a) What is Overfitting, and How Can You Avoid It?

Overfiiting is when the model describes random error and noise rather than the underlying relation.

You can avoid it by adding more training sets, remove irrelevant features, and prune leaf nodes on decision trees.

b) What is 'training Set' and 'test Set' in a Machine Learning Model? How
Much Data Will You Allocate for Your Training, Validation, and Test Sets?

The training set is the set of data used to train the ML model. The test set the data set used to test the trained ML model against.

Training: 80%, Validation: 10%, Test: 10%

c) How Do You Handle Missing or Corrupted Data in a Dataset?

I would fill in those with the average/mean of the all of values in its corresponding column if the value is numeric. If the value is categorical, I would fill it with the value the appeared the most, mode.

d) Explain the AUC score and why it can be used to evaluate Machine
Learning models?

AUC score measures the area under the tradeoff curve between recall and precision.

It can be used to evaluate ML model by measuring the performance of distinguishing between the positive and negative classes.

e) When is it necessary to have higher precision rather than high recall?

When we have imbalanced class and we need high true positives.

f) What is the difference between supervised and unsupervised learning?

Supervised learning uses labeled data.

Unsupervised learning uses unlabeled data.

What is the utility of unsupervised learning?

It is used to discover interesting hidden patterns or correlations in data.

g) How Will You Know Which Machine Learning Algorithm to Choose for
Your Classification Problem?

By looking at what the problem is asking for, what type of data and values I am working with. Is the data labeled or unlabeled. Are the values discrete or continuous?

h) When Will You Use Classification over Regression?

When is desired predicted output is label or something that classify or identity something rather than a real-vauled number.

i) What is 'ensemble learning'?

Using multiple models rather than one to predict the output. Bagging or random forest.

j) Why do we need backpropagation in Neural Network?

To train feedforward neural networks in areas such as deep learning.

k) Why did Multilayer Perceptron evolve over Single Layer Perceptron?

To work with data that is not linear. Single Layer Perceptron can only work with linear data.