3.2. Student Readout

Student Handout: Understanding Model Evaluation Metrics

Overview

This handout provides a concise summary of key model evaluation metrics for classification and regression models. It also includes examples to help you understand how these metrics are applied in practice.

---

1. Classification Models: Accuracy, Precision, Recall, and F1 Score

Accuracy

- **Definition:** The percentage of correct predictions made by the model out of all predictions.
- **Formula:**

[

\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \times 100

]

Examples:

1. A model predicts 90 out of 100 test samples correctly. Accuracy = 90%.
2. In a spam detection system, 950 out of 1000 emails are classified correctly. Accuracy = 95%.
3. A medical diagnosis model correctly identifies 80 out of 100 cases. Accuracy = 80%.

Precision

- **Definition:** The ratio of true positive predictions to the total positive predictions made by the model.
- **Formula:**

[

\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}

]

Examples:

1. A model predicts 50 positive cases, of which 45 are correct. Precision = 90%.
2. In a fraud detection system, 30 out of 40 flagged transactions are fraudulent. Precision = 75%.
3. A disease detection model identifies 70 out of 100 predicted cases correctly. Precision = 70%.

Recall

- **Definition:** The ratio of true positive predictions to the actual positive cases in the data.
- **Formula:**

[

\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}

]

Examples:

1. A model identifies 80 out of 100 actual positive cases. Recall = 80%.
2. In a cancer screening test, 90 out of 120 actual cases are detected. Recall = 75%.
3. A security system detects 60 out of 70 actual intrusions. Recall = 85.7%.

F1 Score

- **Definition:** The harmonic mean of precision and recall, providing a balance between the two.
- **Formula:**

[

\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}}

]

Examples:

1. A model with precision of 80% and recall of 70% has an F1 score of 74.7%.
2. In a sentiment analysis task, precision is 85% and recall is 80%. F1 score = 82.4%.
3. A model with precision of 90% and recall of 60% has an F1 score of 72%.

---

2. Regression Models: R-squared and Mean Absolute Error

R-squared (R^2)

- **Definition:** A metric indicating how well the model's predictions match the actual data, ranging from 0 to 1.
- **Formula:**

[

R^2 = 1 - \frac{\text{Sum of Squared Errors (SSE)}}{\text{Total Sum of Squares (TSS)}}

]

Examples:

1. A model with $R^2 = 0.85$ explains 85% of the variance in the data.
2. In a housing price prediction model, $R^2 = 0.92$ indicates a strong fit.
3. A sales forecasting model with $R^2 = 0.75$ captures 75% of the data variability.

Mean Absolute Error (MAE)

- **Definition:** The average difference between the predicted values and the actual values.
- **Formula:**

[

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

]

Examples:

1. A model predicts house prices with an average error of $5,000. MAE = $5,000.
2. In a temperature prediction model, the average error is 2°C. MAE = 2°C.
3. A stock price prediction model has an average error of $1.50. MAE = $1.50.

---

3. Analyzing the Power BI AutoML Validation Report

Model Performance Overview

- **Metrics:** Includes accuracy, precision, recall, F1 score (for classification models), or R-squared and MAE (for regression models).

Feature Importance

- **Definition:** Indicates which features had the most influence on the model's predictions.

Training History and Iterations

- **Definition:** Shows how the model improved over time during training.

Conclusion

Understanding these metrics allows you to evaluate and improve your machine learning models effectively. Use this handout as a reference when reviewing model validation reports and making data-driven decisions.

Questions? Feel free to reach out for further clarification or assistance.