QUESTION 1 (5pts)

1. (1pt) What is a holdout procedure

Holdout procedure is a method of splitting original data into training and test data sets

2. (1pt) Describe shortly the main 2 methods of predictive accuracy evaluations

k-fold cross-validation (N - N/k ; N/k)

First step: split data into k disjoint subsets D1, Dk, of equal size, called folds

Second step: use each subset in turn for testing, the remainder for training

Training and testing is performed k times

Leave-one-out (N-1 ; 1)

Leave-one-out is a particular form of cross-validation

We set number of folds to number of training instances, i.e. k = N

For N instances we build classifier (repeat the training - testing) n times

3. (3pt) Show how to perform the 3-fold cross-validation (N - N/3 ; N/3) on the CLASSIFICATION DATA below.

It means SHOW how this method divides data into TRAIN-TEST subsets and how final predictive accuracy is evaluated assuming that you KNOW the predictive accuracy for each division

CLASSIFICATION DATA

<table>
<thead>
<tr>
<th>a1</th>
<th>a2</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>c1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>c2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>c1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>c2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>c3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>c3</td>
</tr>
</tbody>
</table>

We split the data into equal disjoint 3 SUBSETS in any way we choose. Let’s call them A, B, C

We perform learning and testing for each FOLD (it means 3 times)

For each FOLD evaluate iys predictive accuracy

Lets call them PA, PB, PC (for subsets A, B, C used for test sets in the corresponding fold
The final predictive accuracy $P$ is

$$P = \frac{P_A + P_B + P_C}{3}$$

QUESTION 2. (12pts) Classification by Decision Tree

Given a Classification DB

<table>
<thead>
<tr>
<th>O</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>o2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>o3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>o4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>o5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>o6</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>o7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**GOAL:** use the above DB and two - fold cross validation holdout to build a CLASSIFIER using the Decision Tree using the BASIC Algorithm with $a_1$ as the root and $a_2, a_3$ as the consecutive split attributes in the subtrees (if needed).

**Build** your CLASSIFIER using the fold $F_1 = \{o_1, o_2, o_3\}$ for training - and rest for testing and vice-versa:

use $T_2 = \{o_4, o_5, o_6, o_7\}$ for training - and the rest for testing.

Here are steps you must follow.

**STEP 1** (3pts) Build (draw) the Decision Tree with $a_1$ as the root for $F_1 = \{o_1, o_2, o_3\}$

The correct Tree in unique. No partial credit

**Solution**

**General Observation** for all STEPS.

1. MAJORITY VOTING class is $C = 1$ - as it is a majority class in the Training set.

2. The attributes $a_1, a_2, a_3, a_4$ have the following sets of values.

   - $a_1$ has values: 0, 1, 2
   - $a_2$ has values: 0, 1
   - $a_3$ has values: 0, 1, 2
   - $a_4$ has values: 0, 1

The Decision Tree we are building has 2 or 2 branches depending on the root and internal nodes.

We build a tree $T_1$ with the fold $F_1 = \{o_1, o_2, o_3\}$ used for TRAINING

<table>
<thead>
<tr>
<th>O</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>o2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>o3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Tree $T_1$

The tree $T_1$ with root node $a_1$ has 3 branches: $b_1$ for $a_1 = 0$, $b_2$ for $a_1 = 1$, and $b_3$ for $a_1 = 2$ and we consider then one by one.

Branch $b_1$.

The tree $T_1$ branch $b_1$: with root node $a_1$ and $a_1 = 0$ is:

<table>
<thead>
<tr>
<th></th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

We use now, as instructed, the attribute $a_2$ as the split attribute and have a node $a_2$ with the following 2 branches for $a_2 = 0, 1$, respectively.

1. Branch $a_2 = 0$

<table>
<thead>
<tr>
<th></th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

We get a leaf $C = 0$ and a rule:

$$R_1 : a_1(x, 0) \cap a_2(x, 0) \Rightarrow C(x, 0)$$

2. Branch $a_2 = 1$

<table>
<thead>
<tr>
<th></th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

We get a leaf $C = 1$ and a rule:

$$R_2 : a_1(x, 0) \cap a_2(x, 1) \Rightarrow C(x, 1)$$

Branch $b_2$.

The tree $T_1$ branch $b_2$: with root node $a_1$ and $a_1 = 1$

No records on this branch, so by Termination Condition we use MAJORITY VOTE and get a leaf $C = 1$ and a rule:

$$R_3 : a_1(x, 1) \Rightarrow C(x, 1)$$

Branch $b_3$.

