Week 6 Lab

Joel Huang 1002530

1. Association Mining

2. Features

|  |  |
| --- | --- |
| Name | Feature Type (doing this for practice) |
| checking\_status | Discrete interval, there are 3 distinct range labels, and 1 for no checking. |
| Duration | Continuous ratio |
| credit\_history | Discrete nominal, 5 different credit history classes |
| Purpose | Discrete nominal, 11 different purposes |
| credit\_amount | Continuous interval |
| savings\_status | Discrete interval, there are 4 distinct ranges and 1 for unknown |
| Employment | Discrete interval, there are 4 distinct ranges and 1 for unknown |
| installment\_commitment | Discrete interval, ranges from 1-4 |
| Personal\_status | Discrete nominal, there are 5 different labels |
| other\_parties | Discrete nominal, there are 3 different labels |
| residence\_since | Discrete interval, ranges from 1-4 |
| Property\_magnitude | Discrete nominal, there are 4 different labels |
| Age | Discrete interval, ages can actually be considered multiplicatively but there’s not much point... |
| other\_payment\_plans | Discrete nominal, 3 different labels |
| Housing | Discrete nominal, 3 different labels |
| existing\_credits | Discrete interval, ranges from 1-4 |
| Job | Discrete ordinal, unemployed<unskilled<skilled<high qualified |
| num\_dependents | Binary interval, contains either 1 or 2 dependents |
| own\_telephone | Binary nominal, contains either 0 or >0 own telephones |
| foreign\_worker | Binary nominal, either foreign or not |
| class | Binary nominal, either good or bad |

3. The continuous features should first be discretized.

4.

|  |
| --- |
| 1. residence\_since='All' 1000 ==> installment\_commitment='All' 1000 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)  2. installment\_commitment='All' 1000 ==> residence\_since='All' 1000 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)  3. age='All' 1000 ==> installment\_commitment='All' 1000 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)  4. installment\_commitment='All' 1000 ==> age='All' 1000 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)  5. existing\_credits='All' 1000 ==> installment\_commitment='All' 1000 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)  6. installment\_commitment='All' 1000 ==> existing\_credits='All' 1000 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)  7. num\_dependents='All' 1000 ==> installment\_commitment='All' 1000 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)  8. installment\_commitment='All' 1000 ==> num\_dependents='All' 1000 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)  9. age='All' 1000 ==> residence\_since='All' 1000 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)  10. residence\_since='All' 1000 ==> age='All' 1000 <conf:(1)> lift:(1) lev:(0) [0] conv:(0) |

5.

|  |
| --- |
| Best rules found:  1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723 <conf:(0.92)> lift:(1.27) lev:(0.03) [155] conv:(3.35)  2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 <conf:(0.92)> lift:(1.27) lev:(0.03) [149] conv:(3.28)  3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705 <conf:(0.92)> lift:(1.27) lev:(0.03) [150] conv:(3.27)  4. biscuits=t fruit=t vegetables=t total=high 815 ==> bread and cake=t 746 <conf:(0.92)> lift:(1.27) lev:(0.03) [159] conv:(3.26)  5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 779 <conf:(0.91)> lift:(1.27) lev:(0.04) [164] conv:(3.15)  6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725 <conf:(0.91)> lift:(1.26) lev:(0.03) [151] conv:(3.06)  7. baking needs=t biscuits=t vegetables=t total=high 772 ==> bread and cake=t 701 <conf:(0.91)> lift:(1.26) lev:(0.03) [145] conv:(3.01)  8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866 <conf:(0.91)> lift:(1.26) lev:(0.04) [179] conv:(3)  9. frozen foods=t fruit=t vegetables=t total=high 834 ==> bread and cake=t 757 <conf:(0.91)> lift:(1.26) lev:(0.03) [156] conv:(3)  10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877 <conf:(0.91)> lift:(1.26) lev:(0.04) [179] conv:(2.92) |

1. Kmeans clustering

2. SSE is high and constant. 87.32963569427773

3. With feature norm, SSE drops to 7.817456892309574