1 What Does Auto Encoder Do?

Given inputs it exactly reproduces that input as its output.

Can have slight round off error such as 0.7 vs 0.6999.

In the hidden layer there are new compressed features, performs feature reduction.

2 Given 2 class problem where $P(C_1|x) = 0.8$ what does $P(C_2|x) =$?

$P(C_2|x) = 1 - 0.8 = 0.2$

3 Bayes Rule for C1 and C2, what does $P(C_1|x) =$

$P(C_1|x) = \frac{P(x|C_1)*P(C_1)}{P(x)} = \frac{P(x|C_1)*P(C_1)}{P(x|C_1)*P(C_1)+P(x|C_2)*P(C_2)}$

4 1M images of 5 classes, what classifier to use and why?

CNN would be the best choice as it is commonly used for images.

CNNs can handle multi-class problems well whereas some techniques can only handle 2 class problems.

Feature extraction is done for you by the convolution filters.

5 100 numeric data points with 1k feature and 2 classes, what classifier to use and why?

SVM, KNN, Naive Bayes with Logs.

6 How to achieve Non-Linear Solutions?

Using SVM w/ kernel function which allows you to transform the decision space.

Deep Nueral Netowrk, or NN with Signle Hidden Layer

7 10 Fold CV has Type 1 error because?

The training sets contain overlapping data.

8 What does K-Means do?

Creates cluster centers

Unsupervised approach to learning

Given a new example it can assign it to an existing cluster, or it can assign to a cluster and then shift the clusters based on the new example.

9 What is the purpose of ϵ in Support vector regression?

Accepts and ignores small errors up to epsilon and does not count them as being incorrect.

10 How are points on ROC curve computed?

They ROC curve is a plot of False Positive vs True Positive and the points are computed by changing the decision threshold.

11 Techniques to compare 5 classifiers on 10 data sets?

You can use one of the following approaches, Anova, Kruskal Wallis test, or Wilcoxon sum of positive ranks test.

12 What is the purpose of convolutional layers in CNN?

They extract features from the data (images), possibly achieve feature reduction.

13 Machine Learning for Regression?

Search for model that fits the data.

14 What is special about Non-Parametric Models and give 3 examples.

Non-Parametric models do not make any assumptions about the data or their distribution.

Examples: KNN, Histogram, Local Regression

15 Compare Decision Trees vs Ripper(JRIP) or CNN vs Nearest Neighbor.

Decision Trees for classification use a test at each node to improve some impurity measure which defines the goodness of the split. A split is pure if after the split each branch contains a single class of data. As well the decision tree can be converted to a set of rules like ripper but is not setup that way from the start. Finally if a set of rules is extracted they will all have the root node test in each rule. RIPPER on the other hand is a rule based approach that adds rules to maximize the information gain ie the number of instances covered and correctly sorted by each rule.

A convolutional neural network uses convolution filters to automatically perform feature extraction and is able to classify data into multiple classes. Generally a CNN is slow to train but fast for testing. As well a CNN requires a lot of data to perform well. A nearest neighbor approach tries to cluster data based on how similar it is to other points in the dataset.