CSE521 DATA MINING TEST 1 SOLUTIONS SPRING 2023
(70pts)

PART 1: SHORT QUESTIONS TOTAL **10pts**

QUESTION 1 (2pts)

1. Describe shortly all stages of the Data Mining Process (you can also draw a picture).

Solution

DM Process includes data cleaning, data integration, data selection, transformation, data mining proper, pattern evaluation, and knowledge presentation

DM Process MAIN STEPS are:

Preprocessing - includes all the operations that have to be performed before a data mining algorithm is applied

Data Mining Proper is a step in the DM process in which Data Mining algorithms are applied to obtain patterns in data.

Interpretation - discovered patterns are presented in a proper format and the user decides if it is necessary to re-iterate the process.

2. When you re-iterate the DM Process?

Solution

We do so when the discovered patterns are not acceptable. We re-itarete in going back to different level of the Process. Decision to which level we return depends on the interpretation of the results and their acceptance by the user.

QUESTION 2 (5pts)

1. Describe major goals and characteristics of Data Warehouse as opposed to the traditional Data Base Management Systems approach methodology.

Solution

Data Warehouse is organized around major subjects such as customer, supplier, product, and sales

Data Warehouse focuses on the modeling and analysis of data for decision makers instead of concentrating on the day-to-day operations and transaction processing of an organization.

Traditional approach is **query driven.** When a query is posed to a client site, a meta-dictionary is used to translate the query into queries appropriate for individual heterogeneous sites involved, and the results are integrated into a global answer set.

Data Warehouse approach is **update-driven**. High performance Information is integrated in advance and stored in warehouses for direct query and analysis.

It does not contain the most current information. Supports complex multi-dimensional queries.

2. Define and compare shortly OLTP and OLAP systems

Solution

The major task *on-line operational* data base systems is to perform on-line *transaction* and *query processing* .

These systems are called

OLTP - On-Line Transaction Processing Systems - for transactions processing

OLAP - On-Line Analytical Processing systems serve users or knowledge workers providing *data analysis* and acting as **decision support systems**.

3. Describe a Data Cube and its operations

Solution

Data Cube (can be multi-dimensional) is used to represent data along some measure of interest.

Each dimension represents some attribute in the database and the cells in the data cube represent the facts of interest.

For example, they could contain a count for the number of times that attribute combination occurs in the database, or the minimum, maximum, sum or average value of some attribute.

4. What is ROLAP.MOLAP and HOLAP?

Solution

ROLAP - it is Relational OLAP, a relational and specialized relational DBMS to store and manage warehouse data.

ROLAP - it is a Multidimensional OLAP, array-based storage structures providing a direct access to array data structures.

HOLAP - it is a Hybrid OLAP, storing **detailed data** in RDBMS and *aggregate data* in Multi-dimensional DBMS. User access it via MOLAP tools.

QUESTION 3 (3pts)

1. Define **classification data**

Solution

Classification data is a data table (with key attribute removed) with one attribute distinguished as a class attribute.

Class attribute must have a small number discrete values. We distinguish attribute part of the record and call it sample/ tuple and class part as determined by the class label attribute. Each tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute.

2. What is a **classifier**

Solution

CLASSIFIER is a "black box" that had been trained and tested and evaluated. It serves to classify unknown records; i.e tuples for which the class attribute is unknown

3. Define 4 stages of the process of building a Classifier

Solution

Stage 1 - use training data to build the classification patterns structure, It is a training stage.

Stage 2 - optimize parameters settings; can use (N:N) re-substitution- parameter tuning stage

Stage 3 - use test data to compute predictive accuracy/error rate. It is a testing stage.

The accepted results of Stages 1-3 are called learned or basic Classifiers.

Stage 4 - consolidate Stages 1-3 to build the **Classifier** as the final product of the process

QUESTION 4 (5pts)

1. Define Holdout and Repeated Holdout (Random Sampling)

Solution

Holdout is a method of randomly partitioning given data into two independent (disjoined) sets: Training set for model (classifier) construction and Test set for accuracy (predictive accuracy/error) estimation

Repeated Holdout also called Random sampling is repeating holdout $k \geq 2$ times.

