EXAM 2 STUDY GUIDE - WEEKS 5-8

Unsupervised, Semi-Supervised, RL & Attacker AI

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

Colorado State University - Computer Information Systems

How to use this study guide

This guide pulls together everything Exam 2 covers (Weeks 5-8): unsupervised learning and clustering; anomaly detection; semi-supervised and reinforcement learning; and how attackers use AI (with defenses). Study actively - cover the answers and recall them, then redo the self-check questions. Exam 2 is online (asynchronous), 90 minutes, open-notes, no heavy math or coding.

Estimated review time: 45-60 minutes. Pair this with your lecture slides and Handouts 5-7.

What Exam 2 Covers

Exam 2 is non-cumulative and covers Weeks 5-8. The arc: we moved from learning WITHOUT labels (unsupervised), to learning with a FEW labels or from REWARDS (semi-supervised and reinforcement), to how attackers turn AI against us (and how we defend). The exam rewards understanding how these connect - especially matching the right method to a security problem.

Figure 1. The Weeks 5-8 arc: unsupervised learning -> semi-supervised & RL -> attacker AI.

Exam logistics

100 points, 90 minutes, online and ASYNCHRONOUS via Canvas (proctored with LockDown Browser). Format: multiple-choice, true/false, short-answer, and applied scenarios. Open-notes and individual - no collaboration, no AI. No heavy math or coding.

Weeks 5-6 - Unsupervised Learning & Clustering

- Unsupervised learning finds patterns in data with NO labels.
- Clustering groups similar records; k-means needs you to choose k (use the ELBOW method).
- Always SCALE features first - distance-based methods are hijacked by big-range features.
- Hierarchical clustering builds a tree (dendrogram) you CUT to pick the number of clusters.
- A cluster in security = a behavior group (e.g., normal vs unusual traffic) you then interpret.

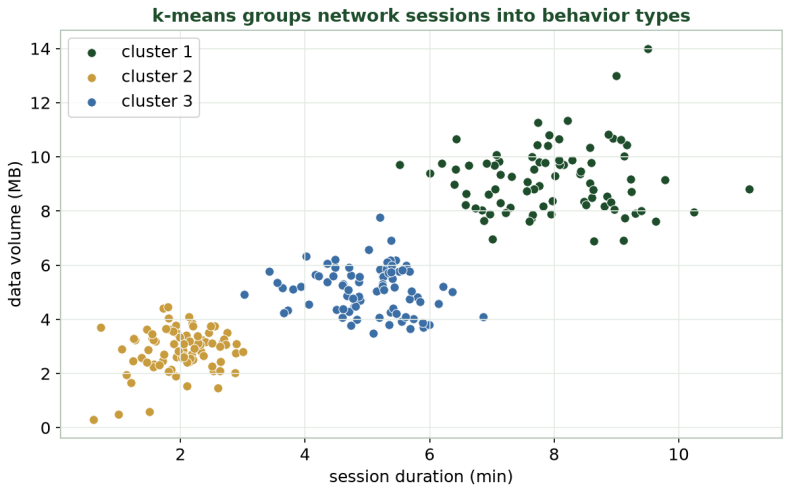

Figure 2. k-means groups network sessions into behavior types (choose k with the elbow method).

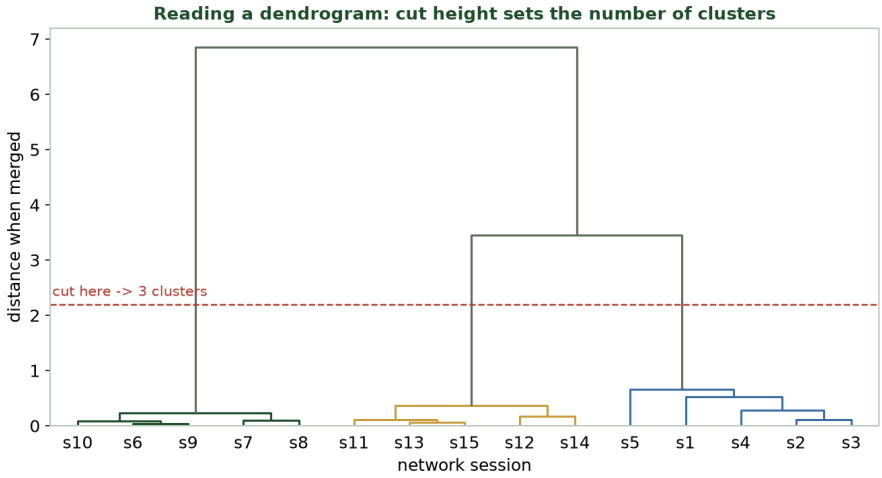

Figure 3. A dendrogram (hierarchical clustering): a horizontal cut sets the number of clusters.

