Intelligent Data Analysis and Probabilistic Inference Coursework Marking Scheme

The mark scheme is given in the coursework handout. The details of how the marks should be awarded are as follows:

1. A partially correct attempt at part 1 will gain a pass mark in the range 50-55.

A reasonable attempt at the first part even if the results are wrong should obtain a pass mark 50%

2. A correct solution to part 1 will obtain a mark in the range 55-65

This would mean that the calculated dependencies for the L1 metric agree with the sample solution and the trees are drawn correctly.

Award 57 for groups of 3
60 for groups of 2
63 for individual solutions

±2 for minor errors or particular clarity
-3 for incorrect trees
-3 for incorrect numeric values

3. A correct solution to both 1 and 2 will obtain a mark in the range 65-75

This would mean that the calculated dependencies for both the L1 and the Mutual Entropy metrics agree with the sample solution and the trees are drawn correctly.

Award 67 for groups of 3
70 for groups of 2
73 for individual solutions

penalties/bonuses as above

4. A correct solution to 1 2 and 3a will obtain a mark in the range 75-85

This requires the same achievement as part 3 together with correctly computed conditional probability matrices.

Award 77 for groups of 3
80 for groups of 2
83 for individual solutions

penalties/bonuses as above

NB because the joint conditional probability matrix in the hepatitis C data set is very large (81 columns) I've told some students they need only print out the first few columns.

5. A correct solution to 1 2 3 and 4 will obtain a mark in the range 85-100

This requires the same achievement as 4 above, plus calculating the MDL metric and experimenting with deleting at least one arc and checking if the MDL score increases or decreases on at least one data set. One group this year implemented a tree search to find the most accurate network. Doing anything like that should receive a very high mark.

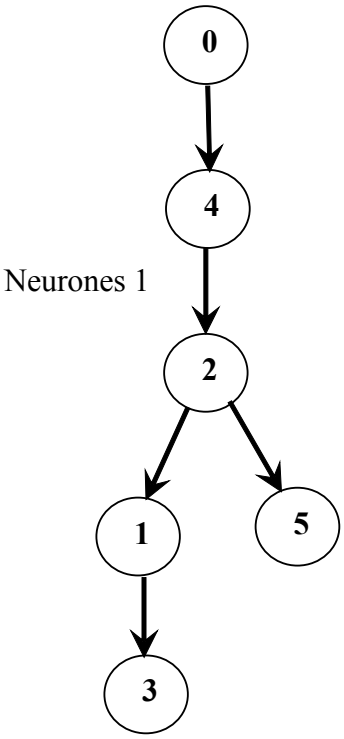
Award 87 for groups of 3
90 for groups of 2
95 for individual solutions
penalties as above, and bonuses depending on the amount of work undertaken.

NB the coursework handout asks for the code listing to be submitted electronically. However, some of the students (especially from other departments) may not have done this. They should not be penalised if they do not. There is no need to check the listings.

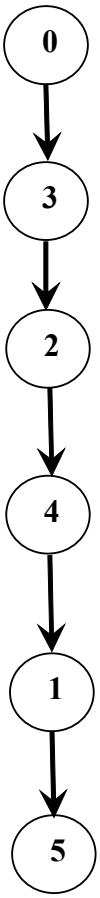
My sample solutions are complete as far as part 2. I've only calculated the MDL score for the spanning tree found in part 2. It is safe to assume that the majority of the students who calculate other networks in part 3 will get the answer right. I'm fairly sure the MDL figures are correct but some students seem to have got different results this year, if there is a consensus against me let me know as they may be right!

The expectation is that most students will score B - A+.