2. Define k-fold cross-validation (N- N/k ; N/k) holdout and repeated k-fold cross-validation holdout

Solution

k-fold cross-validation holdout randomly partition the data into k mutually exclusive subsets $D_1, D_2, \ldots, D_k$ of approximately equal size, called **folds**

At *ith* iteration, uses D_i as **test** set and others as **training** set

3. Define Leave-one-out (N-1 ; 1) holdout.

Solution

Leave-one-out holdout is s a particular form of the k-fold cross-validation (N- N/k ; N/k). We set number of folds to number of training instances, i.e. $k = N$. . Each sample (record) is used the same number of times for training and once for testing.

4. Bootstrap Holdout

Solution

We perform **Bootstrap Holdout** as follows.

Given a data set **D** with **n** records, we form a **training set** by sampling **D** uniformly with replacement **n** times i.e.

each time a tuple is selected, it is equally likely to be selected again and re-added to the training set.

The data tuples that did not make it into the training set end up forming the **test set**

5 .632 boostrap Holdout

Solution

We perform **632 boostrap Holdout** as follows.

A data set with d tuples is sampled d times, with replacement, resulting in a training set of d samples

The data tuples that did not make it into the training set end up forming the test set.

PART 2: PROBLEMS · (55 pts)

PROBLEM 1 (10pts)

1. Classification Rules (5pts)

For the following **formulas** write and use the proper definitions to prove whether they are or they are not **discriminant** or **characteristic rules** in the following dataset DB

CLASSIFICATION DB

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

Characteristic Rule Definition (1pt)

Solution

A characteristic **formula** $CLASS \Rightarrow DESCRIPTION$ is called a **Characteristic Rule** in

the classification dataset **D** if and only if it is TRUE in **D** i.e. when the following holds

$$\{o \in \mathbf{D} : DESCRIPTION\} \cap \{o \in \mathbf{D} : CLASS\} \neq \emptyset$$

Discriminant Rule Definition (1pt)

Solution

A discriminant **formula** $DESCRIPTION \Rightarrow CLASS$ is called a **Discriminant Rule** in the classification dataset **D** if and only if it is TRUE in **D** i.e. when the following conditions hold

1. $\{o \in \mathbf{D}\ \ DESCRIPTION\} \neq \emptyset$

2. $\{o \in \mathbf{D}\ \ DESCRIPTION\} \subseteq \{o \in \mathbf{D}\ \ CLASS\}$

Formulas (3pts)

Solution

f1 $a1 = 1 \cap a2 = 1 \Rightarrow C = 1$

f1 is a Discriminant Rule because it is a Discriminant Formula and $\{o_1\}$ is a subset of $\{o_1, o_5\}$

f2 $C = 1 \Rightarrow a1 = 0 \cap a2 = 1 \cap a3 = 1$

f2 is a Characteristic Formula but **is not** a Characteristic Rule because $\{o :\ a1 = 0 \cap a2 = 1 \cap a3 = 1\} = \emptyset$ and $\{o_1\} \neq \emptyset$

f3 $a1 = 1\ .\Rightarrow C = 1$

f3 is a Discriminant Rule because it is a Discriminant Formula and $\{o_1\}$ is a subset of $\{o_1, o_5\}$

f4 $C = 1 \Rightarrow a1 = 1$

f4is a Characteristic Rule because it is a Characteristic Formula and $\{o_1, o_5\} \cap \{o_1\} \neq \emptyset$

f5 $a1 = 2 \cap a2 = 1 \cap a3 = 1 \Rightarrow C = 0$

f5 is **not** Characteristic Rule because it is **not** a Characteristic **Formula**

2. (3pts) **PROOF**

Prove that in any classification DB the inverse implication to the discriminant rule is a characteristic rule

Solution

By definition, for any classification dataset **D**,

$DESCRIPTION \Rightarrow CLASS$ is **Discriminant Rule** if and only if

1. $\{o \in \mathbf{D}\ DESCRIPTION\} \neq \emptyset$

2. $\{o \in \mathbf{D}\ DESCRIPTION\} \subseteq \{o \in \mathbf{D}\ CLASS\}$.

