MACHINE LEARNING KiTABI

1. What is the purpose of a development (dev) set in machine learning?
A) To train the model
B) To evaluate the final model performance
C) To tune hyperparameters and select models
D) To debug the test set
Answer: C

2. Why is it important for dev and test sets to come from the same distribution?
A) It improves training speed
B) It simplifies error analysis
C) It ensures the team works toward a relevant target
D) It reduces the need for data augmentation
Answer: C

3. What does "single-number evaluation metric" mean in a machine learning project?
A) Using only one metric like accuracy or precision
B) Combining multiple metrics into one formula
C) Evaluating only the test set
D) Measuring dataset size with a single number
Answer: B

4. Which of the following is NOT an example of supervised learning?
A) Image classification
B) Spam detection
C) Clustering customer segments
D) Speech recognition
Answer: C

5. What is error analysis used for in machine learning projects?
A) To train the algorithm faster
B) To identify major error categories for prioritization
C) To generate new test set examples
D) To optimize hyperparameters
Answer: B

6. You are working on a dog classifier, and 10% of errors are due to blurry images. Which action should you prioritize?
A) Improve the quality of blurry data
B) Train the model with more diverse dog breeds

C) Increase the size of the test set
D) Add regularization to the model
Answer: A

7. During error analysis, you find mislabeled examples in your dev set. When should you correct them?
A) When mislabeled data significantly affects decision-making
B) When the dev set is small
C) Always correct mislabeled examples immediately
D) Only if the test set is also mislabeled
Answer: A

8. Which scenario indicates high variance?
A) Training error: 5%, Dev error: 20%
B) Training error: 20%, Dev error: 21%
C) Training error: 15%, Dev error: 15%
D) Training error: 1%, Dev error: 2%
Answer: A

9. How can you reduce bias in a machine learning model?
A) Add regularization
B) Increase the training data size
C) Increase the model complexity (e.g., larger neural network)
D) Use cross-validation
Answer: C

10. Which method is most effective for reducing high variance?
A) Adding more training data
B) Reducing the model size
C) Decreasing the dev set size
D) Using a smaller test set
Answer: A

11. What is the main drawback of using multiple evaluation metrics simultaneously?
A) Reduces the test set size
B) Makes it harder to compare models
C) Increases training complexity
D) Leads to overfitting
Answer: B

12. Which metric combines precision and recall into a single number?
A) Accuracy
B) Mean Squared Error
C) F1 Score
D) AUC-ROC
Answer: C

13. What should you do if your dev set distribution is very different from your test set distribution?
A) Use the dev set to tune hyperparameters
B) Match the test set distribution during data collection
C) Collect more training data
D) Increase regularization
Answer: B

14. Which approach is recommended for large datasets in the big data area?
A) 70% training, 30% test split
B) Smaller dev and test sets but larger absolute sizes
C) Equal-sized dev, test, and training sets
D) Only use a dev set and skip the test set
Answer: B

15. A model has the following results: Training error: 10%, Dev error: 25%. What should you do next?
A) Increase model size
B) Add more training data
C) Add regularization
D) Use early stopping
Answer: B

16. Which learning curve indicates a high bias problem?
A) Both training and dev errors are high and close
B) Training error is low, dev error is high
C) Training error decreases with more data
D) Dev error remains constant as training size increases
Answer: A

17. What is a potential drawback of end-to-end learning?
A) It simplifies the pipeline
B) Requires more labeled data
C) Reduces variance
D) Improves interpretability
Answer: B

18. End-to-end learning is most suitable for tasks where:
A) Intermediate representations are easy to define
B) Feature engineering is complex
C) Data is limited
D) Models are interpretable
Answer: B

19. A speech recognition model has: Training error: 15%, Dev error: 30%. Optimal human error rate: 14%. What is the primary issue?
A) High bias
B) High variance
C) Low data quality
D) Overfitting
Answer: B

20. Your dev set contains mislabeled examples that make up 30% of the total errors. What action should you take?
A) Ignore the mislabeled data
B) Correct the mislabeled examples
C) Collect more data without correcting the current set
D) Adjust the model to handle mislabeled data
Answer: B

