|  |
| --- |
| ***Indian institute of space science and technology thiruvananthapuram*** |
| Data Mining |
| Assignment 4 |
|  |
| **VAISAKH S** |
| **SC14M0047** |

|  |
| --- |
|  |

1) Linear Regression With Regularization



Figure 1 : Datapoints being plotted



Figure 2 : J(w) curve -----degree 2



FIGURE 3: DEGREE 2-REGRESSION CURVE



FIGURE 4: TRAINING ERROR , TESTING ERROR PLOT -DEGREE 2



FIGURE 5: COST CURVE -DEGREE 3



FIGURE 6 : REGRESSION CURVE :DEGREE 3



FIGURE 7: TRAINING ERROR TESTING ERROR-DEGREE 3



FIGURE 8: COST CURVE- DEGREE 7



FIGURE 9: COST CURVE -DEGREE 7



FIGURE 10 : DEGREE 7- TRAINING TESTING ERROR

**2) BOSTON DATA --- COST CURVE**



**COST CURVE WITH REGULARIZATION**



**COST CURVE WITHOUT REGULARIZATION**

**3)K NEAREST NEIGHBOURHOOD**



**ROC Curve(With K=5)**

**4) Decision tree using WEKA Software**

Evaluation on test split

Correctly Classified Instances: 5584 (85 .7494 %)  
Incorrectly Classified Instances: 928 (14.2506 %)  
Kappa statistic: 0.5732

K&B Relative Info Score : 288671.614 4  
K&B Information Score : 2308.5563 bits 0. 3545 bits/instance  
Class complexity | order 0: 5100.5935 bits 0.7833 bits/instance  
Class complexity | scheme : 40494.5293 bits 6.2184 bits/instance  
Complexity improvement(Sf ) : -35393.9358 bits -5.4352 bits/instance  
Mean absolute error : 0.1917  
Root mean squared error : 0.3191  
Relative absolute error : 52 .867%  
Root relative squared error : 75 .4903 %  
Total Number of Instances : 6512

Detailed Accuracy By Class

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | WEIGHTED AVG |
| TP RATE | 0.936 | 0.6 | 0.857 |
| FP RATE | 0.4 | 0.064 | 0.322 |
| PRECISION | 0.885 | 0.739 | 0.851 |
| RECALL | 0.936 | 0.6 | 0.857 |
| FMEASURE | 0.91 | 0.662 | 0.852 |
| ROC AREA | 0.89 | 0.89 | 0.89 |
| CLASS | <=50K | >50K | - |

Confusion Matrix

|  |  |
| --- | --- |
| Class <=50K | Class > 50K |
| 4674 | 322 |
| 606 | 910 |

**5)Apriori Algorithm**

Support count of Individual

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| M | O | N | K | E | Y | D | A | U | C | I |
| 3 | 4 | 2 | 5 | 4 | 3 | 1 | 1 | 1 | 2 | 1 |

Since minimum support is 60%

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| M | O | K | E | Y |
| 3 | 4 | 5 | 4 | 3 |

C2 is formed

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| M,O | M,K | M,E | M,Y | O,K | O,E | O,Y | K,E | K,Y | E,Y |
| 1 | 3 | 2 | 2 | 4 | 4 | 2 | 4 | 3 | 2 |

L2 obtained

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| M,K | O,K | O,E | K,E | K,Y |
| 3 | 4 | 4 | 4 | 3 |

C3 formed

|  |
| --- |
| M,K,E |
| 4 |

Rules formed

O,K} *⇒*E  
{O,E}*⇒*K  
{K,E}*⇒*O

[O,K,E]/[O,K] = 4/4 = 1

[O,K,E]/[O,E] = 4/4 = 1

[O,K,E]/[K,E] = 4/4 = 1

All are strong association since confidence>=80%