Performance Metrics

What is Good Performance?

Sina Moghimi, Autumn 2024

Outline

- Classification Metrics
- Regression Metrics
- Segmentation Metrics
- Computational Efficiency Metrics
- Training and Optimization Metrics

Classification Metrics

- **Accuracy**
  - The ratio of correctly predicted labels to the total number of predictions

$$Accuracy = \frac{\#\ correct\ prediction}{\#\ Total\ Predictions}$$

  - Suitable for **balanced datasets** but less informative for imbalanced datasets
- **Precision**
  - Measures the proportion of true positive predictions out of all positive predictions

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)}$$

  - Important for imbalanced datasets where false positives or false negatives carry significant weight
- **Recall (sensitivity)**
  - Measures the proportion of true positives correctly identified out of all actual positives

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

  - Important for imbalanced datasets where false positives or false negatives carry significant weight

Classification Metrics

- **F1-Score**
    - The harmonic mean of precision and recall

$$F1 - Score = 2.\frac{Precision\,.\,Recall}{Precision + Recall}$$

    - Important for imbalanced datasets where false positives or false negatives carry significant weight
- **ROC-AUC (Receiver Operating Characteristic - Area Under Curve)**
    - Evaluates the trade-off between true positive rate (TPR) and false positive rate (FPR) across different thresholds
    - Useful for **binary classification** and evaluating **model discrimination capability**

Regression Metrics

- **Mean Squared Error (MSE)**
  - Measures the average squared difference between predicted and actual values

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

  - Penalizes larger errors more than smaller ones
- **Mean Absolute Error (MSE)**
  - Measures the average absolute difference between predicted and actual values

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

  - Provides a more interpretable measure compared to MSE
- R^2 **(Coefficient of Determination)**
  - Measures how well the predictions explain the variance in the actual data

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

  - Indicates the goodness of fit for regression tasks

МФТИ

Segmentation Metrics

- **Intersection over Union (IoU)**
  - Measures the overlap between predicted and ground truth regions

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

  - Common in image segmentation tasks to evaluate boundary precision
- **Dice Coefficient**
  - Measures the similarity between predicted and ground truth masks

$$Dice = \frac{2.|predictected \cap Ground\ Truth|}{|predictected| + |Ground\ Truth|}$$

  - Particularly useful for medical image analysis
- **Pixel Accuracy**
  - Proportion of correctly classified pixels

$$Pixel\ Accuracy = \frac{Correct\ Pixels}{Total\ Pixels}$$

Computational Efficiency Metrics

- **FLOPs (Floating Point Operations Per Second)**
  - Measures the computational complexity of the model by counting the number of operations
- **Inference Time**
  - Time taken for a model to process input data and produce an output
- **Model Size**
  - Measures the storage requirement for the model, typically in megabytes (MB) or gigabytes (GB)
  - Or in terms of **# of Parameters**

МФТИ

Training and Optimization Metrics

- **Loss Function**
  - Tracks the error during training
    - **Cross-Entropy Loss**: For classification tasks
    - **Mean Squared Error**: For regression tasks
- **Learning Curves**
  - Plots of training and validation loss/accuracy over epochs, used to monitor overfitting or underfitting