CYS5550: Introduction to Artificial Intelligence

Week 4 Assessment: Learning from Data Foundations of Statistical Learning

Section A: Multiple Choice (60 points — 2 points each)

AI Paradigms: Knowledge-Driven vs. Data-Driven (Questions 1-10)

1. Which of the following best characterizes knowledge-driven AI?
   - B) Human expertise encoded explicitly as rules
2. The "knowledge acquisition bottleneck" in expert systems refers to:
   - B) Difficulty extracting and encoding expert knowledge
3. According to Tom Mitchell's definition, a computer program learns if:
   - B) Its performance on a task improves with experience
4. In supervised learning, the training data consists of:
   - B) Input-output pairs (examples with labels)
5. Which scenario is best suited for unsupervised learning?
   - B) Discovering natural groupings in customer data
6. Why did data-driven approaches outperform rule-based systems for handwriting recognition?
   - B) Handwriting has too much variability for manually specified rules
7. Reinforcement learning differs from supervised learning in that:
   - B) It learns from rewards/penalties through interaction
8. Which is an advantage of knowledge-driven AI over data-driven AI?
   - C) Explainability and verifiable reasoning
9. Neuro-symbolic AI attempts to
   - B) Combine neural networks with symbolic reasoning
10. In the handwriting recognition example, the MNIST dataset contains:
    - B) 60,000 labeled images of handwritten digits

Statistical Learning Theory (Questions 11-20)

11. Generalization in machine learning refers to:
    - B) Performance on new, unseen data

12. Overfitting occurs when a model:
    - B) Performs well on training data but poorly on test data
13. Underfitting is characterized by:
    - B) High training error and high test error
14. Underfitting is characterized by:
    - B) High training error and high test error
15. High variance in a model means
    - B) The model's predictions change significantly with different training data
16. PAC learning theory addresses the question:
    - B) Can we learn a good hypothesis with high probability from finite data?
17. In PAC learning, the parameter ε (epsilon) represents:
    - B) The maximum allowable error
18. In PAC learning, the parameter δ (delta) represents:
    - B) The probability of failure (1 - confidence)
19. Sample complexity in PAC learning refers to:
    - B) The number of examples needed to learn with specified accuracy and confidence
20. According to PAC theory, which factor does NOT increase sample complexity?
    - D) Faster computer hardware

VC Dimension and Model Selection (Questions 21-30)

21. The VC dimension of a hypothesis class measures:
    - C) The "flexibility" or capacity to fit different patterns
22. A hypothesis class H "shatters" a set of points if:
    - C) It can realize every possible labeling of those points
23. The VC dimension of linear classifiers in 2D is:
    - B) 3
24. Why can't a linear classifier in 2D shatter 4 points in the XOR pattern?
    - B) No single line can separate opposite corners with same labels
25. If a model has a very high VC dimension, it:
    - C) Requires more training data to generalize well
26. The purpose of a validation set is to:

- C) Tune hyperparameters and select the best mode

27. Why should you never tune your model based on test set performance?
    - B) It "contaminates" the test set and biases the final evaluation
28. In 5-fold cross-validation:
    - B) Each data point is used for testing exactly once
29. Precision measures:
    - B) The proportion of predicted positives that are truly positive
30. Recall (sensitivity) measures:
    - B) The proportion of actual positives correctly identified

Section B: True/False (10 points — 1 point each)

Write T for True or F for False.

1. _False_ Knowledge-driven AI systems can automatically improve their performance by observing more examples.

2. _True_ In supervised learning, the algorithm is given the correct answers (labels) during training.

3. _False A model with 100% training accuracy is always the best choice for deployment.

4. _True_ Overfitting is more likely when the model is complex and the training dataset is small.

5. __False__ The bias-variance tradeoff suggests that the simplest model is always the best.

6. _False_ VC dimension measures how many training examples are needed, not model complexity.

7. _True_ A linear classifier in 2D can correctly separate any arrangement of 3 points into two classes.

