

Name: Muhammad Abdulleh

Assignment No 4

Roll no: 22P-9371

(Data Science)

Section: BCS-6B

Question 1:  $[\theta' = \theta - \alpha \cdot (h_\theta(x_i) - y_i) x_i]$

Sol

Scenario 1: Keep  $\theta_0$  constant, update  $\theta_1$

$$\theta_1' = \theta_1 - \alpha \cdot (h_\theta(x_i) - y_i) x_i, \quad h_\theta(x_i) = \theta_0 + \theta_1 x_i = 0.5 + \theta_1 x_i$$

Iteration 1:

$$x_1 = 800, y_1 = 150$$

$$h_\theta(x_1) = 0.5 + 1 \cdot 800 = 800.5$$

$$\text{Error} = h_\theta(x_1) - y_1 = 800.5 - 150 = 650.5$$

$$\theta_1' = 1 - (0.01) \cdot (650.5) \cdot 800 = -5203$$

Iteration 2:  $(x_2 = 1000, y_2 = 180)$

$$\text{current } \theta_1 = -5203$$

$$h_\theta(x_2) = 0.5 + (-5203) \cdot 1000 = -5202999.5$$

$$\text{Error} = h_\theta(x_2) - y_2 = -5202999.5 - 180 = -5203179.5$$

$$\theta_1' = -5203 - (0.01) \cdot (-5203179.5) \cdot 1000 = 52026592$$

Iteration 3:  $(x_3 = 1200, y_3 = 200)$

$$\text{current } \theta_1 = 52026592$$

$$h_\theta(x_3) = 0.5 + (52026592)(1200) \approx 6243196400$$

$$\text{Error} = h_\theta(x_3) - y_3 = 6243196400.5 - 200 \approx 6243196200$$

$$\theta_1' = 52026592 - (0.01)(6243196200.5)(1200) = -7.4918 \times 10^{11}$$

Iteration 4:  $(x_4 = 1500, y_4 = 250)$

$$\text{current } \theta_1 \approx -7.49182 \times 10^{11}$$

$$h_\theta(x_4) = 0.5 + (-7.49182 \times 10^{11})(1500) \approx -1.12377 \times 10^{15}$$



$$\text{Error} = -1.12377 \times 10^{15} - 280 = -1.12377 \times 10^{15}$$

$$\theta'_1 = -7.49182 \times 10^{14} - (0.01)(-1.12377 \times 10^{15}) \cdot 1800$$

$$\approx 1.68566 \times 10^{16}$$

$\therefore$  Values are too long to process

Iteration 5: ( $x_5 = 1800, y_5 = 280$ )

$$h_{\theta}(x_5) = 0.5 + (1.68566 \times 10^{16})(1800) \approx 3.0339 \times 10^{19}$$

$$\text{Error} = h_{\theta}(x_5) - y_5 = 3.0339 \times 10^{19} - 280 \approx 3.0339 \times 10^{19}$$

$$\theta'_1 = 1.68566 \times 10^{19} - 0.01(3.0339 \times 10^{19}) \cdot 1800 = -5.46118 \times 10^{20}$$

$$\text{Finally: } \theta_0 = 0.5, \theta_1 = -5.46118 \times 10^{20}$$

~~Note: The values are diverging, likely due to the large learning rate or the nature of updating only  $\theta_1$~~

Scenario 2: update  $\theta_0$ , keep  $\theta_1$  constant

$$\text{For } \theta'_0 = \theta_0 - \alpha \cdot (h_{\theta}(x_i) - y_i)$$

Iteration 1: ( $x_1 = 800, y_1 = 150$ )

$$h_{\theta}(x_1) = 0.5 + 1 \cdot 800 = 800.5$$

$$\text{Error} = h_{\theta}(x_1) - y_1 = 800.5 - 150 = 650.5$$

$$\theta'_0 = 0.5 - (0.01)(650.5) = -6.005$$

Iteration 2: ( $x_2 = 1000, y_2 = 180$ )

$$\text{current } \theta_0 = -6.005$$

$$h_{\theta}(x_2) = -6.005 + 1 \cdot (1000) = 993.995$$

$$\text{Error} = h_{\theta}(x_2) - y_2 = 993.995 - 180 = 813.995$$

$$\theta_0^0 = -6.005 - 0.01(813.995) = -14.14495$$

Iteration 3: ( $x_3 = 1200, y_3 = 200$ )



$$\text{current } \theta_0 = -14.14495$$

$$h_0(x_3) = -14.14495 + 1 \cdot (1200) = 1185.85505$$

$$\text{Error} = h_0(x_3) - y_3 = 985.85505$$

$$\theta_0' = -14.14495 - 0.01(985.85505) = -24.0038005$$

Iteration 4: ( $x_4 = 1500, y_4 = 250$ )

