Trevor Truog

Machine Learning Homework 2

Chart 1

Chart 2

1. The accuracy of the random classifier is about .506
2. With infinite depth, the accuracy is about 0.76
3. The best depth limit to use is depth 5 (~.78 accuracy)
4. Yes we do see overfitting in the data set. With increased depth we have increased accuracy (see Chart 1). Also, looking at Chart 2, we see that the tree with depth 5 has much better accuracy when it trains on larger portions of the overall data. This indicates overfitting.
5. Looking at Chart 2, we see that the larger portion of data we allocate as training data, the better we predict the label in general.
6. The training data experiment tells Sex and First Class are the best features at predicting the label, and that we predict the label correctly about 78% of the time.

Random classifier accuracy: 0.5061295971978984

Accuracy Based on Depth:

Depth:

0: 0.593216783216783

1:0.7716083916083913

2: 0.7728671328671328

3: 0.7744055944055944

4: 0.774755244755245

5: 0.7804895104895102

6: 0.7623776223776219

7: 0.7661538461538457

8: 0.7631468531468529

9: 0.7616783216783216

10: 0.7658041958041955

BEST DEPTH = 6

.05: 0.7201325478645062

.10: 0.7436236391912909

.15: 0.7566392092257004

.20: 0.7598601398601401

.25: 0.7640485074626863

.30: 0.7644999999999998

.35: 0.7654408602150538

.40: 0.772750582750583

.45: 0.7727226463104323

.50: 0.7756862745098035

.55: 0.7779503105590062

.60: 0.7767132867132864

.65: 0.7746400000000004

.70: 0.7793953488372093

.75: 0.7801117318435757

.80: 0.7868518518518521

.85: 0.7775000000000004

.90: 0.7806944444444446

.95: 0.7913888888888887