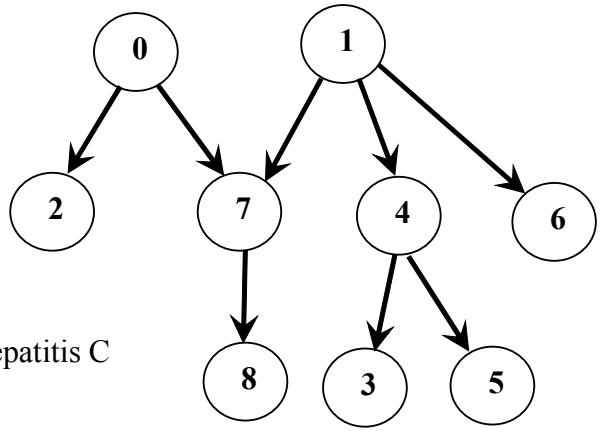
Return date: Tuesday 9th December if possible



Neurones 1



Neurones 2



Hepatitis C

NB Some solutions will number the nodes from 1 rather than 0

HepatitisC.txt

L1 Dependency Matrix:

```
1.0572 0.1477 0.138  0.1185 0.1381 0.1214 0.1243 0.2106 0.0987
0.1477 1.5384 0.0872 0.1594 0.2012 0.1322 0.2381 0.255  0.0418
0.138  0.0872 0.9892 0.097  0.0605 0.0305 0.0687 0.0635 0.0219
0.1185 0.1594 0.097  1.1699 0.7183 0.5138 0.101  0.0987 0.0351
0.1381 0.2012 0.0605 0.7183 1.501  0.7885 0.1296 0.1819 0.0622
0.1214 0.1322 0.0305 0.5138 0.7885 1.259  0.0885 0.1351 0.1067
0.1243 0.2381 0.0687 0.101  0.1296 0.0885 1.576  0.1789 0.0257
0.2106 0.255  0.0635 0.0987 0.1819 0.1351 0.1789 1.0318 0.1196
0.0987 0.0418 0.0219 0.0351 0.0622 0.1067 0.0257 0.1196 0.6785
```

Mutual Entropy Dependency Matrix:

```
1.5298 0.0446 0.0257 0.048  0.034  0.0237 0.0391 0.0861 0.0157
0.0446 2.372  0.0094 0.0604 0.0688 0.0303 0.0714 0.0826 0.0027
0.0257 0.0094 0.9922 0.0117 0.0072 0.0016 0.0070 0.0049 5.0E-4
0.048  0.0604 0.0117 1.6972 0.5389 0.2745 0.0316 0.0324 0.0058
0.034  0.0688 0.0072 0.5389 2.4112 0.6058 0.0406 0.0505 0.0084
0.0237 0.0303 0.0016 0.2745 0.6058 1.8318 0.0251 0.0414 0.0161
0.0391 0.0714 0.0070 0.0316 0.0406 0.0251 2.6292 0.0629 0.0038
0.0861 0.0826 0.0049 0.0324 0.0505 0.0414 0.0629 1.4888 0.032
0.0157 0.0027 5.0E-4 0.0058 0.0084 0.0161 0.0038 0.032  0.7537
```

Prior probabilities node 0

```
0.651  0.188  0.038  0.01   0.00   0.00   0.00   0.00   0.102
```

Prior probabilities node 1

```
0.028  0.126  0.334  0.091  0.117  0.018  0.00   0.0    0.284
```

Transposed link matrix from node 2 to parent(s) 0
The columns are the states of the child node

```
0.511  0.489
0.656  0.344
0.778  0.222
0.813  0.188
1.0    0.0
0.333  0.667
0.833  0.167
1.0    0.0
0.482  0.518
```

Transposed link matrix from node 3 to parent(s) 4
The columns are the states of the child node

```
0.838  0.0    0.054  0.0    0.0    0.0    0.0    0.0    0.027  0.081
0.274  0.453  0.039  0.00   0.034  0.0    0.00   0.0    0.017  0.173
0.339  0.288  0.036  0.00   0.00   0.0    0.0    0.00   0.015  0.297
0.447  0.071  0.082  0.047  0.00   0.0    0.0    0.00   0.018  0.324
0.453  0.015  0.088  0.022  0.00   0.0    0.00   0.022  0.015  0.372
0.333  0.012  0.099  0.0    0.0    0.0    0.037  0.0    0.012  0.506
0.5    0.0    0.042  0.0    0.0    0.0    0.042  0.0    0.0    0.417
0.2    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.8
0.00   0.00   0.00   0.0    0.0    0.00   0.0    0.0    0.0    0.983
```

