Automated Leaf Classification

Zach Stecher

Due: 9/22/16

Project Proposal

After searching on Kaggle, I've decided to participate in the leaf classifaction competition. I chose this project mainly because the possible applications of automated plant identification through leaf classification are far-reaching and include important areas like medical research and food production analysis-areas human society will need to get much better at in the coming years if we don't want to end up like the dinosaurs.

According to the Kaggle description of evaluation, to goal for this competition is to minimize Multi-class LogLoss. The submission file will be a set of computed probabilities for each image that lists the probability of a given image being each class of plant. The probabilities do not need to sum to 1, but must be within 0 and 1.

For this project, code is already available to achieve a confidence level of around 96 percent using the random forest algorithm. My goal is to find a way to increase this confidence level to at least 98 percent, and hopefully 99 percent or more. After some light research, it seems like the neural network direction has been promising. I plan to replicate some of the experiments run by other participants and self-evaluating the logloss using a script provided on Kaggle to get an idea for how well existing classification algorithms solve this problem before making any modifications.

Once I've selected a classifier It's simply a matter of finding a way to increase its accuracy to the point of satisfaction. Ultimately, I would like to make a submission to this competition and finish somewhere in the top twenty-five percent.