21. How does regularization affect a model?
A) Reduces bias, increases variance
B) Increases bias, reduces variance
C) Increases both bias and variance
D) Reduces both bias and variance
Answer: B

22. Which method should you prioritize when computational power is limited, and you need to reduce training time?
A) Use a smaller model
B) Collect more training data
C) Increase the number of epochs
D) Use dropout
Answer: A

23. Why is comparing to human-level performance useful?
A) Helps identify avoidable bias
B) Ensures better generalization
C) Reduces the need for evaluation metrics
D) Eliminates variance issues
Answer: A

24. What is the optimal error rate also known as?
A) Training error
B) Test error
C) Bayes error rate
D) Variance error
Answer: C

25. What is the main purpose of a dev set in machine learning?
a) Training the model
b) Tuning hyperparameters
c) Evaluating the final model
d) Storing unused data
Answer: b

26. Which factor drives machine learning progress significantly?
a) Manual feature engineering
b) Regularization techniques
c) Scale of data and computation
d) Optimizing gradient descent
Answer: c

27. What does overfitting imply in machine learning?
a) High bias
b) High variance
c) Low variance
d) Low bias
Answer: b

28. Which metric is an example of a single-number evaluation?
a) Precision and Recall
b) Accuracy
c) Confusion Matrix
d) ROC Curve
Answer: b

29. What is the goal of regularization in machine learning?
a) Increase model complexity
b) Reduce variance
c) Reduce bias
d) Increase training error
Answer: b

30. Why is a single-number evaluation metric preferred?
a) It simplifies model comparison.
b) It increases model accuracy.
c) It improves computation speed.
d) It reduces data requirements.
Answer: a

31. What is a typical sign of high bias in a model?
a) High training error
b) High dev set error
c) Large gap between training and dev set error
d) Low training error
Answer: a

32. Which of these helps address high bias in a model?
a) Adding more data
b) Increasing model size
c) Adding regularization
d) Decreasing model size
Answer: b

33. What does the term "avoidable bias" refer to?
a) The difference between training error and optimal error
b) The error in the dev set
c) The variance in predictions
d) Noise in the data
Answer: a

34. When should dev and test sets come from the same distribution?
a) Always
b) When the training set is large
c) When optimizing for generalization
d) When building a specific application
Answer: c

35. What is Artificial Intelligence primarily focused on?
A) Developing hardware for computing
B) Designing user-friendly interfaces
C) Creating systems that can simulate human intelligence
D) Building faster computer networks
Answer: C

36. Which of the following is NOT an application of AI?
A) Self-driving cars
B) Facial recognition
C) Web development frameworks
D) Chatbots like Siri or Alexa
Answer: C

37. What is the main feature of Narrow AI?
A) It is a type of AI that can perform a single task effectively.
B) It has intelligence comparable to that of humans.
C) It learns without any human supervision.
D) It creates new AI models.
Answer: A

38. Which algorithm is best suited for identifying patterns in unlabelled datasets?
A) Supervised Learning
B) Unsupervised Learning
C) Reinforcement Learning
D) Regression
Answer: B

39. What is an outlier in data analysis?
A) A data point that fits the expected trend.
B) A data point significantly different from others.
C) A missing data point in the dataset.
D) A synthetically generated data point.
Answer: B

40. What does the sigmoid function output range from?
A) 0 to 1
B) 0 to 100
C) -1 to 1
D) 0 to infinity
Answer: A

41. Which of the following describes the purpose of the Turing Test?
A) To evaluate the efficiency of a machine learning algorithm
B) To determine if a machine can exhibit intelligent behavior equivalent to a human
C) To classify AI into General and Narrow categories
D) To test the computational speed of a machine
Answer: B

42. What type of data is used in supervised learning?
A) Completely unlabelled data
B) A mix of labelled and unlabelled data
C) Data with input-output pairs (labelled data)
D) Data generated by reinforcement methods
Answer: C

43. What is one main advantage of AI in medical research?
A) AI reduces the cost of surgeries.
B) AI accelerates drug discovery and analysis.
C) AI eliminates the need for human doctors.
D) AI automatically cures diseases.
Answer: B

