­­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 and the exercises   
as well as the coding portion of the assignment. Now let’s formulate your programming into well-reasoned 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 Unit 5 Assignment page online. Information about the grading rubric is available on any of the unit assignment pages online. Do not hesitate to contact your facilitator if you have any questions about the assignment.*

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 evaluating different models and choosing the best one from a set of candidate models. Different models or different hyperparameters can lead to varying predictions and have different costs and efficiencies. Therefore, it is crucial to find the most suitable model that can accurately represent the given data and make correct predictions with low cost and high performance. |

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 a computing evaluation metrics on examples that were not part of model training. This helps us choose a model that generalizes well to new data, preventing us from picking one that simply memorizes the training set and won’t perform well in real-word data. |

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

|  |
| --- |
| Cross-validation is a resampling method that uses different portions of the data to train and validate the model on different partitions of the data. This ensures a more robust estimate of a model’s generalizability compared to a simple training-validatdion split, as all data points get a chance to be used for both training and testing. |

1. What is the difference between feature engineering and feature selection? What are the benefits of feature selection?

|  |
| --- |
| Feature engineering involves creating new features or transforming existing ones to improve model performance for better analysis. Feature selection focuses on choosing the most relevant features from the existing set. This helps reduce the model complexity, improve interpretability, and increase performance by removing irrelevant or redundant features. |

1. What are the differences among the classification evaluation metrics accuracy, precision, and recall?

|  |
| --- |
| Accuracy measures the overall correct predictions; calculates the percentage of positive and negative examples correctly classified ((TP+TN)/(TP+TN+FP+FN)). Precision focuses on how many positive predictions are truly positive (TP/(TP+FP)). Recall emphasized catching actual positives that were correctly classified as positive (TP/(TP+FN)). The best metric depends on the cost of different errors in your specific problem. |

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