Week 6 - Anomaly Detection

An anomaly is a rare record that fits no normal pattern. Anomaly detection learns 'normal' and flags the rest - with NO labels - so it can catch novel attacks. The Isolation Forest flags points that are isolated in only a few random splits and gives every point a score; you set 'contamination' (the expected anomaly fraction). Local Outlier Factor (LOF) flags points sparse relative to their neighbors.

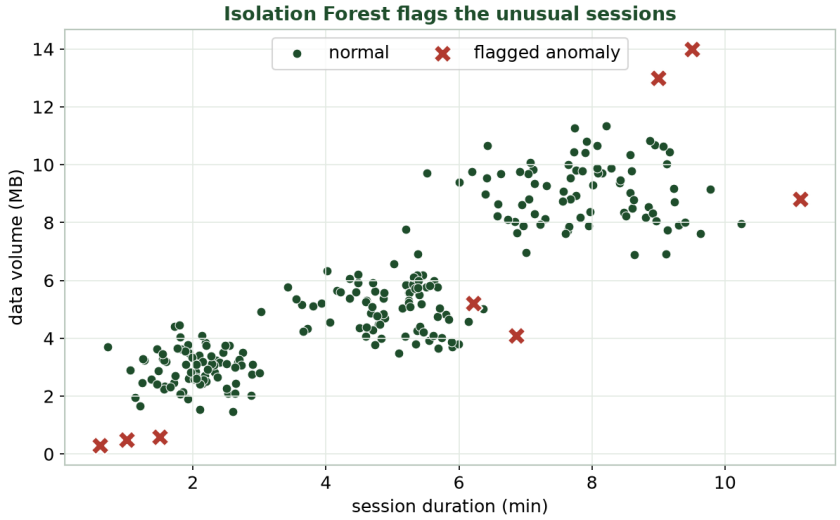

Figure 4. An Isolation Forest flags the unusual sessions - novel activity with no labels needed.

Week 7 - Semi-Supervised & Reinforcement Learning

- Semi-supervised: a FEW labels + lots of unlabeled data; labels spread to similar points.
- In scikit-learn, mark unlabeled rows with -1. It helps most when labels are SCARCE.
- Reinforcement learning (RL): an agent learns a POLICY from rewards by interacting - no dataset.
- Exploration vs. exploitation: try new actions vs. use known-good ones.
- Security is a defender-attacker GAME; RL can learn adaptive defenses.

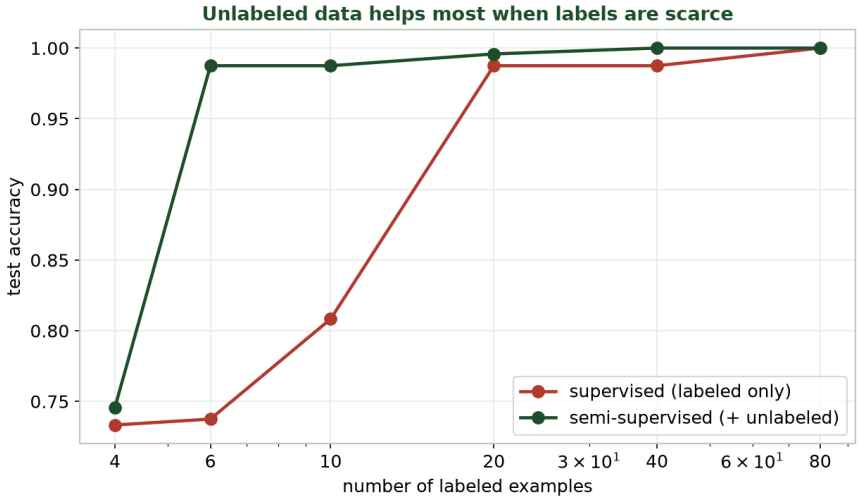

Figure 5. Semi-supervised learning helps most when labels are scarce (security's normal state).

