1. To calculate the information gain for age, we must first calculate the Info on the whole dataset. The calculation is as follows:

If I were to split the data set into smaller portions, I would choose the age attribute. The information gain is greater for age than it is for salary and we should always choose the attribute with the most information gain.

1. (a) Sensitivity: Suited best when trying to minimize the false negatives during classification. For example, when attempting to classify if a patient has a certain deadly disease, it is better to use the 1-Sensitivity. This measures the #FN/#P and this is something that we want to minimize.   
     
   (b) Specificity: This can be used for classifying spam email. It would be bad to classify an email that is spam as not spam, but it would be even worse if we classify an important email as spam. In this case, looking at the specificity would be better than looking at sensitivity or accuracy. Having a high specificity rate means that the rate of false positives is low, meaning the chances of having an important email classified as spam is low.   
     
   (c) Accuracy: More appropriately used when classification of false positives and false negatives are weighted the same. For example, with the classification of the glass dataset, false positives and false negatives are weighted equally and thus both are errors equally as bad. In this case, it makes sense to just simply use the accuracy.
2. Based on the tests conducted for over 100 different folds, the classifier I would prefer is classifier B. In the methods for performance, there are two values we want to minimize; one is the mean error. If the mean error is low, it means that on average, the classifier makes less mistakes when classifying new data. The second value that we want to minimize is the variance. It is not desirable for the variance to be high as it means that the classifier’s performance is not consistent with the dataset used for training.   
     
   However, I would like to see the variance of classifier B on the first dataset and the variance of classifier A on the second dataset. This would help better choose which classifier to use as it provides more evidence.
3. The rules are not mutually exclusive. There is a range (41 - 50) in age that is satisfied by both rules and causes an overlap for tuples that have age in that range. For them to be mutually exclusive, the rules should not have any overlap.  
   These rules are, however, exhaustive as they cover every possible age; every record would be covered by at least one of the rules. This stands assuming every tuple has the age attribute non-empty, otherwise the rules would not be exhaustive.