$$\text{Current } \theta_0 = -24.0038005$$

$$h_0(x_4) = -24.0038005 + 1 \cdot 1500 = 1475.9961995$$

$$\text{Error} = h_0(x_4) - y_4 = 1225.9961995$$

$$\theta_0' = -24.0038005 - 0.01(1225.9961995) = -36.263465$$

Iteration 5: ( $x_5 = 1800, y_5 = 280$ )

$$\text{Current } \theta_0 = -36.263465$$

$$h_0(x_5) = -36.263465 + 1 \cdot 1800 = 1763.736534505$$

$$\text{Error} = h_0(x_5) - y_5 = 1483.736534505$$

$$\theta_0' = -36.263465 - 0.01(1483.736534505) = -51.10083084005$$

$$\text{Finally: } \theta_0 = -51.10083084005$$

$$\theta_1 = 1 \text{ (unchanged)}$$

Scenario 3: update both  $\theta_0, \theta_1$

$$\theta_0' = \theta_0 - \alpha(h_0(x_i) - y_i) \leftarrow x_i = 1$$

$$\theta_1' = \theta_1 - \alpha(h_0(x_i) - y_i)x_i$$

Iteration 1:

$$h_0(x_1) = 0.5 + 1 \cdot 800 = 800.5$$

$$\text{Error} = 650.5$$

$$\theta_0' = 0.5 - 0.01 \cdot 650.5 = -6.005$$

$$\theta_1' = 0.1 - 0.01 \cdot 650.5 \cdot 800 = -5203$$

Iteration 2:

~~Already have calculations:~~

$$\text{current } \theta_0 = -6.005, \theta_1 = -5203$$

$$h_{\theta}(x_2) = -6.005 + (-5203) \cdot 1000 = -5203006.005$$

$$\text{Error} = h_{\theta}(x_2) - y_2 = -5203186.005$$

$$\theta_0' = -6.005 - 0.01(-5203186.005) = 52025.8805$$

$$\theta_1' = -5203 - 0.01(-5203186.005) = \cancel{52025.8805} (1000) \\ = 52026657.05$$

Iteration 3:

$$\text{current } \theta_0 = 52025.8805, \theta_1 = 52026657.05$$

$$h_{\theta}(x_3) = \theta_0 + \theta_1 x_3 = 62432013845.85505$$

$$\text{Error} = h_{\theta}(x_3) - y_3 = 62432013645.85505$$

$$\theta_0' = \theta_0 - 2(\text{error}) \approx -624268110.6035$$

$$\theta_1' = \theta_1 - 2(\text{error})x_3 \approx -7.44184163 \times 10^{11}$$

could not compute  $I_4$  and  $I_5$  because values are too large negatively and positively after ~~com~~ calculation so i am not understanding how is this beneficial for SDG. However updated  $\theta_0$  and  $\theta_1$  after 5 iterations

$$\text{are } \boxed{\begin{aligned} \theta_0 &\approx -3.03449 \times 10^7 \\ \theta_1 &\approx -5.461553 \times 10^{22} \end{aligned}}$$



Question 2:-

Answer:-

The ethics of data science revolves around ensuring responsible, fair and transparent use of data while minimizing harm and respecting individual rights. Below is a concise explanation of Key ethical considerations:

### Privacy and Consent:

- Respecting individual Privacy: Data Scientist must protect personally identifiable information (PII) and ensure data collection complies with regulations like GDPR or CCPA. Anonymization and encryption are critical to safeguarding privacy.
- Informed Consent: Individuals should be fully informed about how their data will be used and provide explicit consent. Deceptive practices or hidden data collection violate trust.

### Fairness and Bias:

- Avoiding Discrimination: Algorithms can perpetuate or amplify biases present in training data leading to unfair outcomes. Data scientist must audit and mitigate biases.

## Transparency:

Models, especially in high-stakes domains like healthcare or criminal justice should be interpretable to stakeholders. Black-box model erode trust.

## Data Quality and integrity:

Poor quality or manipulated data can lead to flawed insights, harming decision making. Ethical data science prioritize rigorous data validation and cleaning. Data should be obtained ethically avoiding sources that exploit individuals or violates laws

## Security:

Robust cybersecurity measures are essential to prevent breaches that could expose sensitive information. Data Scientist should consider how data misuse could harm individuals and take preventive steps

## Social impact:

Data science should prioritize societal goals such as improving healthcare or education, while minimizing risks like surveillance or loss



autonomy. Large scale data processing and I training consume significant energy, raising ethical questions about environmental sustainability.

### Practical Implications:

- Adopt frameworks like the IEEE ethically Aligned Design or the Data Ethics canvas to guide decision making
- Adhere to laws and standards e.g. GDPR, HIPAA etc.
- Involve communities affected by data driven decisions to ensure their perspective are considered

In summary, ethical data science requires balancing technical proficiency with moral responsibility, prioritizing privacy, fairness transparency etc while mitigating the risks of harm or misuse.