EXAM 1 PRACTICE - WEEKS 1-4

Practice Exam

CIS 350: AI for Cybersecurity - Smart Defense for the Digital Business

Colorado State University - Computer Information Systems

About this document

This is a PRACTICE exam to help you prepare for Exam 1. The questions are different from the real exam but cover the same Weeks 1-4 material in the same format and length (100 points; 90 minutes): multiple choice, true/false, short answer, and applied scenarios. Try it closed-book and timed, then check the separate answer key. The real Exam 1 is an online exam in Canvas with LockDown Browser.

Total: 100 points, 90 minutes. Part A multiple choice (20 x 2 = 40), Part B true/false (10 x 1 = 10), Part C short answer (5 x 4 = 20), Part D applied scenarios (3 x 10 = 30). No heavy math, no coding.

Part A - Multiple Choice (20 x 2 = 40 pts)

1. Which is the BEST example of protecting INTEGRITY?

A. Encrypting a laptop's hard drive
B. Keeping a backup server for uptime
C. Using a checksum to detect if a file was altered
D. Installing a firewall

2. A distributed denial-of-service (DDoS) attack most directly threatens:

A. Confidentiality
B. Availability
C. Integrity
D. Non-repudiation

3. 'A weakness that an attacker could exploit' defines a:

A. Vulnerability
B. Threat
C. Risk
D. Control

4. An attacker phones an employee pretending to be IT support to get a password. This is:

A. Malware
B. A denial-of-service attack
C. A software vulnerability
D. Social engineering

5. Multi-factor authentication (MFA) is primarily which kind of control?

A. Prevention
B. Detection
C. Response
D. Recovery

6. Which is an example of AI that is NOT machine learning?

A. A spam filter trained on labeled emails
B. A fraud detector that learns from past transactions
C. A hand-coded rule-based chatbot that follows fixed if/then rules
D. A model that clusters similar logins

7. In a dataset of emails, the column marking each email 'spam' or 'legit' is the:

A. Feature
B. Row
C. Index
D. Label

8. Which of these is a source of security data?

A. Firewall and system logs
B. A printed employee handbook
C. A whiteboard sketch
D. A hallway conversation

9. Adding more data helps a model only if the data is:

A. Encrypted
B. Relevant and good quality
C. Stored offline
D. Collected in one day

10. 'Structured data' usually means data that is:

A. Written in paragraphs
B. Made of images only
C. Organized in rows and columns like a table
D. Impossible to analyze

11. Supervised learning means the model learns from:

A. Labeled examples with known answers
B. Only unlabeled data
C. No data
D. Hand-written rules only

12. The main reason to hold out a test set is to:

A. Speed up training
B. Estimate performance on data the model has not seen
C. Reduce the number of features
D. Make the dataset larger

13. A model scoring 100% on training data but poorly on new data is:

A. Underfitting
B. Perfectly tuned
C. Overfitting
D. Unsupervised

14. In a decision tree, an internal (non-leaf) node represents:

A. A final prediction
B. A random choice
C. The training data size
D. A yes/no question about a feature

15. Increasing k in k-nearest neighbors generally makes the decision boundary:

A. Smoother
B. Noisier
C. Disappear
D. Always perfect

16. Predicting the dollar cost of a data breach is an example of:

A. Classification
B. Clustering
C. A confusion matrix
D. Regression

17. A model produces TP = 40, TN = 30, FP = 20, FN = 10. Its accuracy is:

A. 70%
B. 40%
C. 60%
D. 80%

18. A spam filter sends your CEO's genuine email to the spam folder. This is a:

A. True positive
B. False positive
C. True negative
D. False negative

19. A model has high recall when it:

A. Rarely raises false alarms
B. Rarely misses real positives
C. Runs quickly
D. Uses few features

20. Raising the decision threshold (requiring more confidence to flag) typically:

A. Increases false positives
B. Has no effect
C. Reduces false positives but increases false negatives
D. Eliminates all errors

Part B - True / False (10 x 1 = 10 pts)

21. Confidentiality means keeping data usable and available at all times. (T / F)
22. Phishing is a form of social engineering. (T / F)
23. A rule-based program learns patterns from data on its own. (T / F)
24. In supervised learning, each training example includes a known label. (T / F)
25. Overfitting means a model performs poorly on its training data. (T / F)
26. k-nearest neighbors predicts using the labels of nearby points. (T / F)
27. Classification predicts a numeric value on a scale. (T / F)
28. A false positive means a legitimate item was wrongly flagged. (T / F)
29. With imbalanced data, accuracy can be high even if the rare class is missed. (T / F)
30. Recall measures how many of the flagged items were actually correct. (T / F)

Part C - Short Answer (5 x 4 = 20 pts)

31. Explain the CIA triad and give a real example of violating each of the three goals.

32. Give two differences between machine learning and rule-based software, and name one security task where ML is the better choice.

33. What is overfitting, and how does a train/test split help you detect it?

34. Explain the difference between a false positive and a false negative in fraud detection, and give the business cost of each.

35. A model has 99% accuracy on a dataset where 99% of transactions are legitimate. Explain why this accuracy may be meaningless, and name a better metric.

Part D - Applied Scenarios (3 x 10 = 30 pts)

36. A law firm's email filter (10 pts)

A law firm turned its spam filter up very high. Now some genuine client emails are landing in the spam folder, and staff are missing messages.

(a) Which error type is this - false positive or false negative? (2 pts)
(b) Is precision or recall the low metric here? (2 pts)
(c) Should they raise or lower the threshold, and what trade-off does that create? (3 pts)
(d) Name one non-model change (process or people) that would reduce the harm. (3 pts)

37. A hospital device-monitoring system (10 pts)

A hospital wants to flag unusual behavior from networked medical devices. They have lots of activity logs but almost no labeled examples of real attacks.

(a) Can they train a supervised classifier well right now? Why or why not? (3 pts)
(b) What would the data need for a supervised approach? (2 pts)
(c) If medical devices become unusable during an incident, which CIA goal is most at stake? (2 pts)
(d) What kind of approach suits 'find the unusual points without labels'? (3 pts)

38. A bank deploying a fraud model (10 pts)

A bank built a fraud detector and wants to deploy it. Fraud is rare (about 1% of transactions).

(a) Why should they test on held-out data before deploying? (2 pts)
(b) They report only accuracy. Name two other metrics they should report and why. (3 pts)
(c) Because fraud is rare, what problem does that create for accuracy? (2 pts)
(d) They must pick an alert threshold. What business trade-off do they weigh? (3 pts)

End of practice exam. Check your answers with the separate answer key.