Machine Learning Evaluation Methods

By Soham Banerjee

October 2025

1.0 Introduction

Model evaluation is a fundamental part of the machine learning workflow. It quantifies how well a model performs and generalizes to unseen data. Mathematical evaluation metrics allow comparison, optimization, and interpretation of performance across regression and classification models.

2.0 Regression Evaluation Metrics

2.1 Mean Absolute Error (MAE)

MAE measures the average magnitude of the prediction errors without considering their direction. It is defined as:

$$MAE = (1/n) \sum_{i=1}^{n} |y_a - \hat{y}_a|$$

2.2 Mean Squared Error (MSE)

MSE is a quadratic measure that penalizes larger errors more heavily. It is defined as:

$$MSE = (1/n) \sum_{i=1}^{n} (y_a - \hat{y}_a)^2$$

2.3 Root Mean Squared Error (RMSE)

RMSE = \u221A((1/n) \u2211_{i=1}^{n} (y_a - $\hat{y}_a$)^2)

RMSE restores the metric to the same unit as the target variable, improving interpretability.

2.4 R^2 Score (Coefficient of Determination)

$$R^2 = 1 - [\sum(y_a - \hat{y}_a)^2 / \sum(y_a - \check{y})^2]$$

R^2 evaluates how much variance in the dependent variable is predictable from the independent variables.

3.0 Classification Evaluation Metrics

3.1 Accuracy

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

Accuracy measures the proportion of correct predictions among all predictions. It can mislead when classes are imbalanced.

3.2 Precision and Recall

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

Precision quantifies correctness among predicted positives, while Recall measures coverage of actual positives.

3.3 F1 Score

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

The F1 score is the harmonic mean of Precision and Recall, balancing both metrics and useful for imbalanced data.

3.4 ROC-AUC Score

The ROC curve plots True Positive Rate (TPR) vs False Positive Rate (FPR). AUC is the area under this curve.

AUC = integral of TPR with respect to FPR over [0,1]

3.5 Confusion Matrix

The confusion matrix summarizes prediction results:

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

4.0 Cross-Validation

Cross-validation estimates how a model generalizes. In K-Fold Cross Validation, the data is divided into K subsets. The model is trained K times, each time leaving one subset for validation and averaging the results:

CV = (1/K) ∑_{k=1}^{K} M_k

This approach reduces variance and prevents overfitting, ensuring more reliable model evaluation.

Generated by Soham Banerjee's assistant. For educational and project use.