Solutions

Question 1

a. \( P(\text{toothache}) = 0.108 + 0.016 + 0.012 + 0.064 = 0.2 \)

b. \( P(\text{catch}) = 0.108 + 0.016 + 0.072 + 0.144 = 0.34 \)

c. \( P(\text{cavity} \mid \text{catch}) = \frac{P(\text{cavity} \land \text{catch})}{P(\text{catch})} = \frac{0.108 + 0.072}{0.34} = 0.5294 \)

d. \( P(\text{catch} \mid \text{toothache} \lor \text{cavity}) = \frac{P(\text{catch} \land (\text{toothache} \lor \text{cavity}))}{P(\text{toothache} \lor \text{cavity})} = \frac{0.108 + 0.016 + 0.072}{0.108 + 0.016 + 0.012 + 0.064 + 0.072 + 0.008} = 0.70 \)

Question 2

The Bayesian network corresponding to the set up is given below:

The CPT for coin selection is

<table>
<thead>
<tr>
<th>C</th>
<th>P(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1/3</td>
</tr>
<tr>
<td>b</td>
<td>1/3</td>
</tr>
<tr>
<td>c</td>
<td>1/3</td>
</tr>
</tbody>
</table>
The CPTs for $X_1$, $X_2$ and $X_3$ is same and is given as:

| $C$ | $X_i$ | $P(X_i | C)$ |
|-----|-------|-------------|
| a   | H     | 0.3         |
| b   | H     | 0.5         |
| c   | H     | 0.6         |

For the observed sequence of 2H and 1T, the coin most likely to have been drawn is given by value of $C$ with greatest posterior probability.

Using Bayes Rule,

$$P(C | 2H, 1T) = \frac{P(2H, 1T | C) P(C)}{P(2H, 1T)}$$

Since, $P(C)$ is equal for all values of $C$ and $P(2H, 1T)$ is not dependent on $C$, we can ignore them leading to

$$P(C | 2H, 1T) \propto P(2H, 1T | C)$$

Since $X_1$, $X_2$ and $X_3$ are conditionally independent given $C$, the RHS from above can be written as:

$$= P(H | C) P(H | C) P(T | C)$$

For $C = a$,  \[ P(H | a) P(H | a) P(T | a) = 0.3 \times 0.3 \times 0.7 = 0.063 \]

For $C = b$,  \[ P(H | b) P(H | b) P(T | b) = 0.5 \times 0.5 \times 0.5 = 0.125 \]

For $C = c$,  \[ P(H | c) P(H | c) P(T | c) = 0.6 \times 0.6 \times 0.4 = 0.144 \]

Hence, Coin c is most likely to have been drawn.
Question 3
For the first iteration with 9 features, the Information Gain is highest for Pat (0.541) and hence it is selected as the root of tree.
For the second iteration, Estimate, Price, Type, Reservation all have the same top ranked information Gain (Information Gain: 0.251). We select attribute Price for the dataset which has (Pat = Full). [Other values None and Some have same labels for all remaining dataset]
For the third iteration, we select attribute Est (Information Gain: 0.5) for the dataset which has (Pat=Full and Price=$) [Other values $$ and $$$ either have same labels or have no data remaining]
For the fourth iteration, we select attribute Bar (Information Gain: 1.0, same as Type and Fri) for the dataset which has (Pat = Full and Price = $ and Est=30-60) [Other values 10-30 and >60 have same labels for remaining dataset]
The attribute Bar completely partitions the dataset with value Yes representing True label and value No representing False label.
The decision tree is represented below:
Question 4

The value of attribute used is:

- Alt: No=0.1, Yes=0.9
- Bar: No=0.1, Yes=0.9
- Fri: No=0.1, Yes=0.9
- Pat: None=0.1, Some = 0.5, Full=0.9
- Price: $=0.1, $$=0.5, $$$=0.9
- Rain: No=0.1, Yes=0.9
- Res: No=0.1, Yes=0.9
- Type: French=0.1, Thai=0.3, Burger=0.6, Italian=0.9
- Est: 0-10=0.1, 10-30=0.3, 30-60=0.7, >60=0.9
- WillWait (Output Label): No=0.1, Yes=0.9

