October 1, 2022
Homework 2
Sean Collings
#A02278024

1 Problem 1

1.

$$\begin{aligned}
B[\bar{X}] &= E[\bar{X}] - \mu && (1)\\
&= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - \mu && (2)\\
&= \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] - \mu && (3)\\
&= \frac{1}{n}\sum_{i=1}^{n} \mu - \mu && (4)\\
&= \frac{1}{n} n\mu - \mu && (5)\\
&= 0 && (6)
\end{aligned}$$

2.

$$\begin{aligned}
Var[\bar{X}] &= Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] && (7)\\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var\left[X_i\right] && (8)\\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 && (9)\\
&= \frac{1}{n^2} n\sigma^2 && (10)\\
&= \frac{\sigma^2}{n} && (11)\\
& && (12)
\end{aligned}$$

3.

$$\begin{aligned}
MSE(\bar{X}) &= E[(\bar{X}-\mu)^2] && (13)\\
&= E[\bar{X}^2 - 2\bar{X}\mu + \mu^2] && (14)\\
&= E[\bar{X}^2] - 2E[\bar{X}\mu] + E[\mu^2] && (15)\\
&= Var(\bar{X}) + E[\bar{X}]^2 - 2E[\bar{X}\mu] + \mu^2 && (16)\\
&= Var(\bar{X}) + (E[\bar{X}] - \mu)^2 && (17)\\
&= Var(\bar{X}) + B[\bar{X}]^2 && (18)\\
&= \frac{\sigma^2}{n} + 0 && (19)\\
& && (20)
\end{aligned}$$

4. It can be shown that $B[\hat{s}^2] = \frac{n-1}{n}\sigma^2 - \sigma^2 \neq 0$ which is biased.

$$\begin{aligned}
B[\hat{s}^2] &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] - \sigma^2 && (21)\\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}((X_i-\mu) - (\bar{X}-\mu))^2\right] - \sigma^2 && (22)\\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}((X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}-\mu) + (\bar{X}-\mu)^2)\right] - \sigma^2 && (23)\\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{1}{n}\sum_{i=1}^{n}2(X_i-\mu)(\bar{X}-\mu) + \frac{1}{n}\sum_{i=1}^{n}(\bar{X}-\mu)^2\right] - \sigma^2 && (24)\\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\bar{X}-\mu)^2\right] - \sigma^2 && (25)\\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right] - E\left[(\bar{X}-\mu)^2\right] - \sigma^2 && (26)\\
&= \sigma^2 - \frac{\sigma^2}{n} - \sigma^2 && (27)\\
&= \frac{n-1}{n}\sigma^2 - \sigma^2 \neq 0 && (28)
\end{aligned}$$

An unbiased estimator of σ^2 would be $\hat{s}_u^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ because

$$\begin{aligned}
B\left[\hat{s}_u^2\right] &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] - \sigma^2 && (29)\\
&= \frac{n}{n-1}E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] - \sigma^2 && (30)\\
&= \frac{n}{n-1}\left[\frac{n-1}{n}\sigma^2\right] - \sigma^2 && (31)\\
&= \sigma^2 - \sigma^2 && (32)\\
&= 0 && (33)
\end{aligned}$$

2 Problem 2

1. Consider the perceptron

$$f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

Multiplying the weights w and bias b of this perceptron by some $c > 0$ will never change the output for x. This is because multiplying by a positive constant will not change the parity of $w \cdot x + b$, only the magnitude. In other words, a positive $w \cdot x + b$ will remain positive and a negative $w \cdot x + b$ will remain negative.

This can easily be generalized to an entire neural network of perceptrons. If each individual perceptron is not affected by the scalar c, then neither will the behavior of the entire network.

2.

$$\lim_{c\to\infty} \frac{1}{1+e^{-cz}} \text{ where } z = w \cdot x + b \neq 0$$

Because z may be positive or negative, there are two possible results as c increaes

$$\lim_{c\to\infty} \frac{1}{1+e^{-cz}} = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if} z < 0 \end{cases}$$

This is the same behavior as the perceptron network above. Thus, the limit of a sigmoid is equal to a perceptron when $z \neq 0$.

If, for one of the perceptrons, $z = 0$ the results of the limit change for that neuron

$$\lim_{c\to\infty} \frac{1}{1+e^{-cz}} = \frac{1}{2} \text{ where } z = 0$$

In this case, the sigmoid can no longer approximate the perceptron as c tends to infinity.

3. Perceptron Network

X	Output
000	0
001	1
010	0
011	1
100	1
101	1
110	1
111	1

4. Sigmoid Network (to 2 decimal places)

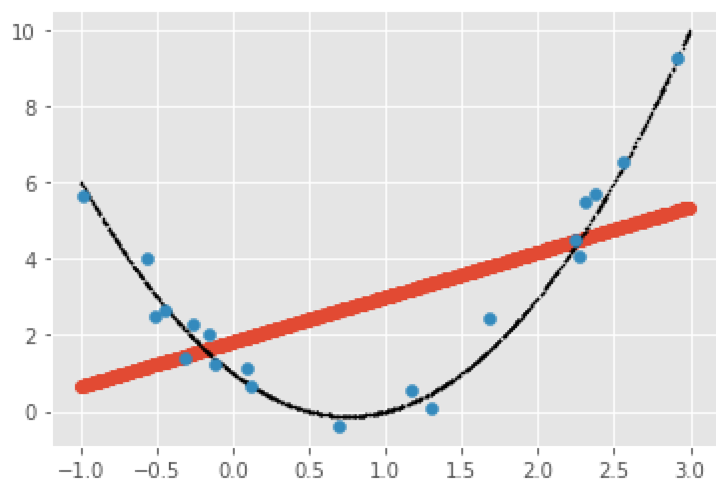
X	Output
000	0.57
001	0.59
010	0.57
011	0.59
100	0.65
101	0.65
110	0.64
111	0.64

