1. What is the difference between regression and classification?
Regression is used when the output variable is continuous to predict a numerical value, whereas classification is used when the output variable is categorical in order to assign an input to a specific category or class.

2. What is a confusion table? What does it help us understand about a model's performance?
A confusion table is a table used to evaluate the performance of a classification model by comparing predicted and actual class labels. It contains the true and false positives along with the true and false negatives, and helps us assess precision, recall, accuracy, and the F1 score.

3. What does the SSE quantify about a particular model?
The SSE quantifies the total squared differences between actual values and predicted values in regression models. It represents how well a model fits the data, and a lower SSE correlates to a better fit, while a higher SSE means more errors and a worse fit.

4. What are overfitting and underfitting?
Overfitting occurs when a model learns noise and details in the training data that do not generalize to unseen data. It performs extremely well on training data but poorly on test data. Underfitting occurs when the model is too simple to capture the underlying pattern in the data, leading to high errors in both training and test sets.

5. **Why does splitting the data into training and testing sets, and choosing k by evaluating accuracy or SSE on the test set, improve model performance?**
Splitting data into training and testing sets prevents overfitting by ensuring the model is evaluated on unseen data (so we can't just map values to their answers, i.e. there is a layer of obfuscation that allows for a part of this data to be truly tested), improving generalization. In knn, choosing k based on test accuracy or SSE helps balance bias and variance: small k overfits by capturing noise, while large k underfits by oversmoothing patterns. Cross-validation further refines k by testing performance across multiple data splits, leading to a better model that performs well on new data.

6. **With classification, we can report a class label as a prediction or a probability distribution over class labels. Please explain the strengths and weaknesses of each approach.**
Class Label Prediction is simple, direct, and easy to interpret- however, class label prediction provides no confidence level; all predictions are absolute (you are only classified with one, absolute label).
Probability Distribution Prediction allows confidence in predictions and is useful when making threshold-based decisions, however, this is more complex and may require threshold tuning to make final decisions.