The data looks something like:

<table>
<thead>
<tr>
<th>Alt</th>
<th>Bar</th>
<th>Fri</th>
<th>Pat</th>
<th>Price</th>
<th>Rain</th>
<th>Res</th>
<th>Type</th>
<th>Est</th>
<th>WillWait</th>
</tr>
</thead>
<tbody>
<tr>
<td>.9</td>
<td>.1</td>
<td>.1</td>
<td>.5</td>
<td>.9</td>
<td>.1</td>
<td>.9</td>
<td>0.1</td>
<td>.1</td>
<td>.9</td>
</tr>
<tr>
<td>.9</td>
<td>.1</td>
<td>.1</td>
<td>.9</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>0.3</td>
<td>.7</td>
<td>.1</td>
</tr>
<tr>
<td>.1</td>
<td>.9</td>
<td>.1</td>
<td>.5</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>0.6</td>
<td>.1</td>
<td>.9</td>
</tr>
<tr>
<td>.9</td>
<td>.1</td>
<td>.9</td>
<td>.9</td>
<td>.1</td>
<td>.9</td>
<td>.1</td>
<td>0.3</td>
<td>.3</td>
<td>.9</td>
</tr>
<tr>
<td>.9</td>
<td>.1</td>
<td>.9</td>
<td>.9</td>
<td>.1</td>
<td>.9</td>
<td>.1</td>
<td>0.1</td>
<td>.9</td>
<td>.1</td>
</tr>
<tr>
<td>.1</td>
<td>.9</td>
<td>.1</td>
<td>.5</td>
<td>.5</td>
<td>.9</td>
<td>.9</td>
<td>0.9</td>
<td>.1</td>
<td>.9</td>
</tr>
<tr>
<td>.1</td>
<td>.9</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.9</td>
<td>.1</td>
<td>0.6</td>
<td>.1</td>
<td>.1</td>
</tr>
<tr>
<td>.1</td>
<td>.9</td>
<td>.1</td>
<td>.5</td>
<td>.5</td>
<td>.9</td>
<td>.9</td>
<td>0.3</td>
<td>.1</td>
<td>.9</td>
</tr>
<tr>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.1</td>
<td>.9</td>
<td>.1</td>
<td>0.6</td>
<td>.9</td>
<td>.1</td>
</tr>
<tr>
<td>.1</td>
<td>.9</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>0.3</td>
<td>.1</td>
<td>.1</td>
</tr>
</tbody>
</table>

A neural network with three hidden nodes is used. The weight learned are:

Linear Node 0 (Output Node)

Inputs  Weights
Threshold  -0.71236416
Node 2    -0.65187498
Node 1    -0.37276145
Node 3    0.19314144

The matrix from Input layer to hidden layer is as follows (dimension 9 x 3):

\[
\begin{bmatrix}
-0.11730051, & 0.39617276, & -0.60786122, \\
-0.27942939, & -0.7500252, & -0.4375444, \\
-0.44351322, & -0.1556574, & -0.61527552, \\
0.05487265, & 0.22576856, & 1.98704303, \\
-0.41779634, & 0.50819454, & -0.59614038, \\
0.24097866, & -0.26898639, & 0.64122423, \\
-0.50836124, & -0.4530991, & 1.13619388, \\
0.66195047, & -0.32410729, & -0.54054069, \\
0.53207649, & 1.20609996, & -1.81732653
\end{bmatrix}
\]

With the bias vector being
[-0.65187498, -0.37276145, 0.19314144]

Time taken to build model: 1179.81 seconds

This is the weights learned by the neural network during back-propagation.

Number of Epochs: 500 (Iterations: 500 * 13)
Learning rate: 0.01
Momentum: 0.9

Training set score: 0.916667
Total training time: 0.13102197647094727 seconds