Therefore, $\{o \in \mathbf{D}\ DESCRIPTION\} \cap \{o \in \mathbf{D}\ CLASS\} \neq \emptyset$.

what **proves** that $CLASS \Rightarrow DESCRIPTION$ is a **Characteristic Rule**

3. (2pts) **EXAMPLE**

Given classification DB

Find a simple condition (example) under which the inverse implication to a characteristic rule

is ALWAYS a discriminant rule

Solution

Take any characteristic rule $CLASS \Rightarrow DESCRIPTION$ such that $\{o : DESCRIPTION\} = \{o\tilde{O}\}$.

By the definition of characteristic rule so we have that $\{o \in \mathbf{D} : DESCRIPTION\} \cap \{o \in \mathbf{D}\ CLASS\} \neq \emptyset$,

so $\{o\tilde{O}\} \cap \{o \in \mathbf{D}\ CLASS\} \neq \emptyset$ and $\{o\tilde{O}\} \subseteq \{o \in \mathbf{D}\ CLASS\}$.

This **proves** that $DESCRIPTION \Rightarrow CLASS$ is a **Discriminant Rule** for any characteristic rule

$CLASS \Rightarrow DESCRIPTION$, such that $\{o : DESCRIPTION\} = \{o\tilde{O}\}$.

This is not the only EXAMPLE!

You can also just give a simple example of few records DB

PROBLEM 2 DECISION TREES (25pts)

Part 1 (5pts)

1. List and explain shortly the Decision Tree Algorithm Attribute Selection Measures.

Solution

Attribute selection measures are heuristic procedures for selecting the attribute that ÒbestÓ discriminates the given tuples according to class.

Some attribute selection measures, like the **Gini Index** enforce the resulting tree to be binary. Others, like the **Information Gain** allow multi-way splits. They allow for two or more branches to be grown from a node. In this case the branches represent all the (discrete) values of the nodes attributes.

Some other attribute selection measures are: **Gini Ratio**, CHAID (measure based on a test for independence) and CART (finds multivariate splits based on a linear combination of attributes)

2. Describe in your words what is their role in the Decision Tree Construction and which kind of trees Òthey produce". Draw a picture as an example.

Solution

Book page 334m Figure 8.4

Part 2 (20pts) BUILDING DECISION TREE CLASSIFIER

Given Classification DB

O	a1	a2	C
o1	1	1	1
o2	0	0	0
o3	0	1	0
o4	0	0	0
o5	1	1	1
o6	1	1	0
o7	0	0	0
o8	1	0	1

Use the above DB and **repeated two fold cross validation holdout** to build a CLASSIFIER using the DT BASIC Algorithm with **a1** as the root. Follow **Stages 1- 4** of the process of building a classifier.

Remember that division into 2-folds is an ARBITRARY partition of records into 2 disjoint sets; so there may be many answers depending on the partitions.

Repeat the two fold cross validation holdout 4 times.

Build your final CLASSIFIER using the following folds for each repetitions round **1.- 4.**.

1. **f1** = $\{o1, o2, o3, o4\}$ for training - rest for testing.

2. **f2** = $\{o5, o6, o7, o8\}$ for training - rest for testing.

3. **f3** = $\{o1, o3, o5, o7\}$ for training - rest for testing.

4. **f4** = $\{o2, o4, o6, o8\}$ for training - rest for testing.

Here are the **STEPS** you must follow

STEP 1 (10pts)

Follow **Stages 1-3** to build, for each repetitions round **1.- 4.**, a learned classifier (base classifier, learned model) and name them **F1** , **F2** , **F3** , **F4** , respectively.

Write the learned classifiers **F1 - F4** in as a set of **discriminant rules** in the predicate form.

REMINDER

Basic algorithm TERMINATION CONDITIONS are as follows.