44. What problem can arise from data bias in AI systems?
A) Increased training time for models
B) Difficulty in labeling datasets
C) AI systems making inaccurate or unfair decisions
D) Failure of the Turing Test
Answer: C

44. What is General AI?
A) AI designed to excel at a single task
B) AI that can transfer knowledge across domains and perform any intellectual task like a human
C) AI used specifically for natural language processing
D) AI capable of executing only pre-programmed tasks
Answer: B

45. Which of the following is an example of reinforcement learning?
A) A chatbot responding to customer queries
B) A robot learning to navigate a maze through trial and error
C) An algorithm identifying objects in images
D) A statistical regression model predicting house prices
Answer: B

46. What is the primary goal of feature engineering in machine learning?
A) To clean the data by removing outliers
B) To create new attributes from existing data for better model performance
C) To reduce the size of the dataset
D) To generate synthetic datasets
Answer: B

47. What is the primary difference between supervised and unsupervised learning?
A) Supervised learning works with labelled data, while unsupervised learning works with unlabelled data.
B) Supervised learning is faster than unsupervised learning.
C) Supervised learning only works with numeric data, while unsupervised learning handles textual data.
D) Supervised learning requires more computational power than unsupervised learning.
Answer: A

48. Which algorithm is commonly used in recommendation systems?
A) Naive Bayes
B) Apriori
C) K-Means Clustering
D) Reinforcement Learning
Answer: B

49. What type of AI application is a recommendation system categorized under?
A) Super AI
B) Narrow AI
C) General AI
D) Autonomous AI
Answer: B

50. Which of the following describes the concept of a "hyperplane"?
A) A line separating data in 1D space
B) A multidimensional boundary used in classification tasks
C) A feature extraction method
D) A preprocessing step in machine learning
Answer: B

51. What kind of algorithm would you use for clustering customer data without labels?
A) Logistic Regression
B) Supervised Learning
C) K-Means Clustering
D) Reinforcement Learning
Answer: C

52. What does the correlation coefficient of 0.75 indicate?
A) A weak positive relationship between variables
B) A strong positive relationship between variables
C) A weak negative relationship between variables
D) No relationship between variables
Answer: B

53. What is the purpose of a confusion matrix in machine learning?
A) To evaluate the accuracy of regression models
B) To display the performance of a classification model
C) To generate synthetic data for testing
D) To summarize the correlation between variables
Answer: B

54. What is a primary characteristic of "Narrow AI"?
A) It is capable of thinking like humans.
B) It specializes in specific tasks or domains.
C) It can independently learn across multiple domains.
D) It can replace human intelligence in all fields.
Answer: B

55. Which AI field involves the study of how agents can act in an environment to maximize cumulative reward?
A) Supervised Learning
B) Reinforcement Learning
C) Unsupervised Learning
D) Natural Language Processing
Answer: B

56. What is the main purpose of synthetic data in AI?
A) To replace missing values in datasets
B) To artificially expand datasets for model training
C) To test algorithms in real-world environments
D) To identify outliers in data
Answer: B

57. Which of the following is true about the Turing Test?
A) It is used to measure the speed of AI algorithms.
B) It determines if a machine can simulate human-like intelligence.
C) It assesses the level of automation in manufacturing processes.
D) It tests the accuracy of classification models.
Answer: B

58. Which method is used to handle missing data in a dataset?
A) Feature engineering
B) Data anonymization
C) Synthetic data generation
D) Model deployment
Answer: C

59. Which statement is correct about outliers in a dataset?
A) Outliers are always removed before model training.
B) Outliers should be included in training but not in testing.
C) Outliers depend on the nature of the problem and dataset.
D) Outliers are irrelevant to machine learning models.
Answer: C

60. What is a key challenge in AI ethics?
A) Lack of computing resources
B) Data bias leading to unfair decisions
C) Inability to train deep learning models
D) Over-reliance on supervised learning
Answer: B

61. What is the significance of explainability in AI systems?
A) It improves computational efficiency.
B) It ensures that AI decisions are understandable and transparent.
C) It eliminates the need for feature engineering.
D) It enhances the speed of data processing.
Answer: B

