

METRICS:

MACHINE LEARNING MODELS:

1. **Explain the difference between precision and recall. When would you prefer one over the other?**

Answer: Precision is the ratio of true positives to the sum of true positives and false positives, measuring the accuracy of positive predictions. Recall is the ratio of true positives to the sum of true positives and false negatives, assessing the model's ability to capture all relevant instances. Precision is favoured in scenarios where false positives are costly, while recall is crucial when missing positive instances is more problematic.

2. **What is the F1 score, and why might it be preferred over accuracy in certain situations?**

The F1 score is the harmonic mean of precision and recall. It is preferred over accuracy when dealing with imbalanced datasets, where one class significantly outweighs the other. The F1 score provides a balanced measure, particularly when the cost of false positives and false negatives is asymmetric.

3. **What is the area under the ROC curve (AUC-ROC), and how does it relate to the performance of a binary classification model?**

AUC-ROC represents the area under the receiver operating characteristic (ROC) curve, which visualizes the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity). A higher AUC-ROC indicates better discrimination ability of the model. It is commonly used in binary classification tasks to assess the model's ability to distinguish between positive and negative instances across different probability thresholds.

DEEP LEARNING METRICS:

1. **Explain the concept of mean squared error (MSE) and when it is appropriate to use in deep learning applications.**

MSE measures the average squared difference between predicted and true values. It is often used as a regression metric in deep learning when the goal is to minimize the average squared error between predicted and actual continuous values. MSE is sensitive to outliers and penalizes larger errors more significantly.

2. **What is the role of categorical cross-entropy in deep learning, particularly in classification tasks?**

Categorical cross-entropy is a loss function used in multiclass classification tasks. It quantifies the difference between predicted probability distributions and true class distributions. Minimizing categorical cross-entropy during training results in the model learning to assign high probabilities to the correct class, making it a common choice for classification objectives.

3. **In what scenarios would you choose categorical cross-entropy as a loss function over mean squared error in deep learning?**

Categorical cross-entropy is typically used as a loss function for multiclass classification tasks, where the target is a probability distribution over multiple classes. **Mean squared error is more suitable for regression tasks.** Categorical cross-entropy penalizes deviations from the true class distributions, making it effective for classification objectives.

4. **What is binary cross-entropy, and when is it used as a loss function in deep learning?**

Binary cross-entropy is a loss function commonly used for binary classification tasks. It measures the dissimilarity between predicted probabilities and true binary labels. Minimizing binary cross-entropy encourages the model to produce probability distributions that align with the ground truth for binary classification problems.

5. Explain the concept of the triplet loss function. In what types of tasks is it commonly used, and why?

The triplet loss is used in tasks involving similarity or dissimilarity comparisons, such as face verification or image retrieval. It involves comparing the distance between an anchor sample, a positive sample (similar), and a negative sample (dissimilar). The loss encourages the model to minimize the distance between the anchor and positive samples while maximizing the distance between the anchor and negative samples.

6. When might you choose to use the Huber loss function over mean squared error in deep learning applications?

Huber loss is less sensitive to outliers compared to mean squared error. It is often preferred in regression tasks where there may be noisy data or outliers. Huber loss provides a balance between the robustness of mean absolute error and the smoothness of mean squared error, making it suitable for scenarios where a compromise is needed.

7. Explain the use of the Kullback-Leibler (KL) divergence as a loss function in deep learning. In what types of tasks is it commonly employed?

KL divergence measures the difference between two probability distributions. In deep learning, it is often used in variational autoencoders (VAEs) as part of the loss function to ensure that the learned latent space distribution closely matches a predefined distribution, typically a Gaussian. KL divergence helps regularize the latent space and encourages it to follow the desired distribution.

8. How does the choice of loss function impact the training process and the performance of a deep learning model?

The choice of loss function influences the learning objectives and characteristics of the trained model. Different loss functions prioritize different aspects of the learning process, such as accuracy, robustness to outliers, or adherence to a predefined distribution. The selection should align with the specific goals and nature of the task at hand.

GENERATIVE AI METRICS:

1. How do you evaluate the performance of a generative model, such as a GAN, in generating realistic images?

Evaluation of generative models is challenging. Metrics like Inception Score (IS) and Frechet Inception Distance (FID) are commonly used. IS measures the quality and diversity of generated images, while FID compares the statistics of generated and real images. However, no single metric is perfect, and visual inspection is often essential.

2. What challenges are associated with evaluating the performance of natural language generation models, and how can you address them?

Evaluating language generation is subjective. Metrics like BLEU, ROUGE, and METEOR assess the quality of generated text against reference text. Human evaluation is valuable for assessing aspects like coherence and fluency. Combining multiple metrics and incorporating user feedback provides a more comprehensive evaluation.

3. In what ways can you measure the diversity and novelty of generated samples in a generative model?

Diversity and novelty metrics assess the variety of generated samples. Metrics like Unique-N and Intra-Diversity measure uniqueness within generated samples. Jensen-Shannon Divergence and Kullback-Leibler Divergence assess the diversity between the generated and real data distributions.