

1. You are given a dataset with different characteristics that are used to predict the brand of a race car. The characteristics include maximum speed, length of the car, maximum height of the car, braking time, average level of noise produced by the engine, and average time to complete one lap. Each characteristic was measured for 150 race cars that were participating in a tournament. There were 7 different brands of race cars represented at this tournament.
  - 1.1. How many features does this dataset have?
  - 1.2. How many data samples does this dataset have?
  - 1.3. How many classes/labels does this dataset have?
  - 1.4. Would you use a regression or classification model for this dataset? Why?
  - 1.5. If you selected a linear model for this dataset, how many weight coefficients ( $w$ ) would there be?
2. After looking at the race car dataset, you discover that the average level of noise produced by the engine dataset is missing 25% of the measurements. How do you treat the missing data and why?
3. When using a classification model, explain how would you quantify the prediction accuracy?
4.
  - 4.1. What is the relationship between bias and model complexity?
  - 4.2. What is the relationship between variance and model complexity?
5. So far in this course, we have covered multiple different linear and non-linear models. How do we determine if the model is a good fit, or if it is under/overfitting?
6. In this course, we discussed how to interpret a confusion matrix for binary classification. Confusion matrices can also be applied to the multiclass case. Given the confusion matrix below, answer the following questions:
  - 6.1. How many false positives does class 3 have?
  - 6.2. How many false negatives does class 3 have?
  - 6.3. What is the precision for class 3?
  - 6.4. What is the recall for class 3?

Confusion matrix									
	0	1	2	3	4	5	6	7	8
True label	37	0	0	0	0	0	0	0	0
0	37	0	0	0	0	0	0	0	0
1	0	39	0	0	0	0	2	0	2
2	0	0	41	3	0	0	0	0	0
3	0	0	1	43	0	0	0	0	1
4	0	0	0	0	38	0	0	0	0
5	0	1	0	0	0	47	0	0	0
6	0	0	0	0	0	0	52	0	0
7	0	1	0	1	1	0	0	45	0
8	0	3	1	0	0	0	0	0	43
9	0	0	0	1	0	1	0	0	1
									44
Predicted label	0	1	2	3	4	5	6	7	8

7. List three benefits of using cross-validation.
8. What is one way to increase the recall of a binary classification?
9. You are given a dataset with fifteen features, ten features are continuous, and the remaining five are categorical. Each categorical feature has four different unique values. After you encode this dataset, how many total columns does the feature matrix have?