Unit-3: Evaluating Models

Section A: Basics of Model Evaluation

1. Model evaluation in AI is similar to:
   a) Writing the algorithm
   b) Collecting data
   c) A student's report card
   d) Building a neural network
   Answer: c) A student's report card

2. Why is model evaluation important?
   a) To check only data size
   b) To identify model's strengths and weaknesses
   c) To increase dataset labels
   d) To avoid using training data
   Answer: b) To identify model's strengths and weaknesses

3. What is the main goal of model evaluation?
   a) Minimize error and maximize accuracy
   b) Maximize error and minimize accuracy
   c) Focus only on training dataset
   d) Ignore unseen data
   Answer: a) Minimize error and maximize accuracy

4. Which stage comes after modeling in the AI project cycle?
   a) Problem scoping
   b) Data exploration
   c) Evaluation
   d) Deployment
   Answer: c) Evaluation

5. Overfitting occurs when:
   a) Model is trained and tested on the same data
   b) Model learns from testing dataset
   c) Model ignores training dataset
   d) Model uses too little data
   Answer: a) Model is trained and tested on the same data

Section B: Train-Test Split

6. Train-test split is used in:
   a) Only unsupervised learning
   b) Only reinforcement learning
   c) Any supervised learning algorithm
   d) Rule-based models
   Answer: c) Any supervised learning algorithm

7. Why don't we test a model on training data?
   a) It gives wrong outputs
   b) It may memorize training data and show false accuracy
   c) It requires more computation
   d) It reduces dataset size
   Answer: b) It may memorize training data and show false accuracy

8. The training dataset is used to:
   a) Test predictions
   b) Make the model learn
   c) Compare actual and predicted values
   d) Reduce overfitting
   Answer: b) Make the model learn

9. The testing dataset is used to:
   a) Train the model
   b) Improve data labeling
   c) Evaluate model performance on new data
   d) Remove errors from dataset
   Answer: c) Evaluate model performance on new data

10. What is the main risk of not using a train-test split?
    a) Increased dataset size
    b) Decreased data quality
    c) False confidence in accuracy
    d) No labeled features
    Answer: c) False confidence in accuracy

Section C: Accuracy and Error

11. Accuracy measures:
    a) How many predictions are wrong
    b) Total correct predictions ratio

c) Size of dataset
d) Model complexity
Answer: b) Total correct predictions ratio

12. Error is defined as:
    a) Data labeling process
    b) Wrong prediction compared to actual outcome
    c) Difference between training and testing data
    d) Noise in dataset
    Answer: b) Wrong prediction compared to actual outcome

13. If a model predicts 9 out of 10 correctly, accuracy is:
    a) 80%
    b) 85%
    c) 90%
    d) 95%
    Answer: c) 90%

14. In model evaluation, the target is:
    a) Minimize error
    b) Maximize accuracy
    c) Both (a) and (b)
    d) None of these
    Answer: c) Both (a) and (b)

15. Error Rate = ?
    a) Error ÷ Actual Value
    b) Actual ÷ Predicted
    c) 1 – Accuracy
    d) Both (a) and (c)
    Answer: d) Both (a) and (c)

Section D: Confusion Matrix

16. Confusion matrix compares:
    a) Input vs output
    b) Actual vs predicted values
    c) Features vs labels
    d) Training vs testing
    Answer: b) Actual vs predicted values

17. True Positive means:
    a) Predicted positive is actually negative
    b) Predicted positive is actually positive
    c) Predicted negative is actually positive
    d) Predicted negative is actually negative
    Answer: b) Predicted positive is actually positive

18. True Negative means:
    a) Predicted negative is actually negative
    b) Predicted positive is actually negative
    c) Predicted positive is actually positive
    d) Predicted negative is actually positive
    Answer: a) Predicted negative is actually negative

19. False Positive means:
    a) Predicted positive but actually negative
    b) Predicted negative but actually positive
    c) Predicted correctly
    d) No prediction
    Answer: a) Predicted positive but actually negative

20. False Negative means:
    a) Predicted negative but actually positive
    b) Predicted positive but actually negative
    c) Predicted correctly
    d) Missing dataset values
    Answer: a) Predicted negative but actually positive

Section E: Classification Metrics

21. Classification accuracy = ?
    a) (TP + TN) ÷ (TP + TN + FP + FN)
    b) (TP + FN) ÷ Total
    c) (FP + FN) ÷ Total
    d) TP ÷ (TP + FP)
    Answer: a) (TP + TN) ÷ (TP + TN + FP + FN)

22. Which metric is misleading in unbalanced datasets?
    a) Precision
    b) Recall

c) Accuracy
d) F1 Score
Answer: c) Accuracy

23. Precision formula = ?
    a) TP ÷ (TP + FP)
    b) TP ÷ (TP + FN)
    c) TN ÷ (TN + FP)
    d) (TP + TN) ÷ Total
    Answer: a) TP ÷ (TP + FP)

24. Recall formula = ?
    a) TP ÷ (TP + FP)
    b) TP ÷ (TP + FN)
    c) TN ÷ (TN + FP)
    d) FN ÷ (TP + FN)
    Answer: b) TP ÷ (TP + FN)

25. F1 Score combines:
    a) Accuracy and Error
    b) Precision and Recall
    c) True Positive and True Negative
    d) Error and Overfitting
    Answer: b) Precision and Recall

