UNIT 6 ASSIGNMENT

Special Topics: Ensemble Models and Unsupervised Learning

## Instructions

The questions below will prepare you for future interviews as they relate to concepts discussed throughout the week. You’ve practiced these concepts in the coding activities, exercises, and coding portion of the assignment. Now, let’s formulate your programming into well-thought responses.

Except as indicated, use this document to record all your assignment work and responses to any questions. At a minimum, you will need to turn in a digital copy of this document to your facilitator   
as part of your assignment completion. You may also have additional supporting documents that   
you will need to submit. Your facilitator will provide feedback to help you work through your findings.

**Note:** Though your work will only be seen by those grading the course and will not be used or shared outside the course, you should take care to obscure any information you feel might be of a sensitive or confidential nature.

*Begin your assignment by completing the questions below. Directions to submit your work can be found on the assignment page. Information about the grading rubric is available on any of the course assignment pages online. Do not hesitate to contact your facilitator if you have any questions about the assignment.*

Week 6 Written Portion

# Implementing Ensemble Models and Unsupervised Clustering

Answer the questions below about ensemble models and unsupervised learning.

1. Explain ensemble modeling. What is the advantage of using this technique?

|  |
| --- |
| Ensemble Modeling combines different models and average out the errors that individual models might have. It can leverage the strengths of various models and mitigate their weakness. This increases robustness and reduces overfitting. |

1. Explain what bias and variance are, along with the bias-variance tradeoff.

|  |
| --- |
| Model error is the sum of estimation bias and estimation variance. Estimation bias refers to model rigidity that prevents adaptation data, and estimation variance refers to model flexibility that causes the estimated model to be sensitive to data. As you decrease bias, variance tends to increase and vice versa. So it is important to find a balance and minimize both bias and variance. |

1. Explain the differences among the ensemble methods bagging, boosting, and stacking.

|  |
| --- |
| Bagging trains multiple models on random subsets of the data with replacement, reducing variance. Boosting trains models sequentially, each focusing on examples misclassified by the previous model, enhancing accuracy by reducing bias. Stacking trains a meta-model on the predictions of multiple base models, reducing both variance and bias. |

1. Explain the random forest algorithm and how it relates to decision trees and bagging.

|  |
| --- |
| Random forest algorithm is an ensemble learning method that constructs multiple decision trees, each trained on a random subset of the data and features. It employs bagging to generate diverse training subsets, thereby mitigating overfitting and enhancing generalization through the averaging of these trees' results. |

1. What’s the difference between gradient boosting decision trees and random forest?

|  |
| --- |
| Both are ensemble learning methods. Random forest builds each tree independently and combines their predictions through averaging, however, Gradient boosting decision trees builds sequentially, each one correcting the errors of its predecessors. |

1. What’s the difference between supervised and unsupervised learning?

|  |
| --- |
| Supervised learning is training a model on labeled data, where the algorithm learns to map input data to known output labels. On the other hand, unsupervised learning deals with data which does not have labels. It aims to discover patterns by clustering similar data points. |

1. Give an example of an ML problem where you would use unsupervised clustering.

|  |
| --- |
| Unsupervised clustering can be used in marketing to provide personalized customer engagement. You have data on customer demographics, purchasing behavior, and other relevant features, but you don’t have predefined labels for different types of customers. The unsupervised model can help identify groupings or segments within the customer. |

*To submit this assignment, please refer to the instructions in the course*.