8. _True_ The test set should only be used once, at the very end of model development.

9. _True_ Cross-validation allows every data point to be used for both training and testing.

10. _False_ A model with high precision but low recall catches most positive cases but also has many false alarms.

Section C: Short Answer (30 points — 2 points each)

1. What is the fundamental difference between knowledge-driven and data-driven AI?
   - Knowledge-driven AI makes use of explicitly defined knowledge that includes rules, logic, and expert systems. Data-driven AI uses machine learning techniques or statistics to learn the patterns directly from the large supply of data.
2. Explain why rule-based systems struggled with handwriting recognition while learned systems succeeded.
   - Handwritten text is highly varied with respect to shapes, sizes, and styles. It is difficult to lay out strict guidelines, but learned systems are successful since they are able to learn from a large set of examples of handwritten words.
3. Give an example of a task where knowledge-driven AI would be preferred over data-driven AI, and explain. Why?
   - Legal reasoning and medical diagnosis with expert rules are more appropriate for Knowledge-Driven Artificial Intelligence. These include tasks that need specific reasoning, transparency, and compliance with legal provisions, rather than data-based learning.
4. What is neuro-symbolic AI, and what problem does it aim to solve?
   - Neuro-symbolic artificial intelligence makes an integration of both the neural networks and the symbolic systems of reasoning. It seeks to address the problem facing both the neural models and the symbolic models. Namely, the problem is that the neural models may not be very clear or explicable.

Generalization and Overfitting

5. Explain why generalization, not memorization, is the central goal of machine learning
   - Machine learning models are supposed to generalize well on unseen data, not only on the training set. The model will perform very poorly in real situations when it encounters new data that is different from the training data.
6. A model achieves 99% training accuracy but only 62% test accuracy. What problem does this indicate, and what are two possible causes?
   - This shows overfitting; the possible reasons include a complex model and inadequate and unrepresentative training data.

7. Explain the bias-variance tradeoff in your own words. What happens at each extreme?
   - Bias means errors due to oversimplification of the model, and variance means errors due to overcomplexity of the model. Models that are overly simple due to high bias underfit the data, and those that are overly complex due to high variance overfit the data.
8. How can you tell from training and validation curves whether a model is underfitting?
   - Both training accuracy and validation accuracy will be low and close to each other, which means the model is too simple to capture the underlying pattern.

PAC Learning and VC Dimension

9. In PAC learning, what do the parameters ε and δ represent, and how do they affect sample complexity?
   - ε denotes the permissible error, and δ denotes the confidence level. Lower values of ε and δ require a higher number of training samples.
10. Explain in plain language what it means for a hypothesis class to have VC dimension = 4.
    - It means the model can correctly classify all kinds of labelings of any 4 data points, but it cannot do this for at least one configuration of 5 points.
11. Why does a model with higher VC dimension require more training data to generalize well?
    - Higher VC dimension means the model is more expressive and can fit many patterns, including noise. More data is needed to ensure the model learns the true pattern rather than overfitting.

Model Selection and Evaluation

12. What is the difference between a validation set and a test set? Why do we need both?
    - The set of validation is utilized to adjust the parameters, and the testing set is used only one time to evaluate the performance, ensuring it is unbiased.
13. Explain how 5-fold cross-validation works and why it's useful when data is limited.
    - The data is split into five parts, on which the model is trained four times each and tested on one, for five times each. This gives an efficient usage of the limited data while in no way compromising on the reliability of the performance estimate.

14. A spam filter has 95% precision and 60% recall. What does this mean in practical terms for email users?
    - Most emails marked as spam are truly spam, so few legitimate emails are misclassified. However, the filter misses many spam emails, which still reach the inbox.
15. You are comparing two models: Model A has 85% accuracy with 1% standard deviation across folds; Model B has 87% accuracy with 8% standard deviation. Which would you choose and why?
    - This makes Model A preferable due to its stability and consistency across different data splits. The high variance of Model B suggests unreliability of the performance, even with a somewhat higher accuracy.