Section F: Use Cases of Metrics

26. In spam detection, which metric is more important?
    a) Recall
    b) Precision
    c) Accuracy
    d) Error rate
    Answer: b) Precision

27. In cancer diagnosis, which metric is critical?
    a) Precision
    b) Recall
    c) F1 Score
    d) Accuracy
    Answer: b) Recall

28. In fraud detection, priority should be:
    a) Reducing False Negatives
    b) Reducing False Positives
    c) Maximizing accuracy
    d) Ignoring errors
    Answer: a) Reducing False Negatives

29. In satellite launch weather prediction, we should reduce:
    a) False Positives
    b) False Negatives
    c) True Positives
    d) Accuracy values
    Answer: a) False Positives

30. When unsure whether FP or FN is more important, we use:
    a) Accuracy
    b) Recall
    c) Precision
    d) F1 Score
    Answer: d) F1 Score

Section G: Case-based Calculations

31. If TP = 90, FP = 40, TN = 820, FN = 50, Accuracy = ?
    a) 91%
    b) 90%
    c) 92%
    d) 89%
    Answer: a) 91%

32. If TP = 80, FP = 30, TN = 850, FN = 40, Precision = ?
    a) 80 ÷ 110 = 72.7%
    b) 80 ÷ 120 = 66.6%
    c) 80 ÷ 850 = 9.4%
    d) 80 ÷ 920 = 8.6%
    Answer: a) 72.7%

33. If TP = 120, FN = 60, Recall = ?
    a) 120 ÷ 180 = 66.6%
    b) 120 ÷ 60 = 200%

c) 60 ÷ 120 = 50%

d) 120 ÷ 240 = 50%

Answer: a) 120 ÷ 180 = 66.6%

34. If Precision = 0.8 and Recall = 0.75, F1 Score = ?

a) 0.78

b) 0.82

c) 0.70

d) 0.60

Answer: a) 0.78

35. If a model correctly classifies 900 out of 1000 samples, accuracy = ?

a) 85%

b) 90%

c) 95%

d) 97%

Answer: b) 90%

Section H: Error & Ethical Concerns

36. Absolute error is measured as:

a) Actual – Predicted (with sign)

b) |Actual – Predicted|

c) Predicted ÷ Actual

d) (Predicted + Actual)/2

Answer: b) |Actual – Predicted|

37. Ethical concerns in evaluation include:

a) Bias in datasets

b) Misuse of metrics

c) Unfair predictions

d) All of the above

Answer: d) All of the above

38. If a model wrongly flags many safe videos as unsafe, the main issue is:

a) High recall

b) Low precision

c) High error

d) Overfitting

Answer: b) Low precision

39. If a model misses too many fraud cases, it suffers from:
    a) Low recall
    b) Low precision
    c) High accuracy
    d) Overfitting
    Answer: a) Low recall

40. Which metric is also called **Sensitivity**?
    a) Precision
    b) Recall
    c) F1 Score
    d) Accuracy
    Answer: b) Recall

Section I: Assertions

41. Assertion: Accuracy measures correct predictions.
    Reason: Accuracy is directly proportional to model performance.
    a) Both A and R true, R explains A
    b) Both A and R true, R not explanation
    c) A true, R false
    d) A false, R true
    Answer: a) Both A and R true, R explains A

42. Assertion: Row sums in confusion matrix = total instances of that actual class.
    Reason: This helps calculate precision & recall.
    a) Both A and R true, R explains A
    b) Both A and R true, R not explanation
    c) A true, R false
    d) A false, R true
    Answer: a) Both A and R true, R explains A

43. Assertion: Accuracy is always the best metric.
    Reason: Accuracy works well in unbalanced datasets.
    a) Both A and R true
    b) A true, R false
    c) Both A and R false
    d) A false, R true
    Answer: b) A true, R false

44. Assertion: High precision means fewer false positives.
    Reason: Precision = TP ÷ (TP + FP).
    a) Both A and R true, R explains A
    b) Both A and R true, R not explanation
    c) A true, R false
    d) A false, R true
    Answer: a) Both A and R true, R explains A

45. Assertion: F1 Score balances precision and recall.
    Reason: F1 = Harmonic mean of precision and recall.
    a) Both A and R true, R explains A
    b) Both A and R true, R not explanation
    c) A true, R false
    d) A false, R true
    Answer: a) Both A and R true, R explains A

Section J: Miscellaneous

46. Which metric to use in highly unbalanced fraud detection?
    a) Precision
    b) Recall
    c) Accuracy
    d) Error rate
    Answer: b) Recall

47. A model with high recall but low precision means:
    a) Finds most positives but with many false alarms
    b) Ignores many true positives
    c) Predicts very few negatives
    d) Has very low accuracy
    Answer: a) Finds most positives but with many false alarms

48. High precision but low recall means:
    a) Very accurate positives, but misses many positives
    b) Detects all positives with false alarms
    c) Predicts random labels
    d) Balanced model
    Answer: a) Very accurate positives, but misses many positives

49. Which metric should be used when both FP and FN are equally important?
    a) Accuracy
    b) Precision
    c) Recall
    d) F1 Score
    Answer: d) F1 Score

50. In model evaluation, what should always be the focus?
    a) Context of problem
    b) Dataset balance
    c) Choice of correct metric
    d) All of the above
    Answer: d) All of the above