The tree $T_1$ branch $b_3$: with root node $a_1$ and $a_1 = 2$ is:

<table>
<thead>
<tr>
<th></th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

We get a leaf $C = 2$ and a rule:

$$R_4 : a_1(x, 2) \Rightarrow C(x, 2)$$
STEP 2 (2pts) Write the set $R_1$ of RULES resulting from your tree $T_1$ in in a predicate form and evaluate rules accuracy and predictive accuracy

Solution

The set $R_1$ of rules for the fold $F_1 = \{o_1, o_2, o_3\}$ and Tree $T_1$ is:

$R_1 : a_1(x, 0) \cap a_2(x, 0) \Rightarrow C(x, 0),$  
$R_2 : a_1(x, 0) \cap a_2(x, 1) \Rightarrow C(x, 1),$  
$R_3 : a_1(x, 1) \Rightarrow C(x, 1),$  
$R_4 : a_1(x, 2) \Rightarrow C(x, 2).$

TESTING $R_1$ rules accuracy with $\{o_1, o_2, o_3\}$

<table>
<thead>
<tr>
<th>O</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>o2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>o3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Tuple $o_1$ is classified by $R_2$,  
$o_2$ is classified by $R_4$,  
$o_3$ is classified by $R_1$.

Rules accuracy for $T_1$ is 100%

TESTING predictive accuracy with $\{o_4, o_5, o_6, o_7\}$

<table>
<thead>
<tr>
<th>O</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>o4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>o5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>o6</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>o7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Tuple $o_5$ is classified by $R_1$,  
$o_4$ is misclassified by $R_1$,  
$o_5$ is classified by $R_4$,  
$o_6$ is misclassified by $R_4$,  
$o_7$ is classified by $R_1$.

Predictive accuracy for the fold $F_1$ is 50%.

We reject this result as "My Classifier".

STEP 3 (3pts) Build (draw) the Decision Tree with $a_1$ as the root for $F_2 = \{o_4, o_5, o_6, o_7\}$

The correct Tree in unique. No partial credit
Solution

We build a tree $T_2$ with the fold $F_2 = \{o4, o5, o6, o7\}$ used for TRAINING

<table>
<thead>
<tr>
<th>O</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>o4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>o5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>o6</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>o7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Tree $T_2$

The tree $T_2$ with root node $a_1$ has 3 branches: $b_1$ for $a_1 = 0$, $b_2$ for $a_1 = 1$, and $b_3$ for $a_1 = 2$ and we consider then one by one.

Branch $b_1$.

The tree $T_2$ branch $b_1$: with root node $a_1$ and $a_1 = 0$ is:

<table>
<thead>
<tr>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

We use now, as instructed, the attribute $a_2$ as the split attribute and have a node $a_2$ with the following 2 branches for $a_2 = 0, 1$, respectively.

1. Branch $a_2 = 0$

<table>
<thead>
<tr>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

We use now, as instructed, the attribute $a_3$ as the split attribute and have a node $a_3$ with the following 3 branches for $a_3 = 0, 1, 2$, respectively.

Branch $a_3 = 0$

No records on this branch, so by Termination Condition we use MAJORITY VOTE and get a leaf $C = 1$ and a rule:

$$ R_1 : a_1(x, 0) \cap a_2(x, 0) \cap a_3(x, 0) \Rightarrow C(x, 1) $$

Branch $a_3 = 1$

<table>
<thead>
<tr>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

We get a leaf $C = 0$ and a rule:

$$ R_2 : a_1(x, 0) \cap a_2(x, 0) \cap a_3(x, 1) \Rightarrow C(x, 0) $$

Branch $a_3 = 2$

<table>
<thead>
<tr>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
We get a leaf \( C = 1 \) and a rule:

\[
R3 : \ a1(x, 0) \cap a2(x, 0) \cap a3(x, 2) \Rightarrow C(x, 1)
\]

2. Branch \( a2 = 1 \)

No records on this branch, so by Termination Condition we use MAJORITY VOTE and get a leaf \( C = 1 \) and a rule:

\[
R4 : \ a1(x, 0) \cap a2(x, 1) \Rightarrow C(x, 1)
\]

This terminates the T2, Branch b1.

Branch b2

The T2 branch b2: with root node \( a1 \) and \( a1 = 1 \)

No records on this branch, so by Termination Condition we use MAJORITY VOTE and get a leaf \( C = 1 \) and a rule:

\[
R5 : \ a1(x, 1) \Rightarrow C(x, 1)
\]

Branch b3.

The T2 branch b3: with root node \( a1 \) and \( a1 = 2 \)

\[
\begin{array}{ccc|c}
 a2 & a3 & a4 & C \\
\hline
 1 & 1 & 0 & 2 \\
 0 & 1 & 0 & 1 \\
\end{array}
\]

We use now, as instructed, the attribute \( a2 \) as the split attribute and have a node \( a2 \) with the following 2 branches for \( a2 = 0, \ a2 = 1 \).

