Machine Learning Decision Tree Home Work. Due: Beginning of class Do not staple or attach pages.

Suppose we have 40 training instances and we have two splits that we can choose between.

Root Node [10, 10, 10, 10]

Split choice 1: Left [0, 0, 10, 10], Right[10, 10, 0, 0]

Split choice 2: Left [10, 0, 0, 0] , Right[0, 10, 10, 10]

1. Compute the GINI impurity for all 5 nodes.

Root node: 1-[(¼)2 +(¼)2 +(¼)2 +(¼)2 ] = .75

Split1left: 1-[(1/2)2+(1/2)2] = 1-.5 = .5

Split1right: 1-[(1/2)2+(1/2)2] = 1-.5 = .5

Split2left: 1-[(1)2] = 0

Split2right: 1-[(1/3)2+(1/3)2+(1/3)2] = 1-[.111+.111+.111] = .6666666

1. Compute the weighted GINI for the split choices

Split1: (1/2).5 + (1/2).5= .5

Split2: (1/4)(0)+(3/4)(.666) = .4995 or if u round the .66 to .67 and u get .5000025

1. Which split should you make?

Split2 because it has a lower gini which makes it the more pure split

D) Compute the entropy for all 5 nodes.

Root node: -e(s) = (1/4)log(1/4)+(1/4)log(1/4)+(1/4)log(1/4)+(1/4)log(1/4) = -.5+-.5+-.5+-.5 = -2 = 2

Split1left: (1/2)log(1/2) + (1/2)log(1/2) = (-0.5 +-0.5) = 1

Split1right: (1/2)log(1/2) + (1/2)log(1/2) =(-0.5 +-0.5) = 1

Split2left: -e(s) = 1log1 + 0log0 + 0log0 + 0log0 = 0

Split2right: -e(s) = 0log0 +(1/3)log(1/3)+ (1/3)log(1/3)+ (1/3)log(1/3) = (-1.5849)(1/3)+ -.5849(1/3+ -1.5849(1/3) = (-0.52824717)+( -0.52824717)+( -0.52824717) = 1.58474151

1. Compute the weighted entropy for the split choices.

Split1: (1/2)( 1)+(1/2)( 1) = 1

Split2: (1/4)(0)+(3/4)( 1.58474151) = 1.1885561325

1. Will you make the same split?

No. we will pick split 1 here because the weighted entropy of split1 is lower than that of split2

Bonus: Can you find a split choice where GINI and entropy select different splits?