Assignment 02

Mushfika Rahman

March 01, 2022

1 Learning From Data exercises

Problem 1.12(a)

If your algorithm is to find the hypothesis h that minimizes the in-sample sum of squared deviations, then show that the estimate will be in-sample mean.

We know that in-sample error is for sum of squared deviation is,

$$E_{in}(h) = \sum_{n=1}^{N}(h - y_n)^2$$

To minimize the function we take the partial derivate respect to h and set to 0.

$$\begin{aligned}
\frac{\partial E_{in}(h)}{\partial h} =& \frac{\partial}{\partial h}\sum_{n=1}^{N}(h - y_n)^2 \\
0 =& 2\sum_{n=1}^{N}(h - y_n) \\
Nh =& \sum_{n=1}^{N} y_n \\
h =& \frac{1}{N}\sum_{n=1}^{N} y_n
\end{aligned} \tag{1}$$

Above equation denotes the mean of h. Thus, by minimizing E_{in} we get, $h_{mean} = \frac{1}{N}\sum_{n=1}^{N} y_n$

Problem 1.12(b)

If your algorithm is to find the hypothesis h that minimizes the in-sample sum of absoulate deviations, then show that the estimate will be in-sample median.

We know that in-sample error is,

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

To minimize the function we take the partial derivate respect to h and set to 0.

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

Now, let us consider the only the $\frac{\partial}{\partial h}|h - y_n|$

$$\frac{\partial}{\partial h}|h - y_n| = \frac{\partial}{\partial h}\sqrt{(h-y)^2} = \frac{h-y}{|h-y_n|}$$

Now,

$$\frac{\partial}{\partial h}\sum_{n=1}^{N}|h - y_n| = \sum_{n=1}^{N}\frac{h-y_n}{|h-y_n|}$$

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

So we get make the derivative 0 and get,

$$sign\sum_{n=1}^{N}(h-y_n) = 0$$

Above equation denotes thh h = $med x_1, x_2, ..., x_n$the derivative is equal to zero when the number of positive terms and negative terms are equal.

Problem 1.12(c)

if y_n is perturbed to $y_n + \epsilon$, where $\epsilon \to \infty$, the single data point becomes outlier. h_{med} will remain the same. As y_n contributes to the calculation so h_{mean} would change. .

Problem 2.24(a)

Give the analytical experience for the average function $\overline{g}(x)$

We know that,

$$\begin{aligned}
\overline{g}(x) =&\mathbb{E}(g^D(x)) \\
=&\mathbb{E}(\frac{y_1 - y_2}{x_1 - x_2}x + \frac{y_2 x_1 - y_1 x_2}{x_1 - x_2}) \\
=&\mathbb{E}(\frac{x_1^2 - x_2^2}{x_1 - x_2}x + \frac{x_2^2 x_1 - x_1^2 x_2}{x_1 - x_2}) \\
=&\mathbb{E}((x_1 + x_2)x - x_2 x_1) \\
=&\mathbb{E}(x_1 x) + \mathbb{E}(x_2 x) - \mathbb{E}(x_2 x_1) \\
=&0
\end{aligned} \tag{2}$$

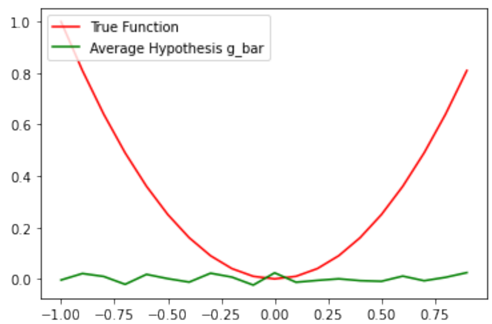
Problem 2.24(b)

Describe an experiment that you could run to determine $\overline{g}(x)$,E$_{out}$, bias , var.

To derive an experiment, firstly we need generate certain example points, in this case number of example point is 1000. Then would calculate the $g^D(x)$ hypothesis, in here the hypothesis is $h(x) = ax + b$. Then would calculate bias, variance for that data samples. This whole process would run for again certain numbers of example let's see it would run again for 1000 times. From the generated hypothesis we would take the avarage of the hypothesis that would be $\overline{g}(x)$. While the hypothesis was taken we calculate the deviation from the true function, which is $f(x) = x^2$. The deviation is the E(E^D_{out}). Again, the experiment would take the mean of the bias, variance. These calculates mean of bias and variance would be bias and variance for the 1000 data samples. Finally the E_{out} comes from adding bias with variance.

Problem 2.24(c)

Plot your $\overline{g}(x)$ with $f(x)$



The plot has $\overline{g}(x)$ with $f(x)$ together.

E_{out} = 0.53564507853184 and

bias+variance = 0.5335426968307737

E_{out} and bias+variance are very close.

