Evaluation Metrics 1

#ML #DLbasics

1. Confusion Matrix:

Actual \ Predicted	+ve	-ve	METRICS
+ve	TP	FN	Sensitivity= (TPR/Recall) $\frac{TP}{TP+FN}$
-ve	FP	TN	Specificity= (TNR) $\frac{TN}{TN+FP}$
METRICS	Precision= $\frac{TN}{TN+FN}$	NPV= $\frac{TN}{TN+FN}$	Accuracy= $\frac{TP + TN}{TP + TN + FP + FN}$

FNR/ Miss Rate/ Type-2 E = $\frac{FN}{TP+FN}$

FPR/ False- Alarm/ Fall-out/ Type 1 E = $\frac{FP}{FP+TN}$

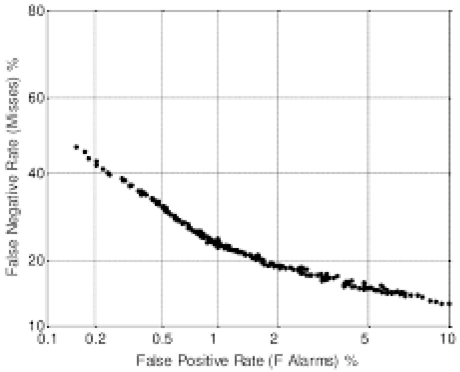
---

2. ROC Curve:

It is also called as the Receiver-Operating Characteristics and is used to evaluate a **binary classifier**. It helps you analyze the trade off between sensitivity and specificity.

- X Axis: False Positive Rate also specificity

- Y Axes: True Positive Rate also known as sensitivity or recall



---

3. DET Curve:

- Also called the detection error trade-off.
- It illustrates the trade off between Fall-Rejection i.e miss and False Acceptance i.e false Alarm.
- It uses a non linear scale on both axes, typically a probit scale.
- The goal is to have the lowest possible error rates. Therefore, the **ideal point is the bottom-left corner (0,0)** of the plot. A curve that is closer to this corner represents a better-performing model.

---

4. Performance Metrics:

1. **F1 Score:**

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN}$$

2. **Log Loss:**

$$\text{Log Loss} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} \log(\hat{y}_{ij})$$

---

5. Precision Recall Tradeoff:

The **precision-recall trade-off** refers to the inverse relationship between precision and recall in classification models, particularly when adjusting the decision threshold used to classify positive and negative instances.

- **Precision** measures the proportion of true positive predictions among all positive predictions made by the model.
- **Recall** measures the proportion of actual positive cases that the model correctly identifies.

**How the Trade-off Works:

- **Increasing precision** (fewer false positives) often leads to **decreasing recall** (more false negatives), because the model becomes more conservative in labeling positives—only the most certain cases are predicted as positive.
- **Increasing recall** (fewer false negatives) typically leads to **decreasing precision** (more false positives), as the model becomes more liberal and labels more cases as positive, including some incorrect ones15678.

Threshold Adjustment:

- Lowering the threshold increases recall but reduces precision.
- Raising the threshold increases precision but reduces recall156.

Visualization:

- The **precision-recall curve** plots precision versus recall for different threshold values, helping to visualize and select the optimal balance for a specific application2356.

Choosing the Right Balance:

- Prioritize **precision** when false positives are costly (e.g., flagging legitimate transactions as fraud).
- Prioritize **recall** when false negatives are costly (e.g., missing actual cases of disease)47.

---

6. Clustering Metrics:

1. **Adjusted Rand Index:** Compares the similarity of cluster assignments with ground truth by considering all pairs of samples and counting pairs assigned in the same or different clusters in the predicted and true clusterings, adjusted for chance.

$$\mathrm{ARI} = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}$$

2. **Silhouette Score:** The Silhouette Score measures how similar a point is to its own cluster compared to other clusters. Values range from -1 to 1.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

3. **Davies-Bouldin Index:** is a clustering evaluation metric that quantifies how well-separated and compact clusters are. Lower values indicate better clustering

$$\mathrm{DB} = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right)$$

---

7. Other Specialized Metrics:

1. **Top-K Accuracy:** Used in multi-class classification, measures if the true label is among the top K predictions.

$$\text{Top-}K \text{ Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left[y_i \in \mathrm{TopK}(\hat{y}_i)\right]$$

2. **Mathews Correlation:** Balanced measure for binary classification, even with imbalanced classes. It's a correlation coefficient between observed and predicted classifications.

$$\mathrm{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

3. **Cohen's Kappa:** Measures agreement between predicted and actual classes, adjusting for chance.

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

$$p_o = \text{observed agreement};\, p_e = \text{expected agreement}$$

$$\Rightarrow \kappa = \frac{\frac{1}{n}\sum_i n_{ii} - \frac{1}{n^2}\sum_i n_{i+}n_{+i}}{1 - \frac{1}{n^2}\sum_i n_{i+}n_{+i}}$$