What is overfitting? Explain it in the context of training set model error rate and model complexity.

Overfitting is the tendency to tailor a model to the training data as much as possible. Doing this results in high variance and low bias. It occurs when a classification begins to memorize the training data rather than learning from a general trend and therefore fails to generalize well to unseen data. It increases the model error rate and model complexity. An overfitting model is confusing the noise in the training data with the signal.

Briefly explain why the accuracy metric may not be a good option to assess the performance of a prediction model, when the classes are of different importance. You can use a real-world example to support your argument.

The accuracy metric may not be a good option for assessment as it is simplistic and it gives equal weights to all classes. When classes are of different importance, we may be willing to tolerate greater overall error in return for better identifying the important class. An example would be in screening for cancer. It would be more important to catch all the patients that have cancer in exchange for a lower accuracy.

What are some of the trade-offs that you make when selecting a classification method? In what ways, it is more advantageous to use decision trees as opposed to other methods?

 Some tradeoffs that you make when selecting a classification method are accuracy, speed, robustness, scalability, and interpretability. Decision trees are discriminative. They are good for interpretability as they reveal the pattern for classification and break down the data by making decisions based on a series of questions. Decision trees also make fast predictions even with large training sets, regarding computational complexity.

Briefly discuss the idea behind the Support Vector Machines approach for classification. You can provide a high level overview of the steps in SVM to facilitate the discussion.

Support Vector Machines consist of efficient algorithms available to find the global minima of the objective function. The objective of the SVM is to maximize the margin, the distance between the decision boundary and the training samples closest to the boundary. They are robust to noise and have high computational complexity. They can be linear or transformed into higher dimensional space.

What is the difference between the one-vs-one and one-vs-rest approaches within the context of multiclass classification? Discuss each approach briefly.

One-vs-one and one-vs-rest are techniques that allow us to extend a binary classifier to multiclass problems.

In OvO, if there are N class, you would need to train  classifiers. When it comes time to make a prediction, the class which received the most votes is selected. OvO is better for algorithms, like SVM, that scale poorly with the size of the training set.

In OvR, one classifier is trained per class (N classes = N classifiers). For each classifier, the class is fitted against all the other classes.and you select the class based on the decision score. OvR is typically the preferred technique.