62. What is a primary limitation of reinforcement learning?
A) It requires labeled data for training.
B) It is computationally intensive and may require a large number of iterations.
C) It cannot handle continuous data.
D) It is limited to supervised learning problems.
Answer: B

63. Which AI approach is commonly used in speech recognition systems?
A) Reinforcement learning
B) Neural networks
C) Clustering
D) Regression
Answer: B

64. In which case can we say the model is underfitting?
a) High training error and low development error
b) Low training error and high development error
c) Both training and development errors are high
d) Development error is zero
Answer: C

65. In the bias-variance tradeoff, what does high bias typically indicate?
a) The model is too complex
b) The model is underfitting the data
c) The model needs more data
d) The model has overfitted to the development set
Answer: B

66. What is an effective way to reduce variance?
a) Increasing the model size
b) Adding more data
c) Removing regularization
d) Adding more labels to the test set
Answer: B

67. During error analysis, which situation is usually prioritized?
a) The most common error categories
b) The least frequent error categories
c) High-accuracy categories
d) Categories unrelated to model performance
Answer: A

68. If 15% of errors in the dev set belong to the "low light" category, what does this suggest?
a) Build a larger model
b) Focus on improvements specific to low-light conditions
c) Expand the training dataset
d) Change the development set
Answer: B

69. Why should development and test sets come from the same distribution?
a) To prevent the model from becoming too complex
b) To increase the size of the training data
c) To ensure consistent evaluation results
d) To simplify data collection
Answer: C

70. When is regularization most helpful?
a) When both training and development errors are low
b) When the model shows high variance
c) When the learning curve flattens
d) When the model has high bias
Answer: B

71. Why is using a single metric beneficial?
a) It increases the model's complexity
b) It accelerates the decision-making process
c) It creates more error categories
d) It improves test set performance
Answer: B

72. What is a possible side effect of increasing model size?
a) Higher training error
b) Reduced variance
c) Overfitting if no regularization is used
d) A less structured learning curve
Answer: C

73. Why are errors below human-level performance significant?
a) To identify bias and variance
b) To show the model aligns with real-world data
c) To reduce the size of the training dataset
d) To simplify the creation of new metrics
Answer: A

74. How is the F1 score calculated?
a) The sum of precision and recall
b) The arithmetic mean of precision and recall
c) The harmonic mean of precision and recall
d) The product of precision and accuracy
Answer: C

75. If training error is low but test error is high, what is it called?
a) The model is overfitting
b) The model is underfitting
c) There is too much training data
d) The test set is too small
Answer: A

76. What should you do if mislabeled examples exist in the dev set?
a) Ignore the development set errors
b) Correct the labeling errors
c) Increase model complexity
d) Add more data to the test set
Answer: B

77. How should the size of the dev set be determined?
a) It should always be 30% of the dataset
b) It should be independent of the training and test sets
c) It should be large enough to detect algorithm differences
d) It should be small to reduce variance
Answer: C

78. If training error is lower than the optimal error rate, what is suggested?
a) Add more data
b) Focus on reducing variance
c) Increase model complexity
d) Reduce the training dataset
Answer: B

Theoretical Questions

1. **Explain the importance of dev and test sets coming from the same distribution.**
Dev (development) and test sets must come from the same distribution to ensure the model's performance on the dev set generalizes to the test set. If the distributions differ, a model optimized for the dev set might not perform well on the test set, leading to inaccurate performance estimates.

2. **Describe the difference between bias and variance in a machine learning model.**
   - **Bias:** Refers to the error introduced by simplifying assumptions in the model. High bias indicates underfitting.
   - **Variance:** Refers to the error introduced by the model's sensitivity to small fluctuations in the training data. High variance indicates overfitting.
   - **Tradeoff:** A model must balance bias and variance to achieve optimal performance.

3. **Why is it important to have a single-number evaluation metric? Provide an example.**
A single-number evaluation metric simplifies comparison and guides decision-making. For instance, using F1-score in classification tasks balances precision and recall, allowing clear evaluation and tuning.

