Validation Metrics

Validation metrics are used to evaluate the performance of a model on a validation dataset. These metrics help assess how well a model generalizes to unseen data. Below are common validation metrics categorized by problem type.

1. Regression Problems

• Mean Absolute Error (MAE):

Measures the average absolute difference between predicted and actual values.

Formula: **MAE = (1/n) Σ|y_i - ÿ_i|**

• Mean Squared Error (MSE):

Averages the squared differences between predicted and actual values.

Formula: **MSE = (1/n) Σ(y_i - ẏ_i)²**

• Root Mean Squared Error (RMSE):

The square root of MSE, with the same units as the target variable.

• R² Score (Coefficient of Determination):

Measures the proportion of variance explained by the model.

```
Formula:
**R<sup>2</sup> = 1 - (Σ(y_i - ẏ_i)<sup>2</sup> / Σ(y_i - ȳ)<sup>2</sup>)**
```

2. Classification Problems

• Accuracy:

Proportion of correct predictions.

```
Formula:
**Accuracy = (Correct Predictions / Total Predictions)**
```

• Precision:

Measures the proportion of true positives among predicted positives.

Formula:

Precision = TP / (TP + FP)

• Recall (Sensitivity):

Measures the proportion of true positives identified.

Formula:
Recall = TP / (TP + FN)

• F1 Score:

Harmonic mean of precision and recall.

```
Formula:
**F1 = 2 × (Precision × Recall) / (Precision + Recall)**
```

• ROC-AUC:

Measures the trade-off between true positive and false positive rates at various thresholds.

• Log Loss (Cross-Entropy Loss):

Evaluates the accuracy of predicted probabilities.

3. Clustering Problems

• Silhouette Score:

Measures how similar an object is to its cluster compared to other clusters.

• Adjusted Rand Index (ARI):

Evaluates similarity between true labels and clustering results.

• Davies-Bouldin Index:

Assesses compactness and separation of clusters.

• Inertia:

Measures how tightly grouped the clusters are.

4. Time Series Problems

• Mean Absolute Percentage Error (MAPE):

Expresses prediction error as a percentage.

Formula:

**MAPE = (100/n) $\Sigma | (y_i - \dot{y}_i) / y_i | **$

• Symmetric Mean Absolute Percentage Error (sMAPE):

Reduces bias for small values in MAPE.

• Mean Squared Logarithmic Error (MSLE):

Penalizes under- and over-predictions logarithmically.

5. Ranking Problems

• Mean Reciprocal Rank (MRR):

Evaluates ranking quality based on the reciprocal of the rank of the first relevant result.

• Normalized Discounted Cumulative Gain (NDCG):

Considers the position of relevant results in a ranked list.

• Precision at k (P@k):

Measures precision for the top-k predictions.

6. Multi-Label Problems

• Hamming Loss:

Proportion of misclassified labels.

```
Formula:
**Hamming Loss = (1/nL) ΣΣ I(y ij ≠ ÿ ij)**
```

• Subset Accuracy:

Measures the percentage of samples where all labels are correctly predicted.

• Macro/Micro Averaged Metrics:

Aggregate metrics across labels (macro) or weight by support (micro).

Summary

The choice of validation metric depends on the problem type, dataset characteristics, and business goals.

From: https://neurosurgerywiki.com/wiki/ - **Neurosurgery Wiki**

Permanent link: https://neurosurgerywiki.com/wiki/doku.php?id=validation_metrics



Last update: 2025/01/15 22:42