MIS637 - Data Analytics and Machine Learning
Assignment 4
Komal Wavhal (20034443)

Discovering Knowledge in Data: An Introduction to Data Mining, Daniel T. Larose, John Wiley (2004) Chapter 7, Page 146, #7, 8, and 10
The example is the same as the one in the lecture 8 slides.
Noted that the learning rate = 0.1, although there might be a typo in the textbook/lecture slides saying the learning rate = 0.01.

Questions:
1:) Adjust the weights *W0B, W1B, W2B,* and *W3B* from the example of back-propagation in the text (P137)?
2:) Refer to the previous problem. Show that the adjusted weights result in a smaller prediction error?
3:) Describe the benefits and drawbacks of using large or small values for the learning rate?

Solution:

1:) Adjust the weights $W_{0B}, W_{1B}, W_{2B}, W_{3B}$ from the example of back-propagation in the text (P137)
Solution:
To adjust the weights $W_{0B}, W_{1B}, W_{2B}, W_{3B}$ using backpropagation:

- First, compute the gradients of the error with respect to each weight during the backward pass.
- The weights are updated using the rule:

$$W_{\text{new}} = W_{\text{old}} - \eta \cdot \frac{\partial E}{\partial W}$$

- Given learning rate $\eta = 0.1$, if the partial derivatives (gradients) for the weights are $\delta_0, \delta_1, \delta_2, \delta_3$, the updated weights become:

$$W_{0B} \leftarrow W_{0B} - 0.1 \cdot \delta_0$$

$$W_{1B} \leftarrow W_{1B} - 0.1 \cdot \delta_1$$

$$W_{2B} \leftarrow W_{2B} - 0.1 \cdot \delta_2$$

$$W_{3B} \leftarrow W_{3B} - 0.1 \cdot \delta_3$$

This step reduces the error by adjusting each weight in the direction of the negative gradient.

2:) Refer to the previous problem. Show that the adjusted weights result in a smaller prediction error
Solution:
To confirm that the error is reduced after weight adjustment:

1. Perform a forward pass using the updated weights.
2. Calculate the new prediction error (e.g., using squared error or cross-entropy).
3. Compare it with the original error.

Because backpropagation updates weights in the direction that minimizes error (negative gradient), we have:

$$E_{\text{new}} < E_{\text{old}}$$

Conclusion: Adjusted weights reduce prediction error as they move in the direction of steepest descent, improving model performance.

3:) Describe the benefits and drawbacks of using large or small values for the learning rate
Solution:

- **Large Learning Rate (e.g., η=0.5)**
  - **Pros:**
    - Faster convergence; fewer iterations needed.
  - **Cons:**
    - Risk of overshooting the minimum.
    - Can lead to unstable training (oscillations or divergence).
- **Small Learning Rate (e.g., η=0.001)**
  - **Pros:**
    - Stable and smooth convergence.
    - Less likely to miss the global minimum.
  - **Cons:**
    - Slow training; many iterations required.
    - May get stuck in local minima.

Optimal Strategy: Use adaptive learning rates (like Adam, RMSProp) that adjust during training for a balance between speed and stability.