

Quiz & Solutions

Multiple-choice questions. The correct answer(s) are indicated.

1. What is data augmentation?

- A) Creating modified copies of existing data
- B) A technique to reduce overfitting ☒
- C) A method that causes overfitting
- D) A technique to generate synthetic training data ☒
- E) None of the above

Explanation:

Data augmentation consists of creating modified or synthetic versions of existing data to enrich the training set. This helps reduce overfitting by exposing the model to varied examples and improving its generalization capabilities.

2. Why are data important for advanced machine learning applications?

- A) It improves model performance
- B) It directly affects model performance
- C) It is essential for hyperparameter tuning
- D) It reduces computational overhead
- E) It is required for both training and evaluating models ☒

Explanation:

Data is fundamental in machine learning: it is used to train models and also to evaluate their effectiveness. Without enough representative and high-quality data, even the best algorithms perform poorly.

3. Why use data augmentation?

- A) To reduce dataset size
- B) To lower data diversity
- C) To increase computational cost
- D) To improve model performance while minimizing the need for new data collection ☒
- E) To intentionally reduce accuracy

Explanation:

Data augmentation improves model performance by increasing data variability without the need to collect new samples, which can be expensive or impractical. This helps in achieving better generalization.

4. What are the main purposes of data augmentation?

- A) Increasing dataset size ☒
- B) Addressing class imbalance ☒
- C) Reducing model complexity
- D) Both A and C
- E) None of the above

Explanation:

Data augmentation increases the amount of training data and helps balance underrepresented classes, which improves the learning process, especially in classification tasks with skewed distributions.

5. What is a typical data augmentation workflow?

- A) Evaluate model performance → choose augmentation technique → apply it → re-evaluate ✓
- B) Measure performance → augment data → shuffle dataset
- C) Measure performance → select technique → augment data
- D) Augment data → evaluate performance → select technique
- E) Shuffle data → measure performance → augment data

Explanation:

A good workflow starts by identifying weaknesses in the current model, choosing appropriate augmentation techniques, applying them, and then measuring the impact on performance.

6. At which stage is data augmentation usually applied?

- A) Validation
- B) Test
- C) Data preprocessing
- D) Training ✓
- E) Hyperparameter optimization

Explanation:

Data augmentation is applied during training to improve model generalization. It is not used during validation or test phases, as it would bias performance metrics.

7. What is a major benefit of data augmentation?

- A) Increased computational cost
- B) Higher risk of noise
- C) Improved model generalization ✓
- D) Limited applicability
- E) Improved fairness

Explanation:

The main advantage of data augmentation is better generalization. By training the model on a wider variety of data, it learns more robust patterns and performs better on unseen data.

8. What is a common drawback of data augmentation?

- A) Improved generalization
- B) Cost-efficiency
- C) Potential to introduce noise or artifacts ✓
- D) Better fairness
- E) Prevention of overfitting

Explanation:

If augmentation techniques are poorly chosen or applied excessively, they can generate unrealistic or noisy data, which may confuse the model rather than help it.

9. What is a risk associated with data augmentation?

- A) Unintended bias or fairness issues
- B) Increased computational requirements
- C) Injection of noisy or unrealistic data
- D) All of the above ✓
- E) None of the above

Explanation:

While beneficial, data augmentation can introduce risks such as computational overhead, unfair model behavior due to biased transformations, or noise that degrades learning.

10. What is a key ethical concern in data augmentation?

- A) Increased hardware demands
- B) Risk of noise
- C) Lack of impact in some domains
- D) Potential bias and fairness issues ✓
- E) Enhanced generalization

Explanation:

Certain augmentation methods may unintentionally favor or disadvantage specific groups in the data, raising fairness and bias concerns—especially in sensitive applications like healthcare or recruitment.