Performance metrics in **Machine Learning (ML)** and **Deep Learning (DL)** are essential for evaluating model effectiveness. The right metric depends on the **problem type** (classification, regression, etc.) and **business goals** (e.g., minimize false positives, maximize recall). Here's a comprehensive breakdown:

**🔹 1. Classification Metrics (Binary & Multiclass)**

| **Metric** | **Description** | **Best For** |
| --- | --- | --- |
| **Accuracy** | Correct predictions / total predictions | Balanced classes |
| **Precision** | TP / (TP + FP) – how many predicted positives are correct | Minimize false positives |
| **Recall (Sensitivity)** | TP / (TP + FN) – how many actual positives are captured | Minimize false negatives |
| **F1-Score** | Harmonic mean of precision & recall | Imbalanced data |
| **Specificity** | TN / (TN + FP) – true negative rate | Medical or anomaly detection |
| **AUC-ROC** | Area under Receiver Operating Characteristics curve | Overall separability |
| **PR-AUC** | Area under Precision-Recall curve | Imbalanced classes |
| **Log Loss (Cross Entropy)** | Penalizes confident wrong predictions | Probabilistic models |

💡 In DL, the same classification metrics apply but are computed on output probabilities from models like neural nets (e.g., softmax/sigmoid outputs).

**🔹 2. Regression Metrics**

| **Metric** | **Description** | **Notes** |
| --- | --- | --- |
| **Mean Absolute Error (MAE)** | Avg. of absolute errors | Interpretable, less sensitive to outliers |
| **Mean Squared Error (MSE)** | Avg. of squared errors | Penalizes larger errors |
| **Root Mean Squared Error (RMSE)** | sqrt(MSE) | In original unit scale |
| **R-squared (R²)** | Proportion of variance explained | 1 = perfect, 0 = no correlation |
| **Adjusted R²** | Adjusted for number of predictors | Good for comparing models |
| **Mean Absolute Percentage Error (MAPE)** | % error | Avoid if target values are near zero |

**🔹 3. Clustering Metrics (Unsupervised Learning)**

| **Metric** | **Description** |
| --- | --- |
| **Silhouette Score** | How well points fit within clusters |
| **Davies–Bouldin Index** | Lower = better cluster separation |
| **Adjusted Rand Index** | Similarity between clusterings |
| **Normalized Mutual Information (NMI)** | Measures agreement with ground truth |

**🔹 4. Other Useful Metrics**

* **Confusion Matrix**: Summary of TP, TN, FP, FN
* **Top-k Accuracy** (e.g., Top-5 Accuracy): Used in DL, especially image classification (e.g., ImageNet)
* **Hamming Loss**: For multi-label classification
* **Cohen’s Kappa**: Agreement between predictions and labels

**🔹 5. Task-Specific DL Metrics**

| **Task** | **Example Metrics** |
| --- | --- |
| **Object Detection** | mAP (mean Average Precision), IoU (Intersection over Union) |
| **Segmentation** | Dice Coefficient, IoU |
| **NLP (e.g., text classification)** | Accuracy, F1, BLEU(bilingual evaluation understud) (for translation), ROUGE(Recall-Oriented Understudy for Gisting Evaluation) (for summarization) |
| **Recommender Systems** | Precision@k, Recall@k, NDCG(Normalized Discounted Cumulative Gain), MAP((mean average precision) |