Assignment Week 5

1. Regression Evaluation Metrics

Mean Absolute Error (MAE)

- **Definition**: Measures the average magnitude of errors in predictions, ignoring their direction.
- **When to Use**:
    - Financial forecasting where errors are interpreted in monetary units
    - Transportation arrival time predictions where early or late predictions are equally problematic
    - Customer wait time prediction for service expectation management

Mean Squared Error (MSE)

- **Definition**: Measures the average of squared differences between actual and predicted values, penalizing larger errors more heavily.
- **When to Use**:
    - Scientific experiments requiring high precision
    - Energy consumption prediction where underestimations could lead to power shortages
    - Renewable energy output forecasting where larger errors have exponentially worse effects on grid stability

Root Mean Squared Error (RMSE)

- **Definition**: The square root of MSE, providing an error metric in the same unit as the target variable.
- **When to Use**:
    - Weather forecasting where understanding impact in real-world terms is critical
    - Air pollution forecasting by government agencies predicting AQI
    - University cafeteria wait time predictions where stakeholders need interpretable units (minutes)

Mean Absolute Percentage Error (MAPE)

- **Definition**: Measures errors as a percentage of actual values.
- **When to Use**:

    - Business performance metrics where percentage-based evaluation is more interpretable
    - Retail demand forecasting for products with varying price points
    - Small business sales forecasting where a $2 error on a $5 item is more significant than on a $200 item

R-Squared (R^2)

- **Definition**: Represents the proportion of variance in the target variable explained by the model.
- **When to Use**:
    - Evaluating model fit for explanatory analysis
    - Music streaming popularity prediction based on multiple factors
    - Environmental science projects analyzing factors affecting water quality in streams

Adjusted R-Squared

- **Definition**: Adjusts R^2 for the number of independent variables, preventing overfitting.
- **When to Use**:
    - Comparing multiple models with different numbers of predictors
    - Stock market return prediction with multiple variables
    - Student performance analysis across different subjects where adding variables may not improve the model

Huber Loss

- **Definition**: A hybrid of MAE and MSE, using MSE for small errors and MAE for large errors, making it robust to outliers.
- **When to Use**:
    - Datasets with outliers where extreme values shouldn't overly influence the model
    - Medical diagnosis prediction to handle abnormal cases
    - Equipment sensor data analysis where occasional spikes can throw off calculations

Log-Cosh Loss

- **Definition**: Similar to MSE but less sensitive to large errors, behaving like MAE for big differences.
- **When to Use**:
  - When a smooth, differentiable loss function is needed for gradient-based optimization
  - Earthquake magnitude prediction to reduce sensitivity to rare extreme events

2. Confusion Matrix and Classification Metrics

Understanding the Confusion Matrix

A confusion matrix is a performance evaluation tool that provides a detailed breakdown of classification model predictions compared to actual values. It consists of four key components:

- **True Positives (TP)**: Correctly predicted positive cases
- **True Negatives (TN)**: Correctly predicted negative cases
- **False Positives (FP)**: Incorrectly predicted positive cases (Type I error)
- **False Negatives (FN)**: Incorrectly predicted negative cases (Type II error)

Why Confusion Matrices Are Useful

- They provide insight into different types of errors a model makes
- They allow calculation of precision, recall, F1 score, and other metrics
- They help optimize classification thresholds
- They reveal model performance issues that accuracy alone might hide

Example AI Model: Medical Tumor Detection

Let's consider a model analyzing medical scans for tumors, classifying images as "tumor present" (1) or "tumor absent" (0).

Test Dataset Results:

Image #	Actual	Predicted
1	1	1
2	0	0
3	1	0
4	0	0

Image #	Actual	Predicted
5	0	1
6	1	1
7	1	0
8	0	0

Resulting Confusion Matrix:

	Predicted: Tumor (1)	Predicted: No Tumor (0)
Actual: Tumor (1)	2 (TP)	2 (FN)
Actual: No Tumor (0)	1 (FP)	3 (TN)

Performance Metrics Calculation

Precision (How many predicted positives were actually positive?)

- Precision = 2 / (2 + 1) = 2/3 ≈ 0.67 or 67%

Recall (How many actual positives were correctly identified?)

- Recall = 2 / (2 + 2) = 2/4 = 0.5 or 50%

F1 Score (Harmonic mean of Precision & Recall)

- F1 = 2 × (0.67 × 0.5) / (0.67 + 0.5) ≈ 0.57 or 57%

Accuracy (Overall correct predictions)

- Accuracy = (2 + 3) / (2 + 3 + 1 + 2) = 5/8 = 0.625 or 62.5%

Specificity (True negative rate)

- Specificity = 3 / (3 + 1) = 3/4 = 0.75 or 75%

Interpretation of Results

1. **Precision (67%)**: When the model identifies a tumor, it's right about 2/3 of the time. This means roughly 1 in 3 patients would undergo unnecessary further testing.
2. **Recall (50%)**: The model only catches half of the actual tumors. This is concerning as missing tumors could delay treatment and worsen outcomes.
3. **F1 Score (57%)**: This moderate score reflects the balance between precision and recall, but isn't sufficient for clinical use.
4. **Accuracy (62.5%)**: The model gets about 5/8 cases right overall. Better than random guessing, but needs improvement.

5. **Specificity (75%)**: The model is better at ruling out tumors in healthy patients than finding tumors in sick ones.

Real-World Application Considerations

In medical contexts like cancer screening, recall might be prioritized over precision, as missing a tumor (false negative) is generally more harmful than a false alarm (false positive). This demonstrates how the application context should influence which metrics are prioritized when evaluating and optimizing a model.

For a delivery time prediction system like Timelytics, RMSE might be preferred because customer frustration increases exponentially with delay time. When predicting whether a package will be delayed, false negatives (predicting on-time delivery when actually late) are particularly damaging to customer satisfaction.

GitHub link for the code : https://github.com/JashpalsinhRana99/Ai-for-manufacturing-course-assignments/tree/main