The recursive partitioning STOPS on a **node N** only when any one of the following conditions is TRUE.

1. All records (samples) for the given **node N** belong to the same class.

We convert node into a **leaf** and label it with this class

2. There are no remaining attributes on which the samples (records in the data table) may be further partitioned.

We apply the **Majority voting**, i.e. we convert the

node N into a leaf and label it with the most common class in D

which is a set of training tuples and their associated class labels.

3. There is no records (samples) left a LEAF is created with **majority vote** for training sample, i.e. we convert the **node N** into a leaf and label it with the most common class in D

REMARK

Students must draw pictures of Trees Constructions - and use the correct application of Terminating Conditions.

They must say which one they have used.

Solution

1. f1 $= \{o1, o2, o3, o4\}$ **TRAINING**

f1 Rules: $R1:\ a1(x,1) \Rightarrow C(x,1), \quad R2:\ a1(x,0) \Rightarrow C(x,0)$

Testing with $\{o5, o6, o7, o8\}$

Tuple o5 is classified by R1, o6 is misclassified by R1, o7 is classified by R2, o8 is classified by R1.

Predictive accuracy for **f1** = 75% and we **ACCEPT** the results as

$$\mathbf{F1} = \{\ R1:\ a1(x,1) \Rightarrow C(x,1), \quad R2:\ a1(x,0) \Rightarrow C(x,0)\ \}$$

F1 Rules Accuracy - testing with Training data $\{o1, o2, o3, o4\}$

Tuple o1 is classified by R1, o2 is classified by R1, o3 is classified by R1, o4 is classified by R2

Rules accuracy for **F1** is 100%

2. f2 $= \{o5, o6, o7, o8\}$ **TRAINING**

f2 Rules $R1:\ a1(x,1) \cap a2(x,1) \ \Rightarrow\ C(x,0), \quad R2:\ a1(x,1) \cap a2(x,0) \ \Rightarrow C(x,0), \quad R3:\ a1(x,0) \Rightarrow C(x,0)$

Testing with $\{o1, o2, o3, o4\}$

Tuple o1 is misclassified by R1, o2 is classified by R3, o3 is classified by R3, o4 is classified by R3.

Predictive accuracy for **f2** is 75% and we **ACCEPT** the results as

$$\mathbf{F2} = \{\ R1:\ a1(x,1) \cap a2(x,1) \ \Rightarrow\ C(x,0), \quad R2:\ a1(x,1) \cap a2(x,0) \ \Rightarrow C(x,0), \quad R3:\ a1(x,0) \Rightarrow C(x,0)\ \}$$

F2 Rules Accuracy - testing with Training data $\{o5, o6, o7, o8\}$

o5 is misclassified by R1, o6 is classified by R2, o7 is classified by R3, o8 is misclassified by R2

Rules accuracy for **F2** is 50%

3. f3 $= \{o1, o3, o5, o7\}$ **TRAINING**

f3 Rules $R1:\ a1(x,1) \Rightarrow C(x,1), \quad R2:\ a1(x,0) \Rightarrow C(x,0)$

Testing with $\{o2, o4, o6, o8\}$

o2 is classified by R2, o4 is classified by R2, o6 is mismisclassified by R1, o8 is classified by R1.

Predictive accuracy for **f3** is 75% and we **ACCEPT** the results as

$$\mathbf{F3} = \{\ 1:\ a1(x,1) \Rightarrow C(x,1),\quad R2:\ a1(x,0) \Rightarrow C(x,0)\ \}$$

F3 Rules Accuracy - testing with Training data $\{o1, o3, o5, o7\}$

o1 is classified by R1, o3 is classified by R2, o5 is classified by R1, o7 is classified by R2

Rules accuracy for **F3** is 100%

4. f4 = $\{o2, o4, o6, o8\}$ **TRAINING**

f4 Rules $R1:\ a1(x,1) \Rightarrow C(x,0),\quad R2:\ a1(x,0) \Rightarrow C(x,0)$

Testing with $\{o1, o3, o5, o7\}$

o1 is misclassified by R1, o3 is classified by R2, o5 is misclassified by R1, o7 is classified by R1.

