Nicholas Fang

Assignment4

Step 1: Using decision tree on this dataset, how accurately you can tell their birth year from their hand postures and thumb styles. How does it affect the evaluation when you include the region while training the model?

Played around with changing the sampling size, ran the decision through training test splits between 70/30, 75/25, and 80/20. The results were similar regardless of the split in this iteration. Will be reporting on using a 75/25 test split. Taking the average accuracy score of 10 runs when predicting brthYr still yielded <1%, with multiple instances of 0% accuracy and the highest yield of ~4.5%.

Adding the regions parameter to the model get a slightly better prediction. On average about ~4.2% when predicting brthYr. There were still a couple of predictions which yield 0% accuracy, with a high of ~8.3% accuracy.

Step 2: Now do the same using random forest (in both of the above cases) and report the difference. Make sure to use appropriate training-testing parameters for your evaluation.

You should also run the algorithms multiple times, measure various accuracies, and report the average (and perhaps the range).

Slightly better results with using a random forest over decision tree. There were less instances of 0% accuracy, but 1 or 2 still popped up when running the program various times. The results for random forest with and without adding the regions parameter yielded about an average accuracy of 4.2%, with a range between 0-8.3%.

Looked into changing the criterion for information gain by entropy into the decision tree process, yielded the same results if not more errors because of more instances of 0% accuracy.

Wondering if I did something wrong or if the data was just not that great a data set to use decision tree or random forest on.