4. **What is error analysis, and how can it guide improvements in machine learning models?**
Error analysis involves manually examining model errors to identify patterns or weaknesses. This process highlights areas needing improvement, such as adding specific features or addressing biases in the data.

5. **Discuss how overfitting can be identified and mitigated.**
   - **Identification:** High training accuracy but low test accuracy indicates overfitting.
   - **Mitigation:** Use techniques like regularization (e.g., L2, dropout), adding more data, or simplifying the model.

6. **Explain the impact of using large neural networks on bias and variance.**
Large networks can reduce bias by better capturing data patterns but are prone to high variance if not regularized properly. Techniques like dropout and weight decay help manage this.

7. **How does human-level performance help define the optimal error rate?**
Human-level performance serves as a benchmark for achievable error. When a model reaches this level, further improvements become challenging and guide the focus toward unavoidable noise in the data.

8. **What is the role of regularization in reducing variance? Provide examples.**
Regularization prevents overfitting by penalizing large weights. Examples:
   - **L1 Regularization:** Adds sparsity by forcing some weights to zero.
   - **L2 Regularization:** Shrinks weights smoothly.

9. **Describe a scenario where adding more training data would not significantly improve model performance.**
Adding more training data won't help if the model has **high bias** (underfitting). For example, a linear regression model trying to predict housing prices in a non-linear scenario will fail because it lacks the complexity to capture patterns. Even with more data, the model can't improve because the issue lies in its simplicity, not data scarcity. The solution is to use a more complex model or better features.

10. **Why might you split a dev set into an Eyeball dev set and a Blackbox dev set?**
An Eyeball dev set is for manual inspection, while a Blackbox dev set is for automated evaluation. This split ensures errors can be analyzed manually while also tracking performance on unseen data.

11. **Explain why dev and test sets must be chosen to represent future data distributions. How does this affect model performance?**
The dev/test sets must resemble future data distributions to ensure the model's real-world applicability. If distributions mismatch, the model may fail to generalize.

12. Describe the process and benefits of conducting error analysis on a dev set. Why is it important to categorize errors?

Process of Error Analysis:
1. Review a sample of dev set errors manually.
2. Identify patterns or common causes of errors (e.g., specific classes, edge cases).
3. Categorize errors into groups (e.g., misclassification of certain categories or poor handling of noisy data).

Benefits:
- Pinpoints weaknesses in the model.
- Guides targeted improvements, such as better data augmentation or feature engineering.
- Prioritizes efforts where fixes will have the most significant impact.

Importance of Categorizing Errors:
Categorizing errors helps identify specific areas to focus on, enabling more efficient and effective model optimization. For example, addressing issues in underrepresented classes can significantly boost overall performance.

13. What is the bias-variance tradeoff? How can it guide decision-making in machine learning?

The **bias-variance tradeoff** is the balance between two types of errors:
- **Bias:** Error from overly simple models that underfit the data.
- **Variance:** Error from overly complex models that overfit the data.

Guiding Decision-Making:
- If **bias is high** (poor training performance): Use a more complex model or add features.
- If **variance is high** (poor test performance): Simplify the model, use regularization, or add more data.

The goal is to find a model that minimizes total error (bias + variance) for optimal performance.

14. Discuss the advantages and challenges of using end-to-end learning in real-world applications.

- **Advantages:** Eliminates the need for manual feature engineering.
- **Challenges:** Requires large labeled datasets and careful tuning to prevent overfitting.

15. How does the use of a single-number evaluation metric streamline machine learning projects? Give examples.

A **single-number evaluation metric** simplifies model comparison and decision-making by providing a clear, objective measure of performance.

How it Streamlines Projects:
- **Consistency:** Ensures all teams optimize toward the same goal.
- **Efficiency:** Makes model evaluation quick and straightforward.
- **Prioritization:** Helps rank models and select the best one.

Examples:
- **F1-Score** in imbalanced classification balances precision and recall.
- **AUC-ROC** summarizes performance across all classification thresholds.
- **Mean Absolute Error (MAE)** in regression gives a single value to compare error magnitudes.

