

## Assignment 06: Theoretical Questions

### Question 1: Gradient Descent Calculation

Given the function  $E(w) = w^2 - 4w + 4$ , perform gradient descent to find the minimum value of  $w$  over three iterations:

1. First, calculate the derivative of  $E(w)$ :

$$E'(w) = 2w - 4$$

2. Starting values:

- Initial  $w_0 = 0$
- Learning rate  $\eta = 0.1$

3. Iteration calculations:

- Iteration 1:

$$w_1 = w_0 - \eta \cdot E'(w_0) = 0 - 0.1 \cdot (2 \cdot 0 - 4) = 0.4$$

- Iteration 2:

$$w_2 = w_1 - \eta \cdot E'(w_1) = 0.4 - 0.1 \cdot (2 \cdot 0.4 - 4) = 0.72$$

- Iteration 3:

$$w_3 = w_2 - \eta \cdot E'(w_2) = 0.72 - 0.1 \cdot (2 \cdot 0.72 - 4) = 0.976$$

After three iterations, the value of  $w$  is approximately 0.976.

### Question 2: Linear Discriminant Function

Given  $w_1 = 2, w_2 = -3, w_0 = 5$ :

1. For point  $(1, 2)$ :

$$g(1, 2) = 2 \cdot 1 - 3 \cdot 2 + 5 = 1$$

2. Classification: Since  $g(1, 2) > 0$ , the point belongs to class C1 (Cancer Positive).

3. For point  $(3, 4)$ :

$$g(3, 4) = 2 \cdot 3 - 3 \cdot 4 + 5 = -1$$

4. Classification: Since  $g(3, 4) < 0$ , the point belongs to class C2 (Cancer Negative).

### Question 3: Gradient Descent and Learning Rate

- a) **Gradient descent** is an optimization algorithm used to minimize a function by iteratively moving towards the minimum value of the function. It is used in machine learning to find the optimal parameters of a model that minimize the cost function.
- b) **Learning rate** determines the size of the steps taken in the gradient descent algorithm. A higher learning rate can cause overshooting of the minimum, while a too-small learning rate can result in a long convergence time. It balances the speed and accuracy of convergence.

### Question 4: Linear Discriminant and Ranking

- a) **Linear discriminant models** attribute importance to input features based on the coefficients associated with each feature. Larger absolute values indicate higher importance in class separation.
- b) **Ranking** in machine learning involves ordering or prioritizing items or instances, unlike classification which categorizes items into predefined classes and regression which predicts a continuous value. Ranking is particularly useful in recommendation systems and search engines.

### Question 5: Decision Trees and Pruning

- a) **Pruning in decision trees** is the process of removing parts of the tree to prevent overfitting. Pruning improves the model's generalization capabilities by reducing its complexity. Methods include cost-complexity pruning and reduced error pruning. Pruning might not be advisable when the dataset is small or the decision tree is already performing well with high generalization.
- b) **Regression with decision trees** can be performed by using the same decision tree algorithm to predict continuous values instead of categories. The tree splits the data into nodes with the target value being the mean or median of the values in each leaf node.
- c) **Rule induction from decision trees** converts the paths from the root to the leaf into a set of if-then rules. This approach enhances model interpretability, allowing users to understand the decision-making process more clearly.

### Question 6: Entropy and Example Calculation

- a) **Entropy** is a measure of randomness or uncertainty in information theory. In decision trees, it quantifies the impurity or unpredictability present in the dataset. It is a crucial metric used to determine the best feature that can split the dataset in each node of the tree.

b) Given the counts for the target variable *Play Tennis*, we have:

- Number of *Yes* instances: 9
- Number of *No* instances: 5
- Total number of instances:  $9 + 5 = 14$

The proportions are:

$$p_{\text{Yes}} = \frac{\text{Number of Yes instances}}{\text{Total number of instances}} = \frac{9}{14}$$
$$p_{\text{No}} = \frac{\text{Number of No instances}}{\text{Total number of instances}} = \frac{5}{14}$$

The entropy  $H(S)$  for the target variable *Play Tennis* is calculated using the formula:

$$H(S) = -p_{\text{Yes}} \log_2(p_{\text{Yes}}) - p_{\text{No}} \log_2(p_{\text{No}})$$

Plugging in the proportions, we get:

$$H(S) = -\left(\frac{9}{14}\right) \log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2\left(\frac{5}{14}\right)$$

Computing the values:

$$\begin{aligned} H(S) &\approx -\left(\frac{9}{14}\right) \log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2\left(\frac{5}{14}\right) \\ &\approx -\left(\frac{9}{14}\right) \times (-0.1520) - \left(\frac{5}{14}\right) \times (-0.4307) \\ &\approx 0.9403 \end{aligned}$$

Therefore, the entropy of the dataset for the target variable *Play Tennis* is approximately 0.9403.