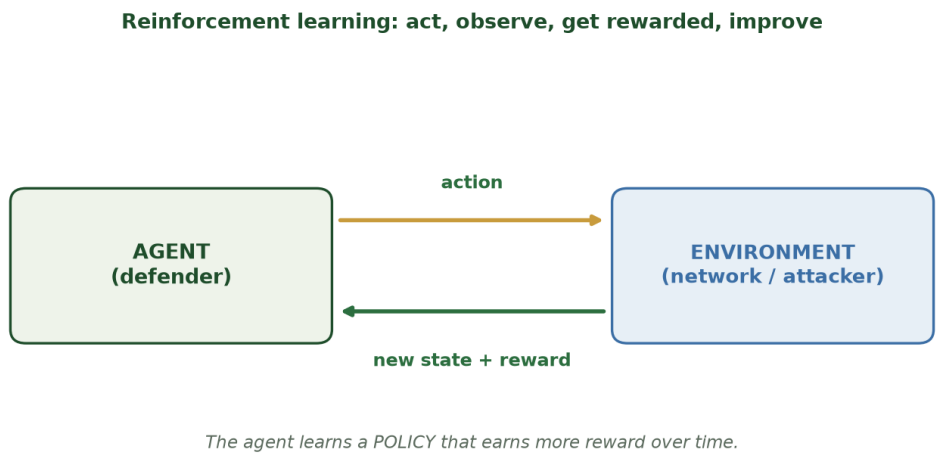

Figure 6. The RL loop: act, observe a new state and reward, update the policy - repeat.

Week 8 - How Attackers Use AI (and Defenses)

AI is dual-use: it makes attacks cheaper, bigger, more personalized, and adaptive. Human-facing attacks include AI phishing, deepfakes/voice fraud, and credential stuffing; machine-facing attacks include automated reconnaissance and adaptive malware. Crucially, every technique maps to a defense - most of which you already learned.

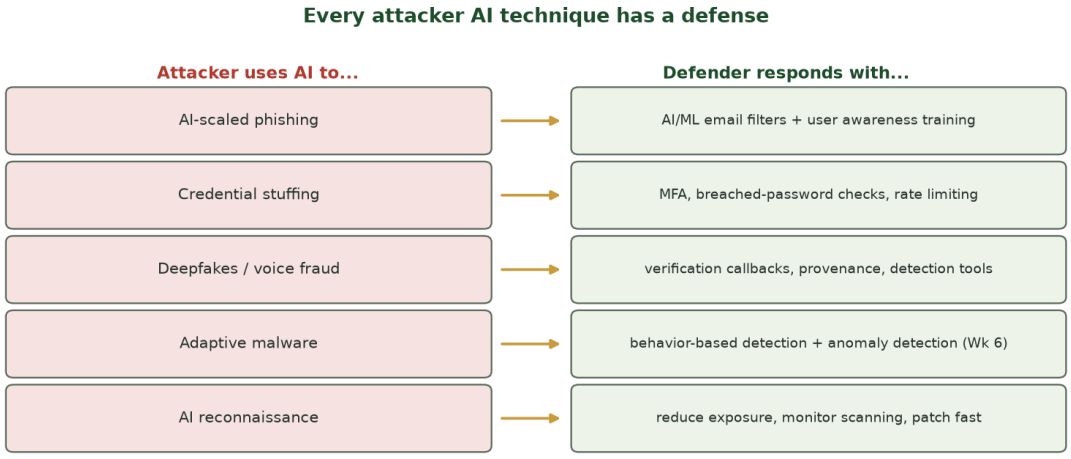

Figure 7. Each attacker AI technique maps to a concrete defense (MFA, ML filters, anomaly detection...).

Key Terms to Know Cold

- Unsupervised learning; clustering; k-means; elbow method; feature scaling.
- Hierarchical clustering; dendrogram; cluster center.
- Anomaly / outlier; Isolation Forest; Local Outlier Factor; contamination.
- Semi-supervised learning; the -1 (unlabeled) convention.
- Reinforcement learning; agent, environment, reward, policy; exploration vs. exploitation.
- Four paradigms: supervised, unsupervised, semi-supervised, reinforcement.

- AI phishing; deepfake; credential stuffing; reconnaissance; adaptive malware.
- MFA; defense in depth (people, process, technology).

Self-Check Questions

1. What makes a task unsupervised, and what does the elbow method choose?
2. Why must you scale features before k-means? What goes wrong if you don't?
3. How does an Isolation Forest decide something is an anomaly?
4. What is semi-supervised learning, and when does it help most?
5. What does a reinforcement-learning agent learn, and from what?
6. Why is security described as a defender-attacker game?
7. Name three ways attackers use AI and a defense for each.
8. Why isn't automated detection alone enough against AI phishing?

If you can answer all eight from memory and match each of the four paradigms to a security problem, you are ready. Remember: AI assistance and collaboration are not permitted on the exam.

References

- scikit-learn User Guide - Clustering, Outlier Detection, and Semi-Supervised Learning.
- R. S. Sutton & A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. (free online).
- NIST, AI Risk Management Framework (AI RMF 1.0), 2023; SP 800-63B (authentication).
- CISA, Phishing Guidance, 2023; MITRE ATT&CK and MITRE ATLAS.
- Liu et al., 'Isolation Forest' (2008); Breunig et al., 'LOF' (2000).
- Chio & Freeman, Machine Learning and Security (O'Reilly).