D[x_1^2x_2^2]]$
$= E_x[x^2E_D[x_1^2 + x_2^2 + 2x_1x_2] - 2xE_D[x_1^2x_2 + x_1x_2^2] + E_D[x_1^2x_2^2]]$

$= E_x[x^2[1/3 + 1/3] - 2x[1/3 \times 0 + 1/3 \times 0] + [1/3 \times 1/3]]$
$= E_x[\frac{2}{3}x^2 + \frac{1}{9}]$
$= \frac{2}{3} \times \frac{1}{3} + \frac{1}{9}$
$= \frac{1}{3}$
Bias$= E_x[(\bar{g}(x) - f(x))^2]$
$= E_x[(0 - x^2)^2]$
$= E_x[x^4]$
$= \frac{1}{5}$
So, $E_D[E_{out}(g^D)] = Variance + Bias = \frac{1}{3} + \frac{1}{5} = \frac{8}{15}$

4 Problem 3.1

For generate 2000 example uniformly for two semi circle, there approximate 1000 example for each class shows in Figure 2

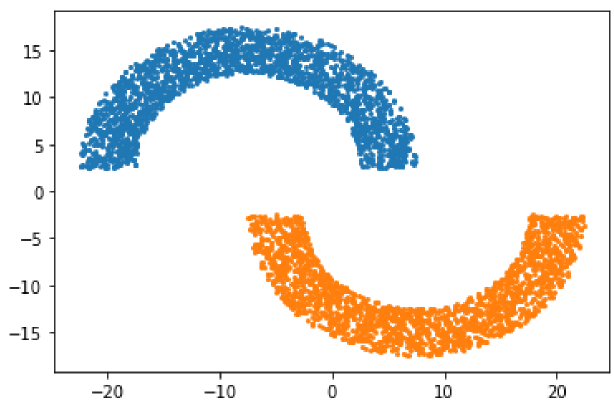

Figure 2: **Generate 2000 example uniformly for two semi circle**

a) After run PLA starting from $w = 0$ until it converge, Figure 3 shows the plot of data and hypothesis where PLA works well to separate the data linearly and achieve close solution.

b) After run Linear regression (for classification), the weight w_{lin} obtain approximate close solution like PLA and the hyperplane separate the data point very well shows in Figure 4

   Figure 5 shows the comparison plot between PLA and Linear regression where both are approximately close.

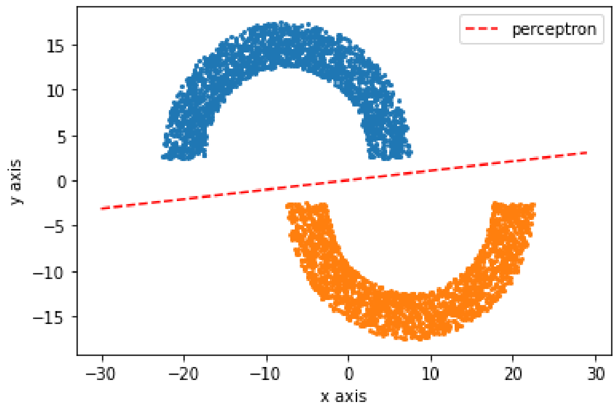

Figure 3: **Applying PLA**

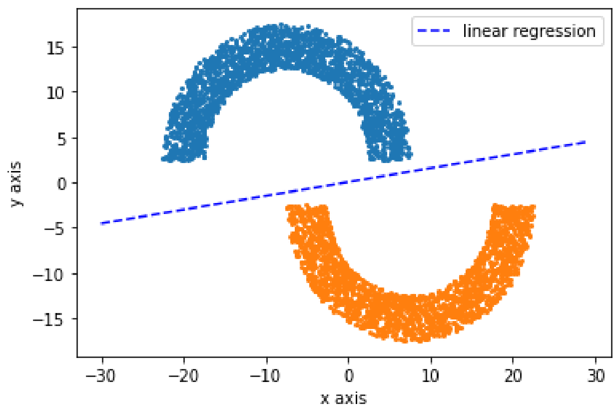

Figure 4: **Applying Linear Regression**

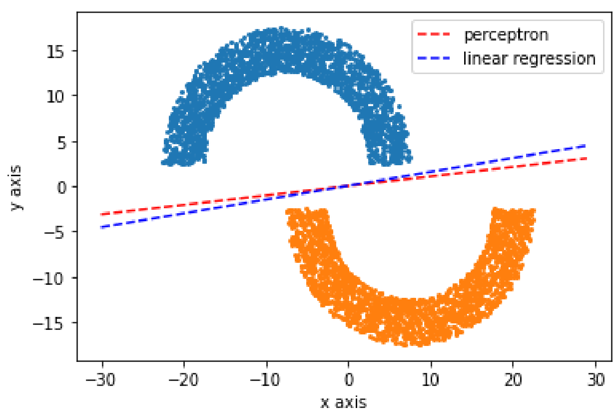

Figure 5: **Applying PLA Linear regression**