Name : Asghar Hussain

Id : 13461

Degree : BS(SE)

Subject : Data Mining

Module : 7th semester

**Question No 1**. Distinguish between classification and regression with the help of relevant scenarios?

**Answer**. There is an important difference between classification and regression. Classification is about predicting a label and regression is about predicting a quantity.

Classification is the process of finding or discovering a model which helps in separating the data into multiple categorical classes. Regression is the process of finding a model or function for distinguishing the data into continuous real values instead of using classes.

Let's take an example, suppose we want to predict the possibility of the rain in some regions on the basis of some parameters. Then there would be two labels rain and no rain under which different regions can be classified.

Let's take the similar example in regression also, where we are finding the possibility of rain in some particular regions with the help of some parameters. In this case there is a probability associated with the rain. Here we are not classifying the regions within the rain and no rain labels instead we are classifying them with their associated probability.

Here is another scenario, if a classification predictive model made 5 predictions and 3 of them were correct and 2 of them were incorrect, then the classification accuracy of the model based on just these predictions would be:

Accuracy = correct predictions / total predictions

Accuracy = 3/5 * 100

Accuracy = 60%

If a regression predictive model made 2 predictions, one of 1.5 where the expected value is 1.0 and the another of 3.3 and the expected value is 3.0.

**Question No 2**. Perform Naïve bayes or decision tree classification for new instance where (SSN = 123-46-4455, Test1 = 85, Test2 = 31 and final = 30) find grade?

**Answer**. Naïve baye's formula = P(A/B) = P(B/A).P(A)/P(B)

Test1 = 85, Test2 = 31 and final = 30, grade = ?

Let suppose grade = X

P(X/test1) = P(85/100).P(100)/100

= 85/100.100/100

= 0.85

P(X/test2) = P(31/100).P(100)/100

= 31/100.100/100

= 0.31

P(X/final) = P(30/100).P(100)/100

= 30/100.100/100

= 0.30

Hence final grade is "C".

---

**Question No 3**. Find a dataset related to any field and perform several classification techniques (Naïve Bayes, Decision tree, SVM, or any other) to predict a class of a new instance using WEKA. Compare the results (Accuracy, Precision, Recall, MARE, MMRE) of classification algorithms in a table.

**Answer**. I am going to apply naïve bayes algorithms to predict that if a fruit has the following properties then which type of fruit it is.

Fruit = yellow, sweet, long

| Fruit | yellow | sweet | long | Total |
|---|---|---|---|---|
| Mango | 350 | 450 | 0 | 650 |
| Banana | 400 | 300 | 350 | 400 |
| Others | 50 | 100 | 50 | 150 |
| Total | 800 | 850 | 400 | 1200 |

Fruit = {yellow,sweet,long} = x

P(a/b) = p(b/a).p(a)/p(b)

P(yellow/mango) = p(mango/yellow).p(yellow)/p(mango)

= 350/800.800/1200/650/1200 = 0.53

P(s/m) = 0.69

P(l/m) = 0

**Banana**

P(y/b) = 1

P(s/b) = 0.75

P(l/b) = 0.875

P(x/b) = 0.65

**Others**

P(y/o) = 0.33

P(s/o) = 0.66

P(l/o) = 0.33

P(x/o) = 0.072

Hence, the fruit is yellow.

**Decision tree :**

| outlook | temperature | humidity | wind | Play |
|---|---|---|---|---|
| Sunny | Hot | High | False | No |
| Sunny | Hot | High | True | No |
| Overcast | Hot | High | False | Yes |

We have to calculate the entropy with respect to a given predictor in order to be able to calculate information gain.

OUTLOOK = sunny, overcast, hot

Play Golf = yes , no

E(playgolf,outlook) = P(sunny)E(2,3) + P(overcast)E(4,0) + P(rainy)E(3,2)

5/14 0.971 + 4/14 0 + 5/14 0.971 = 0.6936

**Support Vector Machine (SVM):**

A 1D Example,

Suppose we have three data points

X = -3 , y = -1

X = -1 , y = -1

X = 1 , y = 1

Many separating perceptions, T[ax + b]

Anything with ax+b = 0 brtween -1 and 2

We can write the margin constraints

a(-3) + b < -1 b < 3a -1

a(-1) + b < -1 b< a -1

a(2) + b>+1 b > -2a + 1

Ex: a = 1, b = 0

Minimize ||a||

a= .66, b = -.33