3 Problem 3

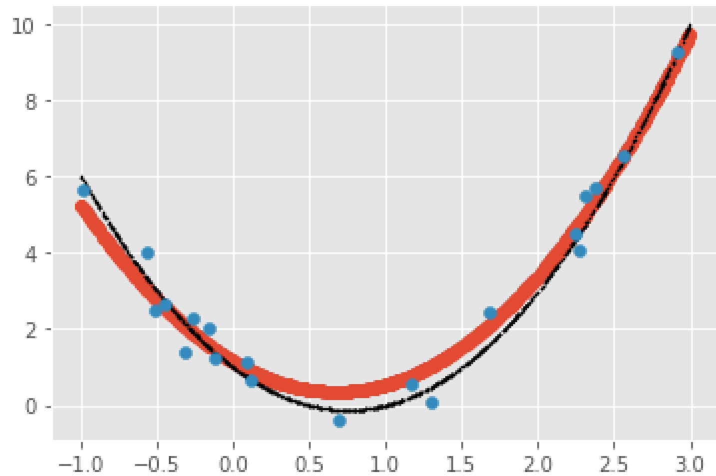
1. Classification: Each output class would represent a kind of tree
2. Regression: This is a perfect $\mathbb{R}^d \to \mathbb{R}$ problem. The network takes in d input variables on an individual's health, and outputs a single value predicting time of death
3. Regression: This problem is not discrete, but continuous, so regression makes more sense
4. Classification: Maps the inputs of browsing history and previous ratings, to a set of classes representing number of stars given: $\mathbb{R}^d \to \{0, 1, 2, 3, 4, 5\}$
5. Classification: The realm of possibilities is discrete: all valid words in the dictionary used

4 Problem 4

1. Linear

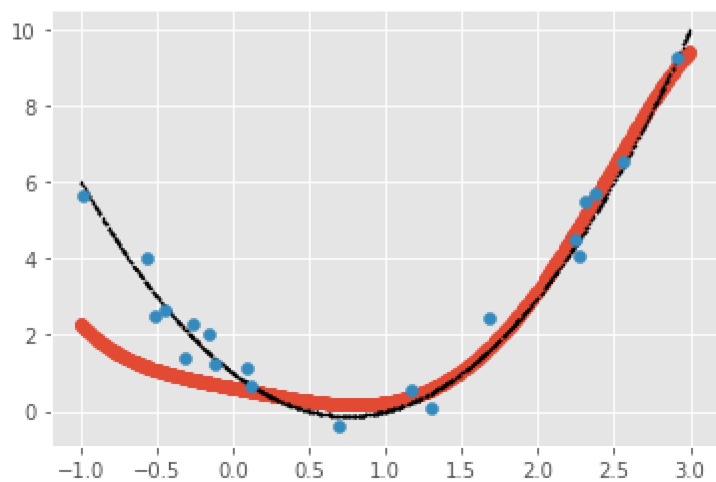


2. Polynomial



This approach approximated the function with these parameters:
$y = 1.7424x^2 - 2.3596x + 1.1556$ where the true function was $y = 2x^2 - 3x + 1$ which is a fair approximation. Notably, it is the furthest off for the linear term.

3. 5-degree polynomial



This approach produced a lower accuracy approximation of the model than the 2nd degree polynomial did. I believe this is because feeding the model the additional information pulls it off the correct track because it's information that will not actually contribute to the real function.

4. 5-degree polynomial

   What follows is the output for each paremeter

```
sigma: 0.1, N:  15, weight decay: 0.0, mse: 14.5104
sigma: 0.1, N:  15, weight decay: 0.2, mse: 13.1583 (Best)
sigma: 0.1, N:  15, weight decay: 0.5, mse: 14.1570

sigma: 0.5, N:  15, weight decay: 0.0, mse: 13.6757
sigma: 0.5, N:  15, weight decay: 0.2, mse: 12.4385 (Best)
sigma: 0.5, N:  15, weight decay: 0.5, mse: 12.9846

sigma:   1, N:  15, weight decay: 0.0, mse: 10.8108
sigma:   1, N:  15, weight decay: 0.2, mse: 11.8270
sigma:   1, N:  15, weight decay: 0.5, mse: 11.0780 (Best)

sigma: 0.1, N: 100, weight decay: 0.0, mse: 13.2202 (Best)
sigma: 0.1, N: 100, weight decay: 0.2, mse: 13.6684
sigma: 0.1, N: 100, weight decay: 0.5, mse: 13.7446

sigma: 0.5, N: 100, weight decay: 0.0, mse: 13.7502
sigma: 0.5, N: 100, weight decay: 0.2, mse: 13.0641 (Best)
sigma: 0.5, N: 100, weight decay: 0.5, mse: 13.4640

sigma:   1, N: 100, weight decay: 0.0, mse: 13.5348
sigma:   1, N: 100, weight decay: 0.2, mse: 13.5143
sigma:   1, N: 100, weight decay: 0.5, mse: 13.2758 (Best)
```