Problem 2.24(d)

Give the analytical experience for the average function $Eout$, bias, var. We know that,

$$\begin{aligned}
Variance(x) =& \mathbb{E}_{\mathbb{D}}(g^D(x) - \overline{g}(x))^2 \\
=& \mathbb{E}_{\mathbb{D}}(g^D(x)^2) - 0 \\
=& \mathbb{E}_{\mathbb{D}}((x_1 + x_2)^2 x^2) - 2\mathbb{E}_{\mathbb{D}}((x_2 + x_1)x_1x_2)x - \mathbb{E}_{\mathbb{D}}(x_2^2 x_1^2) \\
=& \mathbb{E}_{\mathbb{D}}((x_1^2 + x_2^2 + 2x_1x_2)x^2) - 2\mathbb{E}_{\mathbb{D}}((x_2 + x_1)x_1x_2)x + \mathbb{E}_{\mathbb{D}}(x_2^2 x_1^2) \\
=& \mathbb{E}_{\mathbb{D}}((x_1^2 + x_2^2 + 2x_1x_2)x^2) - 2\mathbb{E}_{\mathbb{D}}((x_2^2 x_1 + x_1^2 x_2)x + \mathbb{E}_{\mathbb{D}}(x_2^2 x_1^2) \\
=& \frac{2}{3}x^2 + \frac{1}{9}
\end{aligned} \tag{3}$$

then we get,

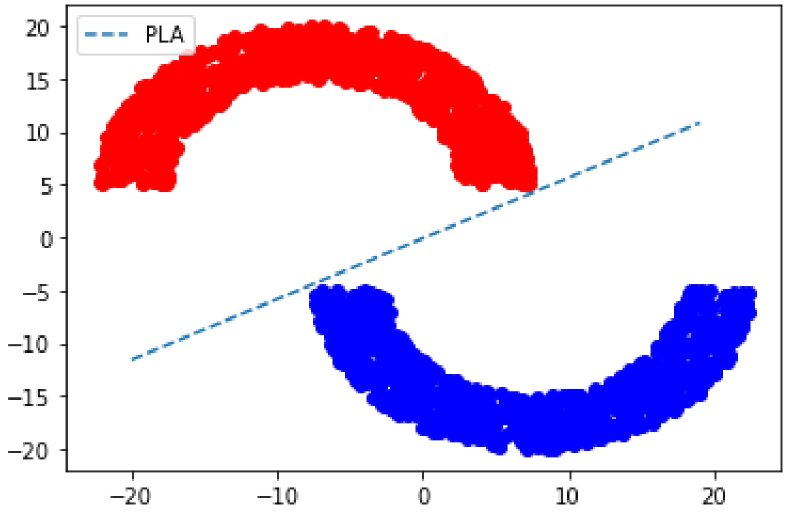
$$
\begin{aligned}
var =& \mathbb{E}_{\mathbb{X}}(\frac{2}{3}x^2 + \frac{1}{9}) \\
=& (\frac{1}{3})
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
bias =& \mathbb{E}_{\mathbb{X}}(\overline{g}(x) - f(x))^2 \\
=& \mathbb{E}_{\mathbb{X}}(0 - f(x))^2 \\
=& \mathbb{E}_{\mathbb{X}}(0 - x^2)^2 \\
=& \mathbb{E}_{\mathbb{X}}(x^4) \\
=& \frac{1}{5} && (5) \\
& && (6)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_{\mathbb{D}}(E_{out}(g^D)) =& bias + var \\
=& (\frac{8}{15})
\end{aligned} \tag{7}
$$

Problem 3.1(a)

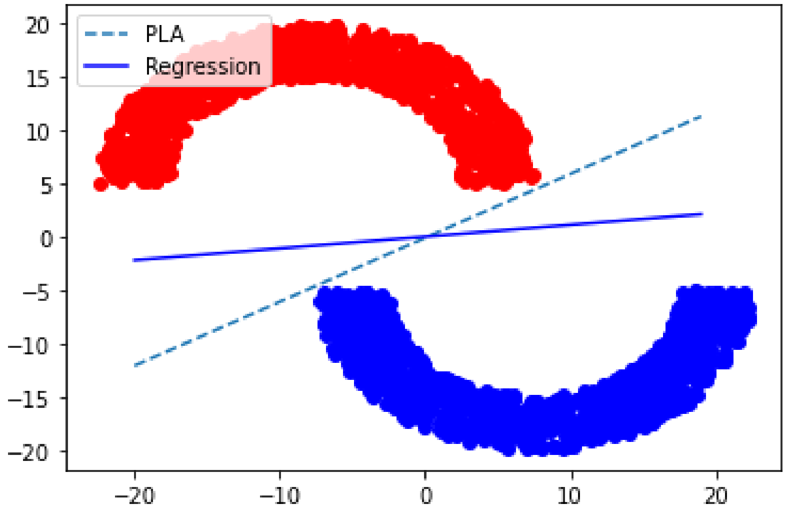
Run the PLA starting from w=0 untill it converges. Plot the data and the final hypothesis.



).png

Problem 3.1(b)

Run the part(a) using the linear regression to obtain w.



The plot shows both the regression and PLA separating boundary. The PLA took longer time to converge than the regression. The regression separated the regions very well. The weights for the regression and PLA was diverse. The weight of regression is very small 0.00735986, -0.066499106 and the bias is also small which as -0.002327.