Predictive accuracy for **f3** = 50% and we **REJECT** the results.

STEP 2 (10pts)

Perform the **Stage 4** as follows: use the learned classifiers from STEP 1 and their predictive accuracy and rules accuracy as metrics to choose ONE as your final CLASSIFIER **F**

All of STEP 1 accepted basic classifiers **F1**, **F2** and **F3** have the same predictive accuracy.

The Rules accuracy of **F2** is 50% to a 100% of **F1** and **F3**.

F1 and **F3** have also the same set of Rules.

We hence adopt

$$\mathbf{F} = \{\ R1:\ a1(x,1) \Rightarrow C(x,1),\qquad R2:\ a1(x,0) \Rightarrow C(x,0)\ \}$$

as the final **CLASSIFIER**

STEP 3 (5pts)

1. Construct as your final CLASSIFIER the **bagged** Ensemble Classifier $\mathbf{F}^*$

2. Use your CLASSIFIERS **F** (from STEP 2) and $\mathbf{F}^*$ to classify the following records and compare the results.

O	a1	a2
o1	0	1
o2	0	0
o3	1	0
o4	1	1

Definition

Bagged Ensamble Classifier $\mathbf{M}^*$

Given **k** learned (basic) classifiers (models) derived from classification data D

$$\mathbf{M1},\ \mathbf{M2},\ \ldots\ \ldots, \mathbf{Mk}$$

To classify a record **o** we proceed as follows: each classifier model **Mi** returns its class prediction and the bagged classifier $\mathbf{M}^*$ counts the **votes** and assigns the **class** with the **most votes** to it.

1. Solution

We build the the **bagged** Ensemble Classifier $\mathbf{F}^*$ from leaned (basic) classifiers **F1**, **F2** and **F3** derived from the STEP 1 following the above definition.

2. Use your CLASSIFIERS **F** and $\mathbf{F}^*$ to classify the following records and compare the results.

O	a1	a2
o1	0	1
o2	0	0
o3	1	0
o4	1	1

Consider record **o1**, i.e. the tuple (0,1)

F1 returns $C = 0$, **F2** returns $C = 0$, **F3** returns $C = 0$

$\mathbf{F}^*$ returns $C = 0$

F returns $C = 0$

Consider record **o2**, i.e. the tuple (0,0)

F1 returns $C = 0$, **F2** returns $C = 0$, **F3** returns $C = 0$

$\mathbf{F}^*$ returns $C = 0$

F returns $C = 0$

Consider record **o3**, i.e. the tuple (1, 0)

F1 returns $C = 1$, **F2** returns $C = 0$, **F3** returns $C = 1$

$\mathbf{F}^*$ returns $C = 1$

F returns $C = 1$

Consider record **o4**, i.e. the tuple (1,1)

F1 returns $C = 1$, **F2** returns $C = 0$, **F3** returns $C = 1$

$\mathbf{F}^*$ returns $C = 1$

F returns $C = 1$

2. We can also consider take the classifiers **F1**, **F2**, **F3** and **F4** derived from the STEP1 (i.e. to accept predoctive accuracy of 50% to build $\mathbf{F}^*$.

PROBLEM 3 (20pts) NEURAL NETWORK

Part 1 (3pts)

1. 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. 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. 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 the number of OUTPUT nodes of any NN for **contrast classification** for **D**?

There are **2** nodes: one for the Target Class and one for the rest of classes. We contrast Target class with rest of the classes.

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

There is only one node - the Target Class

Which is the number of hidden layers?

