CSE 352 QUIZ 2 SOLUTIONS Fall 2018

QUESTION 1

1. Define **classification data**

Data table with one attribute distinguished as a CLASS attribute.Class attribute must have a small number discrete values.

2. Describe **two main** phases of a process of building a **classifier**

Building a classifier consists of two phases: **training** *and* **testing**

We use the training data set to **create patterns***: rules, trees, or to train a Neural or Bayesian network*

We **evaluate** *created patterns with the use of test data*

We **terminate** *the process of building a classifier if it has been trained and tested and the predictive accuracy is on an acceptable level*

3. What is a **classifier**

CLASSIFIER is a "black box", a final product of the above process

QUESTION 2

1. What is a **holdout procedure**

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

2. Describe shortly the main 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. 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 3 (10pts)

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 in the number of INPUT nodes of any NN for **D**?

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

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

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

Which in the number of hidden layers?

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

Which in the number of nodes in the hidden layers?

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

4. Design 3 Neural Networks for the **CLASSIFICATION DATA** from Question 2

One for full classification

There is TWO input nodes, THREE output nodes - at least one hidden layer with number of nods of your choice.

Must draw all **weight with proper indexes** *as i Hmk 3 or Lecture*

Two for contrast learning (for your chosen classes)

There is TWO input nodes, ONE output node - 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

The explanations are above

QUESTION 4 (5pts)

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*