Transposed link matrix from node 4 to parent(s) 1

The columns are the states of the child node

0.0 0.196 0.152 0.043 0.022 0.043 0.0 0.0 0.543
0.014 0.186 0.205 0.095 0.071 0.048 0.029 0.0 0.352
0.029 0.061 0.165 0.109 0.093 0.045 0.013 0.00 0.484
0.033 0.079 0.125 0.105 0.164 0.086 0.026 0.00 0.375
0.0 0.138 0.344 0.051 0.031 0.036 0.00 0.0 0.395
0.0 0.267 0.133 0.033 0.0 0.033 0.0 0.0 0.533
0.0 0.0 0.333 0.0 0.5 0.167 0.0 0.0 0.0
0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111
0.027 0.105 0.208 0.126 0.074 0.046 0.013 0.00 0.394

Transposed link matrix from node 5 to parent(s) 4
The columns are the states of the child node

0.0 0.108 0.432 0.162 0.081 0.135 0.081
0.101 0.369 0.196 0.056 0.022 0.022 0.235
0.303 0.279 0.156 0.036 0.012 0.0 0.213
0.306 0.288 0.124 0.018 0.00 0.0 0.259
0.423 0.285 0.088 0.00 0.0 0.0 0.197
0.519 0.123 0.012 0.0 0.0 0.0 0.346
0.375 0.0 0.0 0.083 0.0 0.0 0.542
0.4 0.0 0.0 0.0 0.0 0.0 0.6
0.00 0.00 0.00 0.0 0.00 0.0 0.99

Transposed link matrix from node 6 to parent(s) 1
The columns are the states of the child node

0.283 0.065 0.043 0.13 0.087 0.0 0.065 0.043 0.0 0.022 0.022 0.239
0.443 0.143 0.038 0.086 0.038 0.0 0.038 0.024 0.0 0.014 0.01 0.167
0.341 0.194 0.022 0.093 0.036 0.0 0.034 0.039 0.0 0.016 0.018 0.208
0.283 0.25 0.02 0.099 0.033 0.013 0.026 0.033 0.0 0.00 0.02 0.217
0.39 0.195 0.031 0.077 0.036 0.00 0.041 0.062 0.0 0.0 0.021 0.144
0.467 0.167 0.033 0.133 0.0 0.0 0.067 0.0 0.0 0.0 0.0 0.133
0.667 0.0 0.0 0.0 0.0 0.0 0.0 0.167 0.0 0.0 0.0 0.167
0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
0.24 0.149 0.017 0.053 0.044 0.00 0.034 0.023 0.00 0.015 0.00 0.402

Transposed link matrix from node 7 to parent(s) 0 1
The columns are the states of the child node

