Question 1 - ID3 Algorithm

S = [9, 5]

Entropy = [9/14(+), 5/14(-)] =

- 9/14(log2)(9/14) - 5/14(log2)(5/14)

Info of all training samples = 0.94



Info[2, 3] = Info[4, 0] = Info[3,2] =

Entropy[⅖(+), ⅗(-)] = entropy[1(+), 0(-)] = Entropy[⅗(+), ⅖(-)] =

- ⅖ (log2)(⅖) - ⅗(log2)(⅗) -log2(1) - 0 - ⅗(log2)(⅗) - ⅖(log2)(⅖)

= 0.97 bits = 0 = 0.97 bits

Average weighted info of subtree = (0.97 \* 5/14) + 0 + (0.97 \* 5/14) = 0.69

Gain (S, Outlook) = 0.94 - 0.69 = 0.29



Info[2, 2] = Info[4, 2] = Info[3,1] =

Entropy[2/4(+), 2/4(-)] = entropy[4/6(+), 2/6(-)] = Entropy[3/4(+), 1/4(-)] =

-1/2 (log2)(1/2) -1/2(log2)(1/2) -⅔(log2)(⅔) - ⅓(log2)(⅓) = -3/4(log2)(3/4) - 1/4(log2)(1/4)

= 1 bits = 0.91 bits = 0.81 bits

Average weighted info of subtree = (1 \* 4/14) + (0.91 \* 6/14) + (0.81 \* 3/14) = 0.85

Gain (S, Outlook) = 0.94 - 0.85 = 0.09



Info[3, 4] = Info[6, 1] =

Entropy[3/7(+), 4/7(-)] = entropy[6/7(+), 1/7(-)] =

-3/7 (log2)(3/7) -4/7(log2)(4/7) - 6/7(log2)(6/7) - 1/7(log2)(1/7) =

= 0.99 bits = 0.59 bits

Average weighted info of subtree = (0.99 \* 7/14) + (0.59 \* 7/14) = 0.79

Gain (S, Outlook) = 0.94 - 0.79 = 0.15



Info[6, 2] = Info[3, 3] =

Entropy[6/8(+), 2/8(-)] = entropy[3/6(+), 3/6(-)] =

-3/4 (log2)(3/4) -1/4(log2)(1/4) - 1/2(log2)(1/2) - 1/2(log2)(1/2) =

= 0.81 bits = 1 bits

Average weighted info of subtree = (0.81 \* 8/14) + (1 \* 6/14) = 0.93

Gain (S, Outlook) = 0.94 - 0.89 = 0.05

2. No, because the independent variable temperature does not impact the value of the dependent variable. It is therefore, overfitting to include them as part of the calculation for the target value.

3. Training Set Classification Accuracy D1-D7

S = [4, 3]

Entropy = [4/7(+), 3/7(-)] =

- 4/7(log2)(4/7) - 3/7(log2)(3/7)

Info of all training samples = 0.99



Info[0, 2] = 0 Info[2, 0] = 0 Info[2,1] =

Entropy[2/3(+), 1/3(-)] =

- 2/3(log2)(2/3) - 1/3(log2)(1/3)

= 0.91 bits

Average weighted info of subtree = 0 + 0 + (0.91 \* 3/7) = 0.39

Gain (S, Outlook) = 0.99 - 0.39 = 0.6



Info[1, 2] = Info[1, 0] = 0 Info[2,1] =

Entropy[1/3(+), 2/3(-)] = Entropy[2/3(+), 1/3(-)] =

-1/3 (log2)(1/3) -2/3(log2)(2/3) -2/3(log2)(2/3) - 1/3(log2)(1/3)

= 0.91 bits = 0.91 bits

Average weighted info of subtree = (0.91 \* 3/7) + 0 + (0.91 \* 3/7) = 0.78

Gain (S, Outlook) = 0.99 - 0.78 = 0.21



Info[2, 2] = Info[2,1] =

Entropy[2/4(+), 2/4(-)] = Entropy[2/3(+), 1/3(-)] =

-1/2 (log2)(1/2) -1/2(log2)(1/2) -2/3(log2)(2/3) - 1/3(log2)(1/3)

= 1 bits = 0.91 bits

Average weighted info of subtree = (1 \* 4/7) + (0.91 \* 3/7) = 0.96

Gain (S, Outlook) = 0.99 - 0.96 = 0.03



Info[3, 1] = Info[1,2] =

Entropy[3/4(+), 1/4(-)] = entropy[1/3(+), 2/3(-)] =

-3/4 (log2)(3/4) -1/4(log2)(1/4) -2/3(log2)(2/3) - 1/3(log2)(1/3)

