Minute May 21

Assign group roles:
We didn't do this in class, so I'll just assign them here. Let me know if that works for you.

1. Recorder and Monitor — Wanli Zhang

2. Coordinator — Martin Guyer

3. Checker — Shangjia Dong

Today's Discussion

1. **Question of interest**: Classify whether a mushroom is poisonous or edible based on their physical attributes. Link: https://archive.ics.uci.edu/ml/datasets/Mushroom

2. Martin looked at the data, and discovered that only one variable, called `stalk-root`, contains missing values. The data description page confirms this: "2480 of them (denoted by "?"), all for attribute 11".

3. Dealing with missing values:

   —Imputation?

   —Ignore variable "stalk-root"?

4. Ideas on how to proceed:

   —Separate data into training and testing sets.

   —Construct classification criterion using the training set, and then verify using the testing set.

   —Use misclassification rate as a measure of goodness.

5. To be done by Friday:

   • Everybody: Look for available classification methods, and see which ones make the most sense to you, and potentially easy to implement. Also look for R packages available corresponding to your preferred method.

   • Martin: upload a copy of original data set to repo.

   • Wanli: upload group contract to repo.