CSE 352 QUIZ 2 SOLUTIONS SPRING 2022

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 **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

a1	a2	C
1	0	c1
0	1	c2
1	1	c1
1	0	c2
0	0	c3
0	0	c3

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 P_A, P_B, P_C *(for subsets* **A, B, C** *used for test sets in the corresponding fold*

The final **predctive accuracy** P is

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

QUESTION 2. (12pts) Classification by Decision Tree

Given a Classification DB

CLASSIFICATION DB

O	a1	a2	a3	a4	C
o1	0	1	1	0	1
o2	2	1	2	0	2
o3	0	0	0	0	0
o4	0	0	2	1	1
o5	2	1	1	0	2
o6	2	0	1	0	1
o7	0	0	1	1	0

GOAL: use the above DB and **two - fold cross validation holdout** to build a CLASSIFIER using the Decision Tree using the BASIC Algorithm with **a1** as the root and **a2** , **a3** as the **consecutive** split attributes in the subtrees (if needed).

Build your CLASSIFIER using the fold **F1** $= \{o1, o2, o3\}$ for training - and rest for testing and vice-versa:

use **T2** $= \{o4, o5, o6, o7\}$ for training - and the rest for testing.

Here are steps you must follow.

STEP 1 (3pts) Build (draw) the Decision Tree with **a1** as the root for

F1 $= \{o1, o2, o3\}$

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 $a1,\ a2,\ a3,\ a4$ have the following sets of values .

a1 has values: 0, 1, 2

a2 has values: 0, 1

a3 has values: 0, 1, 2

a4 has values: 0, 1

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

We build a tree **T1** with the fold **F1** $= \{o1, o2, o3\}$ used for **TRAINING**

O	a1	a2	a3	a4	C
o1	0	1	1	0	1
o2	2	1	2	0	2
o3	0	0	0	0	0

TreeT1

The tree **T1** with **root node** $a1$ has **3** branches: **b1** for $a1 = 0$, **b2** for $a1 = 1$, and **b3** for $a1 = 2$ and we consider then one by one.

Branch b1.

The tree **T1** branch **b1**: with **root node** $a1$ and $a1 = 0$ is:

a2	a3	a4	C
1	1	0	1
0	0	0	0

We use now, as instructed, the attribute $a2$ as the **split** attribute and have a **node** $a2$ with the following 2 branches for $a2 = 0, 1$, respectively.

1. Branch $a2 = 0$

a3	a4	C
0	0	0

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

$$R1: \ a1(x,0) \cap a2(x,0) \Rightarrow C(x,0)$$

2. Branch $a2 = 1$

a3	a4	C
1	0	1

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

$$R2: \ a1(x,0) \cap a2(x,1) \Rightarrow C(x,1)$$

Branch b2.

The tree **T1** 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**:

$$R3: \ a1(x,1) \Rightarrow C(x,1)$$

Branch b3.

The tree **T1** branch **b3**: with **root node** $a1$ and $a1 = 2$ is:

a2	a3	a4	C
1	2	0	2

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

$$R4: \ a1(x,2) \Rightarrow C(x,2)$$

STEP 2 (2pts) Write the set **R1** of RULES resulting from your tree **T1** in in a predicate form and evaluate **rules accuracy** and **predictive accuracy**

Solution

The set **R1** of rules for the fold **F1** $= \{o1, o2, o3\}$ and **Tree T1** is:

$R1: \ a1(x,0) \cap a2(x,0) \Rightarrow C(x,0),$

$R2: \ a1(x,0) \cap a2(x,1) \Rightarrow C(x,1)$

$R3: \ a1(x,1) \Rightarrow C(x,1),$

$R4: \ a1(x,2) \Rightarrow C(x,2).$

TESTING **R1** rules **accuracy** with $\{o1, o2, o3\}$

O	a1	a2	a3	a4	C
o1	0	1	1	0	1
o2	2	1	2	0	2
o3	0	0	0	0	0

Tuple o1 is classified by R2,
o2 is classified by R4,
o3 is classified by R1.

Rules accuracy for **T1** is 100%

TESTING **predictive accuracy** with $\{o4, o5, o6, o7\}$

O	a1	a2	a3	a4	C
o4	0	0	2	1	1
o5	2	1	1	0	2
o6	2	0	1	0	1
o7	0	0	1	1	0

Tuple o5 is classified by R1,
o4 is misclassified by R1,
o5 is classified by R4,
o6 is misclassified by R4,
o7 is classified by R1.

Predictive accuracy for the fold **F1** is 50 %.

We **reject** this result as "My Classifier".

STEP 3 (3pts) Build (draw) the Decision Tree with **a1** as the root for **F2** $= \{o4, o5, o6, o7\}$

The correct Tree in unique. No partial credit

Solution

We build a tree **T2** with the fold $\mathbf{F2} = \{o4, o5, o6, o7\}$ used for **TRAINING**

O	a1	a2	a3	a4	C
o4	0	0	2	1	1
o5	2	1	1	0	2
o6	2	0	1	0	1
o7	0	0	1	1	0

Tree T2

The tree **T2** with **root node** $a1$ has **3** branches: **b1** for $a1 = 0$, **b2** for $a1 = 1$, and **b3** for $a1 = 2$ and we consider then one by one.

Branch b1.

The tree **T2** branch **b1**: with **root node** $a1$ and $a1 = 0$ is:

a2	a3	a4	C
0	2	1	1
0	1	1	0

We use now, as instructed, the attribute $a2$ as the **split** attribute and have a **node** $a2$ with the following **2** branches for $a2 = 0, 1$, respectively.

1. Branch $a2 = 0$

a3	a4	C
2	1	1
1	1	0

We use now, as instructed, the attribute $a3$ as the **split** attribute and have a **node** $a3$ with the following 3 branches for $a3 = 0, 1, 2$, respectively.

Branch $a3 = 0$

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

$$R1: \ a1(x,0) \cap a2(x,0) \cap a3(x,0) \Rightarrow C(x,1)$$

Branch $a3 = 1$

a4	C
1	0

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

$$R2: \ a1(x,0) \cap a2(x,0) \cap a3(x,1) \Rightarrow C(x,0)$$

Branch $a3 = 2$

a4	C
1	1

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 $a1 = 2$

a2	a3	a4	C
1	1	0	2
0	1	0	1

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 $a2 = 0$

a3	a4	C
1	0	1

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

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

2. Branch $a2 = 1$

a3	a4	C
1	0	2

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

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

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

O	a1	a2	a3	a4	C
o4	0	0	2	1	1
o5	2	1	1	0	2
o6	2	0	1	0	1
o7	0	0	1	1	0

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

Rules accuracy for **F2** is 100%

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

O	a1	a2	a3	a4	C
o1	0	1	1	0	1
o2	2	1	2	0	2
o3	0	0	0	0	0

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

Predictive accuracy for **F2** is 66.7%

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

I accept the set **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 a1, a2, ... an** and classes **c1, c2, .. ck**

Which is the number of INPUT nodes of any NN for **D**?

There is **n** *nodes, as many as attributes*

Which is the number of OUTPUT nodes of any NN for full classification for **D**?

There is **k** *nodes, as many as classes*

Which is he number of hidden layers?

There is as many as we want; must be at least one

Which is the number of nodes in the hidden layers?

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

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 wre two possibilities.

1. *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. *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

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

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 weightsl wij 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*