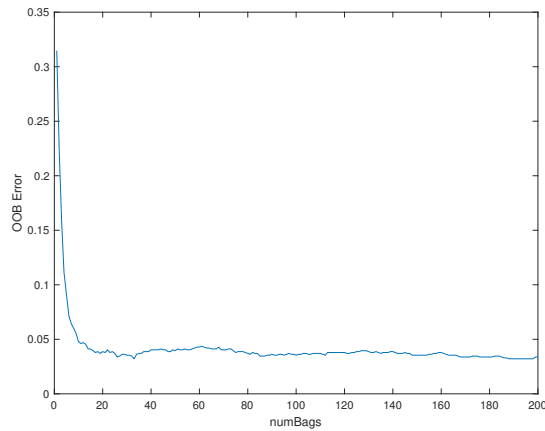


CSE 417T: Homework 4 Solution Sketches

April 1, 2022

Note: These are not intended to be comprehensive, just to help you see what the answers should be.

- (a) The figure of OOB error vs the number of bags would look like the following (there are randomness involved, so as long as the “shape” is similar, it’s fine.)



- (b) The OOB errors for the bagging implementation are reported below:

	Bagged Trees
One-vs-Three	0.0018
Three-vs-Five	0.0313

The test errors for decision trees and the bagging implementation are reported below:

	Decision Trees	Bagged Trees
One-vs-Three	0.0140	0.0116
Three-vs-Five	0.1258	0.0798

As the number of ensembles goes large, the OOB error decreases. This fits the analysis that the variances for the bagged tree decreases. Three-vs-Five is a harder question. OOB error serves as a fine estimation for out-of-sample error.

- (a) The root node of the decision tree after running ID3 is “Stripe”. Let $H(p) = p \log(1/p) + (1 - p) \log(1/(1 - p))$ be the entropy of binary signals in which one of the signals has probability p in the distribution. We can calculate the information gain for choosing each attribute.

The entropy of the data before splitting is $\frac{2}{5} \log(\frac{5}{2}) + \frac{3}{5} \log(\frac{5}{3}) \approx 0.971$.

The average entropy after splitting using “Color” is $\frac{1}{5} \times 0 + \frac{4}{5} [\frac{2}{4} \log(\frac{4}{2}) + \frac{2}{4} \log(\frac{4}{2})] = 0.8$.

The average entropy after splitting using “Stripes” is $\frac{2}{5} \times 0 + \frac{3}{5} [\frac{1}{3} \log(\frac{3}{1}) + \frac{2}{3} \log(\frac{3}{2})] \approx 0.551$.

The average entropy after splitting using “Texture” is $\frac{2}{5} \left[\frac{1}{2} \log \left(\frac{2}{1} \right) + \frac{1}{2} \log \left(\frac{2}{1} \right) \right] + \frac{3}{5} \left[\frac{1}{3} \log \left(\frac{3}{1} \right) + \frac{2}{3} \log \left(\frac{3}{2} \right) \right] \approx 0.951$.

Given the above values, we have

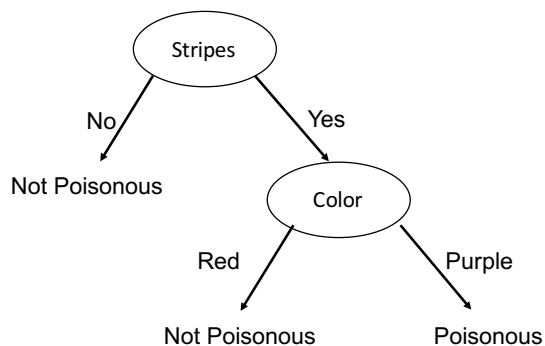
$$Gain(D, \text{Color}) \approx 0.971 - 0.8 = 0.171$$

$$Gain(D, \text{Stripes}) \approx 0.971 - 0.551 = 0.420$$

$$Gain(D, \text{Texture}) \approx 0.971 - 0.951 = 0.020$$

Therefore, splitting using “Stripes” leads to the maximum information gain.

(b) The final decision tree looks like the following:



3. The updates will make positive and negative examples have equal weight after one round. The result suggests that this learner is too weak and won't be able to return anything useful after the first round.
4. Mentioning one or both of the following (or other valid reasonings) is okay:
 - The output of bagged linear regressions is still a linear regression (the sum of linear functions is still a linear function).
 - Bagging helps reduce variance and is therefore more helpful for low-bias high-variance weak learners. Linear regression is a low-variance weak learner and bagging might not help much.