CS 458
Grant Davis
10/06/22

Homework 2

HW2-1:

a) Yes = 4, No = 6
   Entropy = (4/10 * log4/10 + 6/10 * log 6/10)
b) Body temperature: 0.9708 - 0.36095 = 0.60985
   Give Birth: 0.60985
c) Give Birth:
   Classification Error Rate:

Give Birth	Yes	No
Yes	4	0
No	1	5

Error_yes = 1 - max(4/5 , 1/5) = 0.2
Error_no = 1 - max (0/5, 5/5) = 0
Weighted Classification Error = 5/10 * 0.2 + 5/10 * 0 = 0.1
Original Classification Error = 1 - max (4/10, 6/10) =0.4
Classification Error = Error_Origin - Weighted Classification Error = 0.4- 0.1 = 0.3

Four Legged:
Classification Error Rate:

Four Legged	Yes	No
Yes	2	2
No	2	4

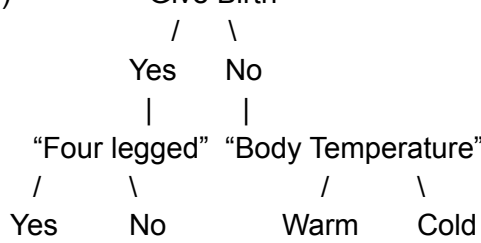
Error_yes = 1 - max (2/4, 2/4) = 0.5
Error_no = 1 - max (2/6, 4/6) = 0.34
Weighted Classification Error = 4/10 * 0.5 + 6/10 * 0.34 = 0.398
Original Classification Error = 1 - max (4/10, 6/10) =0.4
Classification Error = Error_Origin - Weighted Classification Error = 0.4- 0.398 = 0.002

Give Birth is the best option for classification error

d) Parents Gini Index = 1 - (4/10)^2 - (6/10)^2 = 0.48
Give Birth:
Gini_Yes = 1 - (4/5)^2 - (1/5)^2 = 0.32
Gini_No = 1 - (0/5)^2 - (5/5)^2 = 0.45
Weighted Gini Index = 5/10 * 0.32 + 5/10 * 0 = 0.16
Gini Index = Gini Parents - Gini Split = 0.48 - 0.16 = 0.32
Four Legged:
Gini_Yes = 1 - (2/4)^2 - (2/4)^2 = 0.5
Gini_No = 1 - (2/6)^2 - (4/6)^2 = 045
Weighted Gini Index = 4/10 * 0.5 + 6/10 * 0.45 = 0.4664
Gini Index = Gini Parents - Gini Split = 0.48 - 0.46 = 0.0136

Give Birth is the best option for Gini Index

e)



HW2-2. Model Overfitting

a) Optimistic error = T1 = (2+1+1+3+2+1+1+3+1)/73= 16/73 = 0.2197
T2 = (1+2+6+5+3+3)/73= 20/73= 0.27397
b) Pessimistic error (0.5 penalty) = T1 = 0.21917+13*0.5/73 = 0.090
T2 = 0.27396+6*0.5/73 = 0.44
(0.75 penalty) = T1 = 0.21917+13*0.75/73=0.135
T2 = 0.27396+6*0.75/73=0.67
(1 penalty) = T1 = 0.21917+13*1/73=0.18
T2 = 0.27396+6*173=0.89
c) Tree 1 would be based on the generalization error the tree with the lower value is preferred.
d) Tree 2 would be based on Ocala razor choosing the simpler tree when error is similar.