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 is the idea of using and building several ML models to come up with the best prediction for the testing dataset. The advantage is to reduce the model’s estimation bias and variance |

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

|  |
| --- |
| Bias is the difference between the predicted outcome and the actual outcome  Variance tells us the spread of the data meaning how far apart the features of the dataset are or how different each data point is from each other.  Bias-variance trade off is the concept of what happens to a model when there is high bias and low variance or vice versa. The idea is to minimize both. |

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

|  |
| --- |
| Bagging: we use one algorithm but use randomly sampled data with replacement  Boosting: we use a chain of models that aim to reduce errors over multiple iterations  Stacking: combining multiple models and aggregating their output prediction by taking their average for example |

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

|  |
| --- |
| Random forest is an example of bagging. It uses multiple decision trees on different samples of the dataset, and also experiments with different features of each subset of the dataset. In the end after all the models have been run on their respective sub-datasets, their results are usually averaged to one final prediction. |

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

|  |
| --- |
| Random forest uses multiple decision trees on subsets of the original dataset, whereas gradient boosting decision trees uses the original dataset on one tree, and then in the next iteration, uses the previous results to further enhance the prediction, and continues with this process until the defined iterations is met |

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

|  |
| --- |
| Biggest difference is in supervised learning, you are given an expected output when training the data, meaning the correct answer is already outlined for us. For unsupervised learning, the model does not have a clear defined expected output before the model runs. It is the model’s job to identify patterns and classify the input to predict an output |

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

|  |
| --- |
| Maybe if I have a document with letters of an unknown language and I want my ML model to cluster the similiar looking letters together so later, I can try to distinguish and further analyze each letter on its own |

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