UNIT 5 ASSIGNMENT

Choosing Your Model

## 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 project 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 project 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.

*Complete each project part as you progress through the course. Wait to submit the project until all parts are complete. Begin your course project by completing Part One below. A submit button can be found on the final Course Project assignment page. Information about the grading rubric is available on any of the course project assignment pages online. Do not hesitate to contact your facilitator if you have any questions about the project.*

Week 5 Written Portion

# Choosing Your Model

Answer the questions below about selecting the correct models and approaches to solve your machine learning problems.

## Questions:

1. What is model selection and why is performing model selection important?

| Model selection is the process of choosing what algorithm, features and hyper-parameters will comprise your final model. This is important because of the ‘no free lunch’ theorem, which states that no algorithm is universally best for all problems. Thus, one has to explore and test many model options until one can empirically identify the optimal one. |
| --- |

1. What is out of sample validation and why is this key in helping us choose the best performing model?

| Out of sample validation is the process of validating or evaluating a model against data that is not part of the training data. The out-of-sample data is the best proxy for new and unseen data. This process helps to avoid overfitting and to ensure proper generalization of the trained model. |
| --- |

1. What is cross-validation and what is the benefit of performing cross validation?

| Cross-validation is a validation procedure iteratively validates a model against different partitions of the data. This is a powerful to leverage all of one’s data for model training while also holding data out for validation. Cross validation also enables one to compute standard errors on a performance metric. |
| --- |

1. What's the difference between Feature Engineering vs. Feature Selection? What are the benefits of feature selection?

| Feature engineering is the process of creating features, and Feature selection is the process of deciding which features to include or drop from your model. Feature selection helps to avoid overfitting by removing spurious features. It also simplifies computation, as each feature requires compute resources to build and include in a model. |
| --- |

1. What is the difference between the classification evaluation metrics accuracy, precision, and recall?

| Accuracy tells us what % of our predictions are correct overall. Precision tells us what % of our positive predictions are correct. Recall tells us what % of the actual positive cases were predicted to be positive. |
| --- |

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