0.074 0.074 0.0 0.0 0.074 0.778
0.21 0.081 0.00 0.032 0.04 0.629
0.308 0.13 0.018 0.023 0.049 0.472
0.276 0.038 0.0 0.01 0.029 0.648
0.192 0.051 0.0 0.03 0.051 0.677
0.294 0.118 0.0 0.059 0.0 0.529
0.667 0.0 0.0 0.167 0.0 0.167
0.167 0.167 0.167 0.167 0.167 0.167
0.071 0.092 0.018 0.018 0.015 0.785
0.0 0.0 0.0 0.067 0.0 0.933
0.041 0.041 0.0 0.0 0.0 0.918
0.216 0.114 0.0 0.0 0.0 0.67
0.138 0.034 0.0 0.0 0.0 0.828
0.121 0.034 0.0 0.0 0.0 0.845
0.667 0.0 0.0 0.0 0.0 0.333
0.167 0.167 0.167 0.167 0.167 0.167
0.167 0.167 0.167 0.167 0.167 0.167
0.015 0.015 0.0 0.0 0.0 0.97
0.0 0.0 0.0 0.0 0.0 1.0
0.25 0.0 0.0 0.0 0.0 0.75
0.0 0.118 0.0 0.0 0.0 0.882
0.0 0.0 0.0 0.0 0.0 1.0

```
0.0    0.071  0.0    0.0    0.0    0.929
1.0    0.0    0.0    0.0    0.0    0.0
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.133  0.0    0.0    0.0    0.867
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.0    0.0    0.0    0.0    1.0
0.0    0.143  0.0    0.0    0.0    0.857
0.0    0.0    0.0    0.0    0.0    1.0
0.0    0.0    0.0    0.0    0.0    1.0
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.25   0.0    0.0    0.0    0.75
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.0    0.0    0.0    0.0    1.0
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.0    0.0    0.0    0.0    1.0
0.0    0.0    0.0    0.0    0.0    1.0
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.0    0.0    0.0    0.0    1.0
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.0    0.0    0.0    0.0    1.0
0.0    0.0    0.0    0.0    0.0    1.0
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.0    0.0    0.0    0.0    1.0
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.25   0.0    0.0    0.0    0.75
0.0    0.0    0.0    0.0    0.0    1.0
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.0    0.0    0.0    0.0    1.0
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.0    0.0    0.0    0.0    0.0    1.0
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.0    1.0    0.0    0.0    0.0    0.0
0.059  0.471  0.0    0.0    0.0    0.471
0.086  0.569  0.0    0.0    0.0    0.345
0.125  0.125  0.0    0.0    0.0    0.75
0.19   0.333  0.0    0.0    0.0    0.476
0.0    0.5    0.0    0.0    0.0    0.5
0.167  0.167  0.167  0.167  0.167  0.167
0.167  0.167  0.167  0.167  0.167  0.167
0.113  0.081  0.129  0.048  0.0    0.629


Transposed link matrix from node 8 to parent(s) 7
The columns are the states of the child node

0.857  0.143
0.94   0.06
```

```
0.682  0.318
0.759  0.241
0.436  0.564
0.753  0.247

MDL Size   3972.43
MDL Accuracy   -23672.18
MDL Score   27644.61
```

Neurones1.txt

L1 Dependency Matrix:

1.4587 0.7986 1.1405 0.7903 1.1426 0.564
0.7986 1.6983 0.9587 0.9483 0.9452 0.5496
1.1405 0.9587 1.7128 0.9318 1.343 0.6033
0.7903 0.9483 0.9318 1.6446 0.9318 0.4731
1.1426 0.9452 1.343 0.9318 1.5289 0.5093
0.564 0.5496 0.6033 0.4731 0.5093 1.3616

Mutual Entropy Dependency Matrix:

Prior probabilities node 0

0.159 0.295 0.205 0.341

Transposed link matrix from node 1 to parent(s) 2
The columns are the states of the child node

0.143 0.143 0.286 0.143 0.143 0.0   0.0   0.143
0.4   0.2   0.4   0.0   0.0   0.0   0.0   0.0
1.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.333 0.0   0.0   0.167 0.167 0.167 0.167 0.0
0.0   0.111 0.0   0.111 0.111 0.333 0.222 0.111
0.0   0.0   0.0   0.167 0.0   0.0   0.667 0.167
0.0   0.0   0.0   0.0   0.0   0.0   0.4   0.6
0.0   0.0   0.2   0.0   0.0   0.0   0.2   0.6

Transposed link matrix from node 2 to parent(s) 4
The columns are the states of the child node

0.583 0.333 0.083 0.0   0.0   0.0   0.0   0.0
0.0   0.2   0.0   0.6   0.2   0.0   0.0   0.0
0.0   0.0   0.0   0.5   0.5   0.0   0.0   0.0
0.0   0.0   0.0   0.0   1.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   1.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.333 0.333 0.333

Transposed link matrix from node 3 to parent(s) 1
The columns are the states of the child node

0.333 0.667 0.0   0.0   0.0   0.0
0.0   0.667 0.333 0.0   0.0   0.0
0.2   0.4   0.2   0.2   0.0   0.0
0.25   0.0   0.5   0.25   0.0   0.0
0.333 0.0   0.667 0.0   0.0   0.0
0.25   0.25   0.0   0.5   0.0   0.0
0.0   0.0   0.3   0.3   0.2   0.2
0.111 0.0   0.111 0.0   0.222 0.556

Transposed link matrix from node 4 to parent(s) 0
The columns are the states of the child node

