Laboratory Task3

Introduction to Classification Techniques

OBJECTIVES: In this lab, you shall

1. Learn how to apply various data classification techniques

2. Evaluate the performance of your implementation

Note: You may implement your own classification code from scratch or build on available tools.

Task1:

The following table consists of training data from an employee database. The data have been generalized. For example, "31 . . . 35" for age represents the age range of 31 to 35. For a given row entry, count represents the number of data tuples having the values for department, status, age, and salary given in that row.

Department	status	age	salary	count
sales	senior	31..35	46K..50K	30
sales	junior	26..30	26K..30K	40
sales	junior	31..35	31K..35K	40
systems	junior	21..25	46K..50K	20
systems	senior	31..35	66K..70K	5
systems	junior	26..30	46K..50K	3
systems	senior	41..45	66K..70K	3
marketing	senior	36..40	46K..50K	10
marketing	junior	31..35	41K..45K	4
secretary	senior	46..50	36K..40K	4
secretary	junior	26..30	26K..30K	6

Let status be the class-label attribute.

(a) Implement a Decision Tree classification solution

(b) Repeat the same problem using Naïve Bayesian classifier

(c) Evaluate the performance of the two algorithms

Repeat the above problem using ANN:

(d) Design a multilayer feed-forward neural network for the given data. Label the nodes in the input and output layers.

(e) Using the multilayer feed-forward neural network obtained in (a), show the weight values after one iteration of the backpropagation algorithm, given the training instance "(sales, senior, 31..35, 46K..50K)". Indicate your initial weight values and biases and the learning rate used.

Task2:

The MNIST database of handwritten digits, available at http://yann.lecun.com/exdb/mnist/, has a training set of 60,000 examples, of which it includes a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image.

Read the information provided on this site about the content of the dataset.

- Implement the K-nearest neighbor (and/or any other) classification algorithm to recognize handwritten digits.
- Repeat the classification using neural network implementation
- Repeat the classification using SVM
- Discuss the following:
    - Validation method used
    - Selection of training samples
    - Accuracy of your results.