1. In a binary classification (two classes: positive, negative) problem, precision is defined as true positive / (true positive + false positive), whereas recall is defined as true positive / (true positive + false negative). Explain why precision alone (in absence of recall) may not reflect the actual effectiveness of a binary classifier.

Answer: Just illustrate the case of a classifier that labels most instances as negative (a conservative classifier). This classifier will have a high precision, but low recall.

2. In ensemble methods, k classifiers are generated by learning tuples sampled from the dataset D. This seems like wasting a lot of training data as we could utilize the whole dataset D for learning each of the k classifiers. What's the fallacy (error) in this claim?

Answer: If we use the entire training dataset, all classifiers will learn the same concept. Therefore it will just be a single classifier, instead of a collection of them.

3. Naive Bayesian prediction requires each conditional probability be non-zero. Explain how this requirement can be an issue during document classification using words in the documents as features.

Answer: When new words appear in the testing cases. Those words will have zero probability, and will require correction.

4. Report three techniques (one sentence for each) to resolve the class imbalance problem.

Answer: Sampling (to balance out the ratio), Bagging, Boosting.

5. Explain how bagging can be advantageous over boosting for combining classifiers. Hint: Consider noisy data.

Answer: Boosting would propagate noise in case of noisy data, whereas bagging would reduce the noise propagation chances.

6. A linear classifier makes a classification decision based on the value of a linear combination of the features. A decision tree is a non-linear classifier. Provide explanation/argument to support this claim.

Answer: Consider a simple decision tree with just 2 attributes:

If (A = 0 and B = 1) then +
If (A = 0 and B = 0) then -
If (A = 1 and B = 0) then +
If (A = 1 and B = 1) then -

Now, the decision boundary of this classifier is If ((A = 0 and B = 1) or (A = 1 and B = 0)), then +

This decision boundary can not be represented with a linear classifier.