Ensembling

Handout 4 of Introduction to Machine Learning

January 2020

We'll assign each team a number. Your team number determines which ten observations from the testing set you'll predict (Round 1) and vote on (Round 2).

Team #: _____

Round 1: Predict

Pick a random seed to use: _____
Shh. Don't tell any of your teammates.

Use the `get_boot_votes()` function provided to:

- build a single classification tree,
- train it with one bootstrapped sample, and
- predict your 10 observations from the testing set.

Your output will look like this. Use it to fill in the table on the back of this page.

```
get_boot_votes(seed = 0, team = 2020)
```

```
## # A tibble: 10 x 3
##      obs truth      estimate
##    <int> <fct>      <fct>
##  1    68 Remote     Not remote
##  2    69 Remote     Remote
##  3    70 Remote     Remote
##  4    71 Remote     Remote
##  5    72 Remote     Not remote
##  6    68 Not remote Not remote
##  7    69 Not remote Not remote
##  8    70 Not remote Not remote
##  9    71 Not remote Not remote
## 10    72 Not remote Not remote
```

What was your bootstrapped tree's overall accuracy?

(over)

			Round 1: Predict		Round 2: Vote!		
	Obs	**Truth**	**Guess**	**Score**	**Votes Remote**	**Votes Not**	**Majority Vote**
1		Remote		0 / 1			**Remote** / Not
2		Remote		0 / 1			**Remote** / Not
3		Remote		0 / 1			**Remote** / Not
4		Remote		0 / 1			**Remote** / Not
5		Remote		0 / 1			**Remote** / Not
6		Not Remote		0 / 1			Remote / **Not**
7		Not Remote		0 / 1			Remote / **Not**
8		Not Remote		0 / 1			Remote / **Not**
9		Not Remote		0 / 1			Remote / **Not**
10		Not Remote		0 / 1			Remote / **Not**
			Total	/ 10			

Round 2: Vote!

Now, as a team, form a voting committee. You have your tree's votes for those 10 observations already filled in the table above. Your job now is to tally up the number of votes for `Remote / Not Remote` across all team members' trees. Fill in the second half of the table with your team.

What was your team's overall accuracy, across each members' bootstrapped trees?