= 0.81 bits = 0.91 bits

Average weighted info of subtree = (0.81 \* 4/7) + (0.91 \* 3/7) = 0.85

Gain (S, Outlook) = 0.94 - 0.85 = 0.09

Testing Set Classification Accuracy D8-D14

Info[5, 2]

Entropy[5/7(+), 2/7(-)] =

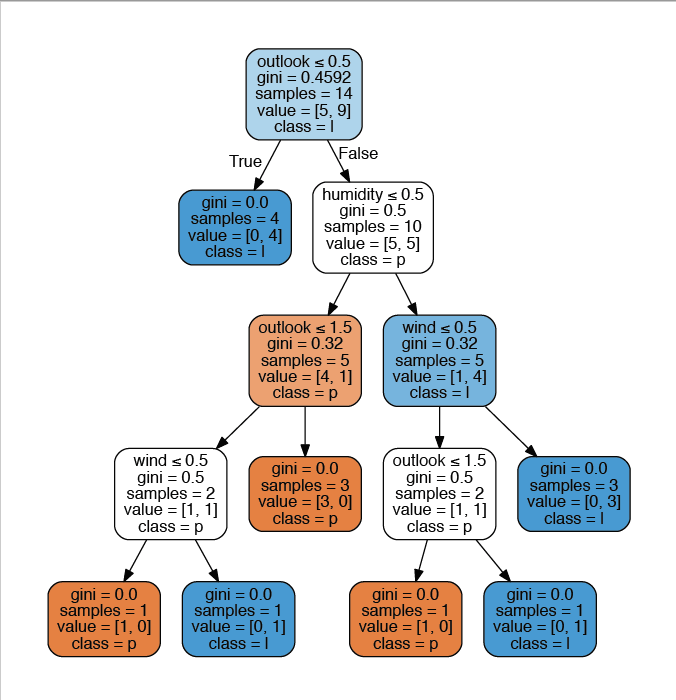
- 5/7(log2)(5/7) - 2/7(log2)(2/7)

= 0.86 bit

4. We can do a post pruning strategy to removing categories or independent variables that don’t impact the value of the dependent variable.

* Once the decision tree classifies the training set, we use a validation set that has different independent variables removed to analyze its impact on the dependent variable.
* If an independent variable is removed or added and the value of the dependent variable remains unchanged or improved then we can decide to remove the independent variable.
* Through this way we can also estimate the error amount and more accurately see the effects of removing/adding variables without the chance of risking overfitting.
* Another strategy is to add a minimum number of samples for a leave or setting the depth maximum level of the tree to avoid the tree from becoming too complex.

5.



Question 2 Neural Network

Activation of unit based on the step-function

y = τ(v) = 1 if v ≥ 0

y = τ(v) = 0 otherwise

P1 X = {1, 0, 0}, W = {2, -4, 1}

τ(v) = τ(x1 \* w1 + x2 \* w2 + x3 \* w3)

τ(v) = τ(1 \* 2 + 0 \* -4 + 0 \* 1)

τ(v) = τ(2)

y = 1 since v is 2 and it is greater than 0

P2 X = {0, 1, 1}, W = {2, -4, 1}

τ(v) = τ(x1 \* w1 + x2 \* w2 + x3 \* w3)

τ(v) = τ(0 \* 2 + 1 \* -4 + 1 \* 1)

τ(v) = τ(-3)

y = 0 since v is -3 and it is less than 0

P3 X = {1, 0, 1}, W = {2, -4, 1}

τ(v) = τ(x1 \* w1 + x2 \* w2 + x3 \* w3)

τ(v) = τ(1 \* 2 + 0 \* -4 + 1 \* 1)

τ(v) = τ(3)

y is 1 since v is 3 and it is greater than 0

P4 X = {1, 1, 1}, W = {2, -4, 1}

τ(v) = τ(x1 \* w1 + x2 \* w2 + x3 \* w3)

τ(v) = τ(1 \* 2 + 1 \* -4 + 1 \* 1)

τ(v) = τ(-1)

y is 0 since v is -1 and it is less than 0

Question 3 - MatLab Iris Data Neural Networks

After going through the Iris Data and training a neural network based on the training data we can see the how well the model matches the testing data simulated values with the actual target values. It had a mean squared error of 1.43 which signifies they have a minimized error rate showing high accuracy in regression. Additionally, the line graph shows both the actual values and prediction values correlating in visual pattern.

