Important Equations in Machine Learning

February 8, 2025

1 Supervised Learning

1.1 1. Linear Regression

$$y = \beta_0 + \beta_1 x + \epsilon \tag{1}$$

- y: Dependent variable (target)
- x: Independent variable (feature)
- β_0: Intercept
- β_1: Slope coefficient
- ϵ: Error term

1.2 2. Multiple Linear Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \tag{2}$$

- y: Dependent variable (target)
- x_i: Independent variables (features)
- β_0: Intercept
- β_i: Coefficients for each feature
- ϵ: Error term

1.3 3. Logistic Regression

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \tag{3}$$

- $P(y = 1|x)$: Probability of class 1 given feature x
- β_0, β_1: Model coefficients
- e: Euler's number (≈ 2.718)

1.4   4. Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (4)$$

- n: Number of observations
- y_i: Actual value
- $\hat{y}_i$: Predicted value

1.5   5. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = \sqrt{MSE} \qquad (5)$$

- n: Number of observations
- y_i: Actual value
- $\hat{y}_i$: Predicted value

1.6   6. Gradient Descent Update Rule

$$\theta_j := \theta_j - \alpha\frac{\partial J}{\partial \theta_j} \qquad (6)$$

- θ_j: Model parameter
- α: Learning rate
- J: Cost function

2   Unsupervised Learning

2.1   7. Entropy (for Decision Trees)

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \qquad (7)$$

- $H(X)$: Entropy of dataset X
- p_i: Probability of class i

2.2   8. Gini Impurity (for Decision Trees)

$$Gini = 1 - \sum_{i=1}^{n} p_i^2 \qquad (8)$$

- p_i: Probability of class i

2.3   9. K-Means Clustering Objective Function

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \tag{9}$$

- J: Objective function (Inertia)
- k: Number of clusters
- C_i: Cluster i
- x: Data point
- μ_i: Centroid of cluster i

3   Other Important Equations

3.1   10. Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{10}$$

- $P(A|B)$: Probability of A given B (Posterior)
- $P(B|A)$: Probability of B given A (Likelihood)
- $P(A)$: Prior probability of A
- $P(B)$: Prior probability of B (Evidence)

3.2   11. Support Vector Machine (SVM) Decision Function

$$f(x) = \text{sign}(w^T x + b) \tag{11}$$

- $f(x)$: Decision function
- w: Weight vector
- x: Feature vector
- b: Bias term

3.3   12. Neural Network Weight Update (Backpropagation)

$$w := w - \alpha \frac{\partial L}{\partial w} \tag{12}$$

- w: Weight parameter
- α: Learning rate
- L: Loss function

3.4   13. TF-IDF (Term Frequency-Inverse Document Frequency)

$$TF - IDF = TF \times IDF \tag{13}$$

$$IDF = \log\left(\frac{N}{DF}\right) \tag{14}$$

- TF: Term frequency
- IDF: Inverse document frequency
- N: Total number of documents
- DF: Number of documents containing the term