# Machine Learning Performance Metrics

## Regression

| Metric | Description | Pros | Cons |
|---|---|---|---|
| **Mean Absolute Error (MAE)** | Measures the average magnitude of errors between predictions and actual. | Simple to interpret; doesn’t heavily penalize large errors. | Doesn’t capture variability very well, less sensitive to outliers. |
| **Mean Squared Error (MSE)** | Measures average squared errors, penalizing large errors more than MAE. | Useful for algorithms that need error gradients (in optimization). | Sensitive to outliers, can disproportionately affect the metric. |
| **Root Mean Squared Error (RMSE)** | Square root of MSE, providing errors in same units as target variable. | Benefits of MSE but is more interpretable. | Sensitivity to outliers, as in MSE. |
| **Mean Absolute Percentage Error (MAPE)** | Expresses errors as a percentage of actual values. | Intuitive for business applications and relative comparisons. | Undefined when actual values are zero; biased towards underestimating errors. |

| | | | |
|---|---|---|---|
| **R-Squared (R2)** | Indicates the proportion of variance in the dependent variable explained by the model. | Easy to interpret, general indicator of model performance. | Can be misleading for small datasets or overfitting, doesn't penalize complexity. |
| **Adjust R-Squared** | Similar to R2, adjusts for the number of predictors, penalizing overfitting. | Accounts for model complexity; useful in feature selections. | Less intuitive to explain than R2. |
| **Akaike Information Criterion (AIC)** | Penalize model complexity while considering goodness of fit. Often used to compare separate models. | Useful for model comparison. | Not a standalone metric. |

## Classification

| Metric | Description | Pros | Cons |
|---|---|---|---|
| Accuracy | Measures the proportion of correct predictions. | Easy to understand; broad indicator of performance. | Misleading for imbalanced datasets (high accuracy with majority class decision). |
| Precision | Proportion of true positives among predicted positives. | Useful when false positives are costly. | Doesn't account for false negatives. |
| Recall (Sensitivity) | Proportion of true positives identified out of all true positives. | Useful when false negatives are costly. | Doesn't account for false positives. |
| F1 Score | Harmonic mean of precision and recall. | Balances precision and recall, especially useful for imbalanced datasets. | Doesn't reflect true negatives; not as intuitive as individual metrics. |
| Specificity | Proportion of true negatives among actual negatives. | Complements recall; useful in medical diagnostics. | Not always emphasized in standard evaluations. |
| ROC/AUC | Plots true positive rate vs false positive rate at various thresholds. | Visualizes trade-offs between sensitivity and specificity. | |

# Formulas

## Regression

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

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

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

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

$$AIC = 2k - 2\ln(L)$$ $$BIC = k\ln(n) - 2\ln(L)$$

## Classification

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

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

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

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

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

$$TruePositiveRate = \frac{TP}{TP+FN} \text{ vs. } FalsePositiveRate = \frac{FP}{FP+TN}.$$