Alan Ou
301320642

Decision Tree Learning

1) **Entropies: Hun = F:**

Act = T: - $(1/4)\log_2(1/4)$ - $(3/4)\log_2(3/4)$ = 0.81
Act = F: - $(0/1)\log_2(0/1)$ - $(1/1)\log_2(1/1)$ = 0

Bar = T: - $(1/3)\log_2(1/3)$ - $(2/3)\log_2(2/3)$ = 0.92
Bar = F: - $(0/2)\log_2(0/2)$ - $(2/2)\log_2(2/2)$ = 0

Fri = T: - $(1/3)\log_2(1/3)$ - $(2/3)\log_2(2/3)$ = 0.92
Fri = F: - $(0/2)\log_2(0/2)$ - $(2/2)\log_2(2/2)$ = 0

Pat = Some: - $(1/1)\log_2(1/1)$ - $(0/1)\log_2(0/1)$ = 0
Pat = Full: - $(0/2)\log_2(0/2)$ - $(2/2)\log_2(2/2)$ = 0
Pat = None: - $(0/2)\log_2(0/2)$ - $(2/2)\log_2(2/2)$ = 0

Price = $$$ - $(0/1)\log_2(0/1)$ - $(1/1)\log_2(1/1)$ = 0
Price = $$: - $(0/0)\log_2(0/0)$ - $(0/0)\log_2(0/0)$ = 0
Price = $: - $(1/4)\log_2(1/4)$ - $(3/4)\log_2(3/4)$ = 0.81

Rain = T: - $(1/3)\log_2(1/3)$ - $(2/3)\log_2(2/3)$ = 0.92
Rain = F: - $(0/2)\log_2(0/2)$ - $(2/2)\log_2(2/2)$ = 0

Res = T: - $(1/4)\log_2(1/4)$ - $(3/4)\log_2(3/4)$ = 0.81
Res = F: - $(0/1)\log_2(0/1)$ - $(1/1)\log_2(1/1)$ = 0

Price = Burger - $(1/3)\log_2(1/3)$ - $(2/3)\log_2(2/3)$ = 0.92
Price = French: - $(0/1)\log_2(0/1)$ - $(1/1)\log_2(1/1)$ = 0
Price = Thai: - $(0/1)\log_2(0/1)$ - $(1/1)\log_2(1/1)$ = 0
Type = Italian: - $(0/0)\log_2(0/0)$ - $(0/0)\log_2(0/0)$ = 0

Est = T: - $(1/3)\log_2(1/3)$ - $(2/3)\log_2(3/3)$ = 0.92
Est = F: - $(0/2)\log_2(0/2)$ - $(2/2)\log_2(2/2)$ = 0

Information gain of all attributes from Hun = F:

Act: 0.72 – (4/5) 0.81 – (1/5) 0 = 0.07
Bar: 0.72 – (3/5) 0.92 – (2/5) 0 = 0.17
Fri: 0.72 – (3/5) 0.92 – (2/5) 0 = 0.17
Pat: 0.72 – (1/5) 0 – (2/5) 0 – (2/5) 0 = 0.72
Price: 0.72 – (1/5) 0 – (0/5) 0 – (4/5) 0.81 ≈ 0.07
Rain: 0.72 – (3/5) 0.92 – (2/5) 0 ≈ 0.17
Res: 0.72 – (4/5) 0.81 – (1/5) 0 ≈ 0.07
Type: 0.72 – (3/7) 0.92 – (1/7) 0 – (1/7) 0 – (0/7) 0 = 0.17
Est: 0.72 – (3/5) 0.92 – (2/5) 0 = 0.17

For Hun = F, the greatest information gain is 0.72 which belongs to the attribute Pat, therefore Pat will be the next attribute chosen by ID3.

2) As calculated above,

For Hun = F, the expected information gain associated with the next attribute Pat is:

= 0.72 – (1/5) 0 – (2/5) 0 – (2/5) 0 = 0.72

3) In a general sense, the expected information gain is calculated by calculating the "expected reduction in entropy due to splitting on A"

Entropy as defined in the class notes is:

$$\mathrm{H}[x] = -\sum_{x} p(x) \log_2 p(x)$$

Information gain as defined in the class notes is:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Let S be the root set and S_v be the subset (the attribute).

Then, in general, calculate the entropy for the root set and all of its attributes. Then for each attribute, use the entropy of the root set and subtract the sum of the entropies of the attribute multiplied by the probability $|S_v| / |S|$. This yields the information gain for each attribute based on the root set S. We then choose the attribute that yields the highest information gain to split.

In this question, I first calculated the entropy for Hun = T and Hun = F because we were given the assumption that ID3 splits first on the Hun attribute.

Then the entropies of all the attributes were calculated in a similar way by first counting the number of trues and number of falses in the “target will wait” column for each choice in the attribute (for example: two choices for Fri, Fri = T and Fri = F); then plugging the number of trues over the total of trues + falses into p(x) for one part of the entropy equation and the number of falses over the total of trues + falses into p(x) for the other part of the entropy equation and finally subtracting the two (for example: if there were 5 trues and 2 falses, then H(x) = - (5/7) $\log_2(5/7)$ - (2/7) $\log_2(2/7)$).

Now that we calculated all the entropies, we fill in the information gain equation, by first putting the entropy for Hun = F in the information gain equation; then for each attribute, I subtracted the entropy of the attribute that we want to test multiplied by the probability $|S_v| / |S|$, from the entropy of the root set (which is entropy of Hun = F = 0.72). This yields the information gain for each attribute.

Trace Backpropagation

1) **Formulas**

$In_i = \sum_{k=0}^{n} w_{k,i} a_k$

$a_i = g(in_i) = g(\sum_{k=0}^{n} w_{k,i} a_k)$ where g is the activation function

$\Delta[i] = g`(in_i) (\sum_k w_{i,k} \Delta_k)$

Weight update for $W_{i,j}$ = $w_{i,j} \leftarrow w_{i,j} + \alpha \times a_i \times \Delta_j$

2)

Iteration	a_c	Δ[c]	a_d	Δ[d]	W_{0c}	W_{ac}	W_{bc}	W_{cd}	W_{0d}
X_1	0.6457	0.0061	0.6210	0.0892	0.3012	0.3012	0.3000	0.3115	0.3178
X_2	0.6459	-0.0115	0.5190	-0.1477	0.2989	0.3012	0.2977	0.2924	0.2883

Iteration X_1

W_{0c} = 0.3 + 0.2 * 1 * 0.0061 = 0.3012

W_{ac} = 0.3 + 0.2 * 1 * 0.0061 = 0.3012

W_{bc} = 0.3 + 0.2 * 0 * 0.0061 = 0.3000

W_{cd} = 0.3 + 0.2 * 0.6457 * 0.0892 = 0.3115

W_{0d} = 0.3 + 0.2 * 1 * 0.0892 = 0.3178

Iteration X_2

W_{0c} = 0.3012 + 0.2 * 1 * -0.0115 = 0.3012

W_{ac} = 0.3012 + 0.2 * 0 * -0.0115 = 0.2989

W_{bc} = 0.3000 + 0.2 * 1 * -0.0115 = 0.2977

W_{cd} = 0.3115 + 0.2 * 0.6459 * -0.1477 = 0.2924

W_{0d} = 0.3178 + 0.2 * 1 * -0.1477 = 0.2883