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

Part 2 (2pts)

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 **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**

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

or a pre- specified number of **epochs** has **expired**

PART 3 (15pts) BUILDING NN CLASSIFIER

Given the following

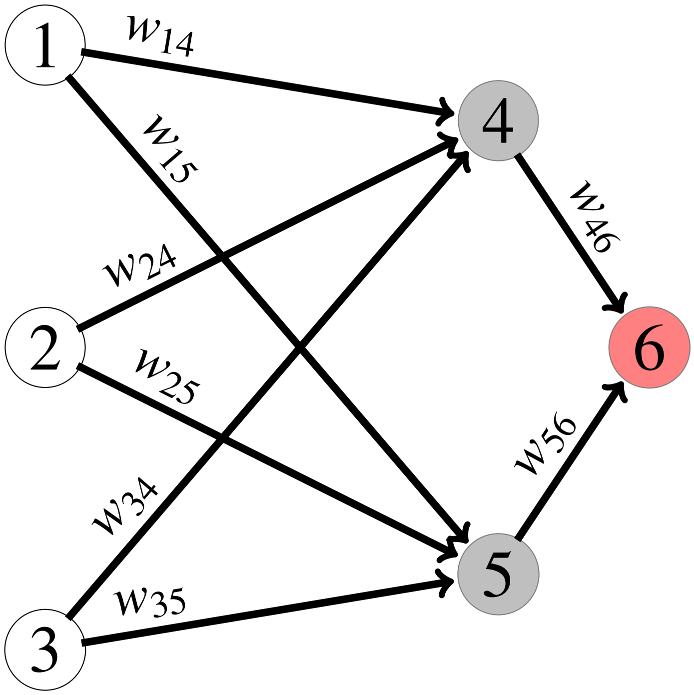
CLASSIFICATION DATA **D** (with objects)

O	a1	a2	a3	C
o1	0.5	0	0.2	1
o2	0	0.3	0	1
o3	0.4	0.1	0.5	0
o4	0.3	0.2	0.2	1
o5	0.1	0	0.4	1
o6	0.4	0.3	0.2	0

1. Design a **full classification** NEURAL NETWORK for **D** with ONE HIDDEN LAYER with 2 NODES

Draw your **full classification** NEURAL NETWORK picture numbering INPUT, HIDDEN LAYER and OUTPUT NODES by numbers 1, 2, ... 6 and specifying corresponding WEIGHTS as in Lecture 9 EXAMPLE

SOLUTION



2. **Perform** only ONE EPOCH **training** with records $o1 - o4$ and with all initial weights and bias set as follows.

$w14 = w15 = w24 = w25 = w34 = 0.500$, $w35 = w46 = w56 = 0.5$, and biases $B4 = B5 = B6 = 0.5$.

Set the learning rate $lr = 0.2$. You should use sigmoid activation function $f(t) = \frac{1}{1+e^{-t}}$ at each output layer.

Present results in the form of tables as in the BOOK and Lecture 9 EXAMPLE.

SOLUTION

Input :

O	a1	a2	a3	C
o1	0.5	0	0.2	1
o2	0	0.3	0	1
o3	0.4	0.1	0.5	0
o4	0.3	0.2	0.2	1

Training: O1

Forward Pass:

Node	Input (I_j)	Output (O_j)
4	$0.5 \times 0.5 + 0 \times 0.5 + 0.2 \times 0.5 + 0.5 = 0.850$	$\frac{1}{1+e^{-0.85}} = 0.701$
5	$0.5 \times 0.5 + 0 \times 0.5 + 0.2 \times 0.5 + 0.5 = 0.850$	$\frac{1}{1+e^{-0.85}} = 0.701$
6	$0.7 \times 0.5 + 0.7 \times 0.5 + 0.5 = 1.201$	$\frac{1}{1+e^{-0.1.2}} = 0.769$

w_{14}	0.500
w_{15}	0.500
w_{24}	0.500
w_{25}	0.500
w_{34}	0.500
w_{35}	0.5
w_{46}	0.5
w_{56}	0.5
θ_4	0.5
θ_5	0.5
θ_6	0.5
lr	0.2

Table 1: Initial Configuration

Error measurement:

Node	$\text{Error}_k(Err_k)$
6	$0.769 \times (1-0.769) \times (1-0.769) = 0.041$
5	$0.701 \times (1-0.701) \times 0.041 \times 0.5 = 0.004$
4	$0.701 \times (1-0.701) \times 0.041 \times 0.5 = 0.004$

