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.

1-[(¼)2 +(¼)2 +(¼)2 +(¼)2 ] = .75

1. Compute the weighted GINI for the split choices
   1. = 0 (left)

1-(1/2) = .5 (right)

1. Which split should you make?

left

D) Compute the entropy for all 5 nodes.

-e(s) = (1/4)log(1/4)+(1/4)log(1/4)+(1/4)log(1/4)+(1/4)log(1/4)

-2+-2+-2+-2 = -8 = 8

E) Compute the weighted entropy for the split choices.

-e(s) = 1log1 + 0log0 + 0log0 + 0log0 = 0 (left)

-e(s) = 0log0 +(1/3)log(1/3)+ (1/3)log(1/3)+ (1/3)log(1/3) = -4.75488

F) Will you make the same split?

Will split left again so yes

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