By focusing on a unified metric, teams can avoid ambiguity and align efforts effectively.

16. Why is it essential to compare a model's performance to human-level performance, especially in tasks like image recognition?

Comparing a model's performance to **human-level performance** is essential because:
1. **Benchmarking:** It sets a realistic goal for what's achievable.
2. **Error Analysis:** If the model performs worse than humans, it highlights areas for improvement.
3. **Understanding Limits:** When the model reaches human-level performance, further improvements may be limited by data noise or labeling inconsistencies.

For example, in **image recognition**, human-level performance helps determine if errors are due to model shortcomings or inherent challenges in the task (e.g., ambiguous images).

17. Provide a step-by-step process for splitting large dev sets into "Eyeball" and "Blackbox" subsets. What are their respective purposes?

Step-by-Step Process:
1. Randomly Split the Dev Set:
→ Select a small portion (e.g., 10-20%) of the dev set as the Eyeball subset.
→ Assign the rest to the Blackbox subset.

2.<u>Ensure Representativeness</u>:
→ Both subsets should represent the same data distribution as the full dev set.

3.<u>Assign Tasks</u>:
→ Use the Eyeball subset for manual inspection and error analysis.
→ Use the Blackbox subset for automated performance tracking.
<u>Purposes</u>:
→ Eyeball Subset: Helps identify patterns and weaknesses by allowing manual examination of model errors.
→ Blackbox Subset: Measures performance consistently over time without manual intervention.

This division helps refine the model efficiently while ensuring robust evaluation.

18. **Discuss how increasing model size and regularization simultaneously can reduce avoidable bias without causing overfitting.**
Increasing model size can reduce **avoidable bias** by allowing the model to learn more complex patterns in the data. However, larger models are prone to overfitting due to high capacity.
To prevent overfitting, **regularization** techniques (e.g., L2 regularization, dropout) are applied simultaneously. Regularization limits the model's ability to fit noise in the training data, ensuring it generalizes well.
This combination strikes a balance:
- Larger model → Reduces bias.
- Regularization → Controls variance.
This approach improves performance without sacrificing generalization.

19. **Explain the importance of learning curves in diagnosing bias and variance problems. Include a scenario as an example.**
Learning curves plot training and validation errors against the amount of training data. They help diagnose:
1. **High Bias**: Both training and validation errors are high and close, even with more data.
   - Example: A simple linear model underfitting a non-linear dataset.
2. **High Variance**: Training error is low, but validation error is high and decreases with more data.

   - Example: A deep neural network overfitting a small dataset.

Importance: Learning curves provide insights into whether to:
- Increase model complexity (for high bias).
- Use regularization or more data (for high variance).

20. **Why might it be necessary to adjust or completely change the dev/test sets or metrics during a project? Give practical examples.**
Adjusting or changing dev/test sets or metrics is necessary when they no longer align with the **project's goals** or **real-world conditions**.

Reasons:
1. **Data Distribution Shift**: If real-world data differs from the dev/test set.
   - Example: A speech recognition system trained on clear audio but used in noisy environments.
2. **Misaligned Metrics**: When current metrics fail to reflect business needs.
   - Example: Switching from accuracy to F1-score in an imbalanced classification task.
3. **New Priorities**: Project focus shifts, requiring updated evaluation criteria.
   - Example: Prioritizing inference speed over accuracy in a real-time application.
Adjustments ensure the model remains relevant and effective for its intended use.

21. **This comprehensive set should cover both practical and theoretical aspects of Machine Learning Yearning and test a variety of skills.**
This comprehensive set of questions integrates **practical skills** (e.g., error analysis, model tuning) and **theoretical concepts** (e.g., bias-variance tradeoff, regularization) from *Machine Learning Yearning*. It ensures:
1. **Foundational Knowledge**: Tests understanding of core principles.
2. **Problem-Solving Skills**: Encourages practical applications, like splitting dev sets or diagnosing errors.
3. **Critical Thinking**: Assesses the ability to evaluate and improve models based on real-world scenarios.
By blending theory with application, it prepares learners for effectively designing and optimizing machine learning systems.