1. Branch \( a2 = 0 \)

\[
\begin{array}{ccc|c}
 a3 & a4 & C \\
\hline
 1 & 0 & 1 \\
\end{array}
\]

We get a leaf \( C = 0 \) and a rule:

\[
R6 : \ a1(x, 2) \cap a2(x, 0) \Rightarrow C(x, 1)
\]

2. Branch \( a2 = 1 \)

\[
\begin{array}{ccc|c}
 a3 & a4 & C \\
\hline
 1 & 0 & 2 \\
\end{array}
\]

We get a leaf \( C = 2 \) and a rule:

\[
R7 : \ a1(x, 2) \cap a2(x, 1) \Rightarrow C(x, 2)
\]

**STEP 4** (2pts) Write the set of RULES R2 resulting from your tree T2 in a predicate form and evaluate the rules accuracy and predictive accuracy.
Solution

The set $\mathbf{R2}$ of rules for the fold $\mathbf{F1} = \{o1, o2, o3\}$ and Tree $\mathbf{T2}$ is:

- $\mathbf{R1}: a1(x, 0) \cap a2(x, 0) \cap a3(x, 0) \Rightarrow C(x, 1)$,
- $\mathbf{R2}: a1(x, 0) \cap a2(x, 0) \cap a3(x, 1) \Rightarrow C(x, 0)$,
- $\mathbf{R3}: a1(x, 0) \cap a2(x, 0) \cap a3(x, 2) \Rightarrow C(x, 1)$,
- $\mathbf{R4}: a1(x, 0) \cap a2(x, 1) \Rightarrow C(x, 1)$,
- $\mathbf{R5}: a1(x, 1) \Rightarrow C(x, 1)$,
- $\mathbf{R6}: a1(x, 2) \cap a2(x, 0) \Rightarrow C(x, 1)$,
- $\mathbf{R7}: a1(x, 2) \cap a2(x, 1) \Rightarrow C(x, 2)$.

**TESTING $\mathbf{R2}$ rules accuracy** with the fold $\mathbf{F2} = \{o4, o5, o6, o7\}$ used for TRAINING

<table>
<thead>
<tr>
<th>O</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>o4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>o5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>o6</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>o7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

- $o4$ is classify by rule $R3$
- $o5$ is classified by $R7$
- $o6$ is classified by $R6$
- $o7$ is classified by $R2$.

**Rules accuracy** for $\mathbf{F2}$ is 100%

**TESTING predictive accuracy** with $\{o1, o2, o3\}$

<table>
<thead>
<tr>
<th>O</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>o2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>o3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- Tuple $o1$ is classified by $R4$,
- $o2$ is classified by $R7$,
- $o3$ is misclassified by $R1$.

**Predictive accuracy** for $\mathbf{F2}$ is 66.7%

**STEP 5** (2pts) Write your answer as ”My Classifier is: ” and Justify your decision.

I accept the set $\mathbf{R2}$ rules as my Classifier; not a very string one - but rules accuracy is 100%.

**Observe** that there is a strong reason to reject it and to give an answer There is no Classifier.

**QUESTION 3** (5pts)
1. (1pts) Give a short general description what is a Neural Network

   Neural Network is a set of connected INPUT/OUTPUT UNITS, where each connection has a WEIGHT associated with it

2. (1pts) Give a short general description how Neural Network learns

   Neural Network learns by adjusting the weights so as to be able to correctly classify the training data and hence, after testing phase, to classify unknown data

3. (1pts) Given a classification data \( D \) with attributes \( a_1, a_2, \ldots, a_n \) and classes \( c_1, c_2, \ldots, c_k \)

Which is the number of INPUT nodes of any NN for \( D \)?

   \( \text{There is } n \text{ nodes, as many as attributes} \)

Which is the number of OUTPUT nodes of any NN for full classification for \( D \)?

   \( \text{There is } k \text{ nodes, as many as classes} \)

Which is the number of hidden layers?

   \( \text{There is as many as we want; must be at least one} \)

Which is the number of nodes in the hidden layers?

   \( \text{There is as many as we want; must be at least one} \)

4. (2pts) Design 3 Neural Networks for the CLASSIFICATION DATA from QUESTION 2

   One for full classification

   \( \text{There is FOUR input nodes, THREE output nodes - at least one hidden layer with number of nods of your choice.} \)

   Two for contrast learning (for your chosen classes)

   **Solution**

   There were two possibilities.

   1. \( \text{There is FOUR input nodes, TWO output nodes - one for your CLASS of choice, the other for NOL CLASS. There is at least one hidden layer with number of nods of your choice} \)

   2. \( \text{There is FOUR input nodes, ONE output node for your CLASS - at least one hidden layer with number of nods of your choice} \)

   The output node represent a CLASS (one of three ) of your choice. Hence you can have 3 choices for OUTPUT node

   \( \text{Must draw all weight with proper indexes as i Hmk 3 or Lecture} \)

   Draw pictures and explain correctness of your topology

   Enumerate all nodes and weights

   \( \text{The explanations are above} \)
QUESTION 4  (3pts)

1. Give a general description of the following STEPS of the Backpropagation Algorithm

Step 1: initialize the weights and biases

Step 2: feed the training sample

Step 3: propagate the inputs forward

Step 4: backpropagate the error

Step 5: backpropagate the weights, biases

Step 6: repeat and apply Terminating Conditions

Step 7: terminate when

all weights $w_{ij}$ in the previous epoch are below some threshold

the percentage of samples misclassified in the previous epoch is below some threshold

a pre-specified number of epochs has expired