📘Machine Learning Formulas by Session (with Symbol Definitions)

🕰️Session 2: Data Preprocessing and Quality

1. Interquartile Range (IQR):

$$\text{IQR} = Q_3 - Q_1$$

Where: - Q_1 : First quartile (25th percentile) - Q_3 : Third quartile (75th percentile)

2. Outlier Detection Bounds:

$$\text{Lower Bound} = Q_1 - 1.5 \times \text{IQR}$$

$$\text{Upper Bound} = Q_3 + 1.5 \times \text{IQR}$$

🕰️Session 3: Linear Regression

1. Hypothesis Function:

$$h(x) = \theta^T x = \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$

Where: - θ : Weight vector (parameters) - x : Feature vector

2. Mean Squared Error (MSE) – Cost Function:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2$$

Where: - m : Number of training examples - $h(x^{(i)})$: Predicted output for the i-th sample - $y^{(i)}$: Actual output for the i-th sample

3. Gradient Descent Update Rule:

$$\theta_j := \theta_j - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Where: - α : Learning rate - $x_j^{(i)}$: j-th feature of the i-th sample

4. Normal Equation (Closed-form Solution):

$$\theta = (X^T X)^{-1} X^T y$$

Where: - X : Feature matrix - y : Target vector

🕰️Session 4: Gradient Descent Variants

(Uses same core update as Session 3; includes Batch, Stochastic, and Mini-batch strategies.)

🕰️Session 5: Basis Functions

1. Polynomial Regression Hypothesis:

$$h(x) = w_0 + w_1 x + w_2 x^2 + \ldots + w_n x^n$$

Where: - w_i : Coefficient for x raised to the i-th power

2. General Linear Model with Basis Functions:

$$h(x) = \sum_{j=0}^{n} \theta_j \phi_j(x)$$

Where: - $\phi_j(x)$: Basis function applied to feature x

🕰️Session 6: Regularization

1. Ridge Regression (L2 Regularization):

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^{n} \theta_j^2$$

Where: - λ : Regularization strength

2. Lasso Regression (L1 Regularization):

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^{n} |\theta_j|$$

3. Elastic Net Regularization:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2 + \lambda \left[r \sum_{j=1}^{n} |\theta_j| + (1 - r) \sum_{j=1}^{n} \theta_j^2 \right]$$

Where: - r : Mixing ratio between L1 and L2 terms

🕰️Session 7: Decision Trees

1. Entropy:

$$H(S) = -p_+ \log_2(p_+) - p_- \log_2(p_-)$$

Where: - p_+ : Proportion of positive class examples - p_- : Proportion of negative class examples

2. Information Gain:

$$\text{Gain}(S, A) = H(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} H(S_v)$$

Where: - S : Entire training set - S_v : Subset of S where attribute A has value v

---

🕰️Session 8: Classification Evaluation & Logistic Regression

1. Confusion Matrix Metrics:

- Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- F1 Score:

$$F1 = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}}$$

2. ROC Curve Metrics: - True Positive Rate (TPR):

$$TPR = \frac{TP}{TP + FN}$$

- False Positive Rate (FPR):

$$FPR = \frac{FP}{FP + TN}$$

3. Logistic Regression – Sigmoid Function:

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

4. Logistic Regression – Cost Function:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right]$$

5. Logistic Regression – Gradient Descent Update Rule:

$$\theta_j := \theta_j - \alpha \cdot \left[\frac{1}{m} \sum (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} + \lambda \cdot \theta_j \right]$$

Where: - α : Learning rate - λ : Regularization parameter

(End of Formula Reference)