

HW6

Problem 1

a.)

Overfitting is the well-known statistical notion of significance.

b.)

One way to prune a tree would be to set a threshold for the entropy, so if a possible value for an attribute has a low enough entropy, then don't split, even if the entropy is greater than 0.

c.)

Pruning will make sure that each of the leaves of the decision tree will not be as close to homogeneity and will cover more examples.

d.)

One alternative to pruning is to limit the set of tests available. Another alternative is to limit the minimum number of examples used to select a split.

Problem 2

$(L-1) / (K-1)$ rounded up to the nearest integer.

Problem 3

a.)

Entropy = 1

b.)

Male player strength: Information gained = 0

Female player strength: Information gained = 0

Player handedness: Information gained = 0.0905

The root will be same handedness.

c.)

The height of the tree is 3 because there are 3 parameters that decide the outcome.

d.)

Error = 0.325

e.)