1.0   0.0   0.0   0.0   0.0   0.0
0.385 0.231 0.154 0.154 0.077 0.0
0.0   0.222 0.444 0.333 0.0   0.0
0.0   0.0   0.0   0.0   0.0   1.0

Transposed link matrix from node 5 to parent(s) 2

```
The columns are the states of the child node

0.0    0.0    0.0    0.571  0.286  0.143
0.0    0.0    0.2    0.2    0.2    0.4
0.0    0.0    0.0    0.0    0.0    1.0
0.0    0.0    0.0    0.5    0.333  0.167
0.0    0.0    0.111  0.333  0.333  0.222
0.0    0.0    0.0    0.0    0.333  0.667
0.0    0.0    0.0    0.0    0.6    0.4
0.0    0.0    0.0    0.0    0.0    1.0
```

MDL Size 548.67
MDL Accuracy -381.26
MDL Score 929.93

```
Neurones2.txt
Mutual Entropy

Dependency Matrix:

1.9098 0.7709 0.8582 1.0731 0.8779 0.2677
0.7709 2.9328 0.7465 0.7804 0.8809 0.5416
0.8582 0.7465 2.6274 1.0954 1.1766 0.3976
1.0731 0.7804 1.0954 2.3474 1.0225 0.3497
0.8779 0.8809 1.1766 1.0225 2.1028 0.3309
0.2677 0.5416 0.3976 0.3497 0.3309 2.1118


Prior probabilities node 0

0.213  0.393  0.246  0.148


Transposed link matrix from node 1 to parent(s) 4
The columns are the states of the child node

0.333  0.125  0.208  0.125  0.083  0.083  0.042  0.0
0.0    0.111  0.278  0.222  0.278  0.056  0.056  0.0
0.0    0.0    0.143  0.143  0.571  0.0    0.0    0.143
0.0    0.0    0.0    0.333  0.0    0.667  0.0    0.0
0.0    0.0    0.0    0.0    0.0    1.0    0.0    0.0
0.0    0.0    0.0    0.0    0.0    0.0    0.5    0.5


Transposed link matrix from node 2 to parent(s) 3
The columns are the states of the child node

0.75   0.0    0.25   0.0    0.0    0.0    0.0    0.0
0.611  0.167  0.167  0.056  0.0    0.0    0.0    0.0
0.105  0.211  0.579  0.105  0.0    0.0    0.0    0.0
0.0    0.0    0.286  0.571  0.0    0.0    0.143  0.0
0.0    0.0    0.0    0.167  0.5    0.333  0.0    0.0
0.0    0.0    0.0    0.0    0.286  0.286  0.286  0.143


Transposed link matrix from node 3 to parent(s) 0
The columns are the states of the child node

0.308  0.615  0.077  0.0    0.0    0.0
0.0    0.417  0.583  0.0    0.0    0.0
0.0    0.0    0.267  0.4    0.267  0.067
0.0    0.0    0.0    0.111  0.222  0.667


Transposed link matrix from node 4 to parent(s) 2
The columns are the states of the child node

0.938  0.063  0.0    0.0    0.0    0.0
0.714  0.286  0.0    0.0    0.0    0.0
0.176  0.706  0.118  0.0    0.0    0.0
0.125  0.375  0.375  0.125  0.0    0.0
0.0    0.0    0.4    0.4    0.0    0.2
0.0    0.0    0.0    0.0    0.25   0.75
0.0    0.0    0.0    0.0    0.0    1.0
0.0    0.0    0.0    0.0    0.0    1.0


Transposed link matrix from node 5 to parent(s) 1
The columns are the states of the child node
```

```
0.0    0.125  0.25   0.25   0.375  0.0
0.0    0.4    0.2    0.2    0.0    0.2
0.091  0.0    0.273  0.455  0.182  0.0
0.0    0.0    0.222  0.444  0.222  0.111
0.0    0.0    0.0    0.455  0.455  0.091
0.0    0.0    0.0    0.167  0.667  0.167
0.0    0.0    0.0    0.667  0.167  0.167
0.0    0.0    0.0    0.0    0.6    0.4
```

MDL Size 554.52
MDL Accuracy -565.13
MDL Score 1119.66