New weights :

Node	Updated Weights
θ_6	$0.5 + 0.2 \times 0.041 = 0.508$
θ_5	$0.5 + 0.2 \times 0.004 = 0.501$
θ_4	$0.5 + 0.2 \times 0.004 = 0.501$
w_{56}	$0.5 + 0.2 \times 0.041 \times 0.701 = 0.506$
w_{46}	$0.5 + 0.2 \times 0.041 \times 0.701 = 0.506$
w_{14}	$0.5 + 0.2 \times 0.004 \times 0.5 = 0.500$
w_{15}	$0.5 + 0.2 \times 0.004 \times 0.5 = 0.500$
w_{24}	$0.5 + 0.2 \times 0.004 \times 0 = 0.500$
w_{25}	$0.5 + 0.2 \times 0.004 \times 0 = 0.500$
w_{34}	$0.5 + 0.2 \times 0.004 \times 0.2 = 0.500$
w_{35}	$0.5 + 0.2 \times 0.004 \times 0.2 = 0.500$

Training: O2

Forward Pass:

Node	Input (I_j)	Output (O_j)
4	0.651	0.657
5	0.651	0.657
6	1.173	0.764

Error measurement:

Node	$\text{Error}_k(Err_k)$
6	0.043
5	0.005
4	0.005

New weights :

Node	Updated Weights
θ_6	0.517
θ_5	0.502
θ_4	0.502
w_{56}	0.512 / 0.511
w_{46}	0.512 / 0.511
w_{14}	0.500
w_{15}	0.500
w_{24}	0.500
w_{25}	0.500
w_{34}	0.500
w_{35}	0.500

Training: O3

Forward Pass:

Node	Input (I_j)	Output (O_j)
4	1.002	0.731
5	1.002	0.731
6	1.266	0.78

Error measurement:

Node	$\text{Error}_k(Err_k)$
6	-0.134
5	-0.013
4	-0.013

New weights :

Node	Updated Weights
θ_6	0.490
θ_5	0.499
θ_4	0.499
w_{56}	0.492
w_{46}	0.492
w_{14}	0.499
w_{15}	0.499
w_{24}	0.500
w_{25}	0.500
w_{34}	0.499
w_{35}	0.499

Training: O4

Forward Pass:

Node	Input (I_j)	Output (O_j)
4	0.849	0.700
5	0.849	0.700
6	1.179	0.765

Error measurement:

Node	Error$_k$(Err_k)
6	0.042
5	0.004
4	0.004

New weights :

Node	Updated Weights
θ_6	0.498
θ_5	0.500
θ_4	0.500
w_{56}	0.498
w_{46}	0.498
w_{14}	0.499 / 0.500
w_{15}	0.499 / 0.500
w_{24}	0.500
w_{25}	0.500
w_{34}	0.499
w_{35}	0.499

3. Perform **testing** your trained network with records $o5 - o6$ - specify your criteria of correct/incorrect classification of test records

Use your ONE EPOCH Basic Model (classifier) to classify a record (0, 0.1, 0.2) - write down your criteria of all decisions used in the process

SOLUTION

Testing o6 value > 0.5 **classified as 1,** < 0.5 **classified as 0.**

O	a1	a2	a3	C
o5	0.1	0	0.4	1
o6	0.4	0.3	0.2	0

Testing: O5

Forward Pass:

Node	Input (I$_j$)	Output (O$_j$)
4	0.750	0.679
5	0.750	0.679
6	1.174	0.764

$0.764 > 0.5$ **thus will be classified correctly**

Testing: O6

Forward Pass:

Node	Input (I_j)	Output (O_j)
4	0.949 / 0.950	0.721
5	0.949 / 0.950	0.721
6	1.216	0.771

0.771 > 0.5 **thus will be mis classified as 1**

Testing: (0, 0.1, 0.2)

Forward Pass:

Node	Input (I_j)	Output (O_j)
4	0.650	0.657
5	0.650	0.657
6	1.152	0.76

0.76 > 0.5 **thus will be classified as 1**