**Q:** When do we need to perform stratified sampling? / How do you handle imbalanced data sets?

Stratified sampling preserves the percentage of samples for each class, especially useful in imbalanced data sets to mitigate sampling bias when selecting your training and test data. However, this is not the only way – or necessarily the best way – to handle imbalanced data. For example, you may not have enough data points for each stratum, e.g. predicting default rate for a credit card, which is less than 1%. It is important to have a sufficient number of instances in your data sets for each stratum, or else the estimate of the stratum's importance may be biased. This means that you shouldn't have too much strata, and each stratum should be large enough. For these, over-/under-sampling or a technique called SMOTE may be more appropriate. Ensemble methods are actually a common method used to generalize a model on imbalanced data. See here for a pretty good intro into the different ways to handle imbalance: https://www.kdnuggets.com/2017/06/7-techniques-handle-imbalanced-data.html

An important thing to point out, regardless of how you derive the training and test data, is that a few of you used an imbalanced data set but only calculated the accuracy. If you are keeping the imbalance, it's important to look at other metrics (precision, recall) and the confusion matrix.

**Q:** Isn't the first question similar to soft voting strategy? (Q1: When using a bagging/pasting ensemble and applying it to a new, test instance, the ensemble classifier aggregates the predictions of all predictors and estimates the statistical mode (i.e., most frequent prediction). Is this similar to the hard or the soft voting strategy?)

No, and almost all of you got this correct. With hard voting you take the majority of votes for a specific class among the classifiers; with soft voting you take each classifier's confidence into account and weigh their decisions accordingly. If the base estimators can estimate class probabilities, then it is soft voting.

**Q:** What does the learning rate do in boosting?

Inspection of the graph may suggest the answer: learning rate is responsible for the contribution of each tree regressor to the ensemble. With a higher rate you see more irregular fitting; with a lower learning rate you see a smoother fitting line. However, lower learning rate implies slower learning ensemble, so you need more trees in the ensemble (cf. 3 estimators in the higher learning rate case, graph on the left, and 200 estimators in the lower learning rate case, graph on the right).