Week 5 : 8-12 February, 2021

Lab Assignment 5

Problem Statement:

A table containing grades earned by students in respective courses is made available to you in (codes folder) 2020_bn_nb_data.txt.

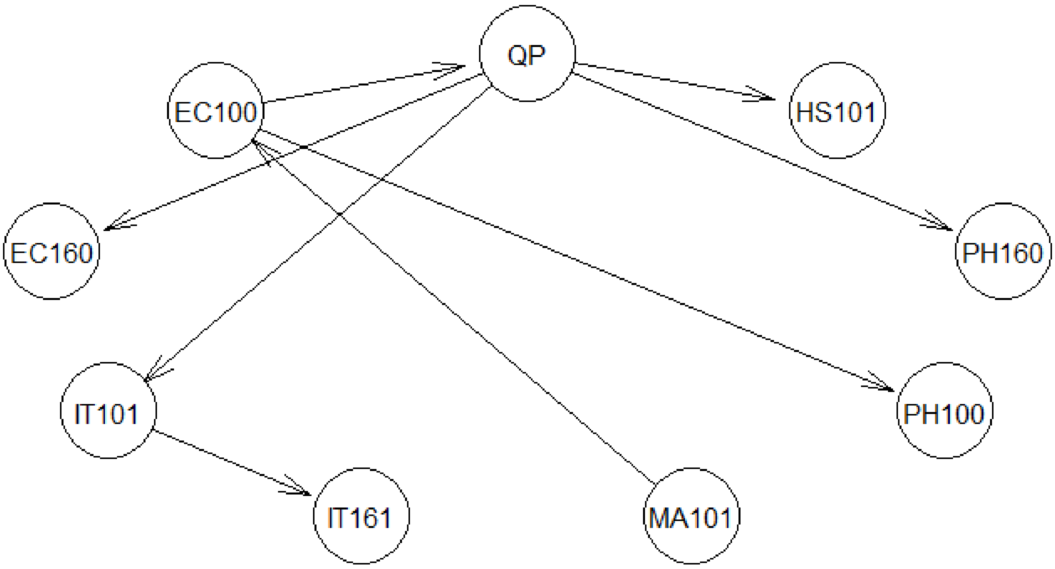
1. **Consider grades earned in each of the courses as random variables and learn the dependencies between courses.**

```
> library(bnlearn)
> library(caret)
> library(e1071)
> mydata <- read.table("C:/Users/ADMIN/Desktop/AI Codes/Week_5/2020_bn_nb_data.txt", header=TRUE)
> mydata.net<-hc(mydata,score="k2",start=NULL)
> print(mydata.net)

  Bayesian network learned via Score-based methods

  model:
   [MA101][EC100|MA101][PH100|EC100][QP|EC100][EC160|QP][IT101|QP][PH160|QP][HS101|QP][IT161|IT101]
  nodes:                                 9
  arcs:                                  8
    undirected arcs:                     0
    directed arcs:                       8
  average markov blanket size:           1.78
  average neighbourhood size:            1.78
  average branching factor:              0.89

  learning algorithm:                    Hill-Climbing
  score:                                 Cooper & Herskovits' K2
  tests used in the learning procedure:  168
  optimized:                             TRUE
```



2. Using the data, learn the CPTs for each course node.

```
> library(bnlearn)
> library(caret)
> library(e1071)
> mydata <- read.table("C:/Users/ADMIN/Desktop/AI Codes/Week_5/2020_bn_nb_data.txt", header=TRUE)
> mydata.net<-hc(mydata,score="k2",start=NULL)
> mydata.net.fit<-bn.fit(mydata.net, mydata)
> print(mydata.net.fit)

  Bayesian network parameters

  Parameters of node EC100 (multinomial distribution)

Conditional probability table:
 
     MA101
EC100         AA         AB         BB         BC         CC         CD         DD          F
   AA 0.75000000 0.07692308 0.03846154 0.01851852 0.00000000 0.00000000 0.00000000 0.00000000
   AB 0.00000000 0.46153846 0.25000000 0.05555556 0.00000000 0.00000000 0.00000000 0.00000000
   BB 0.25000000 0.23076923 0.32692308 0.22222222 0.04081633 0.00000000 0.00000000 0.00000000
   BC 0.00000000 0.15384615 0.28846154 0.27777778 0.32653061 0.00000000 0.00000000 0.00000000
   CC 0.00000000 0.07692308 0.09615385 0.24074074 0.32653061 0.04166667 0.00000000 0.00000000
   CD 0.00000000 0.00000000 0.00000000 0.12962963 0.26530612 0.33333333 0.04761905 0.00000000
   DD 0.00000000 0.00000000 0.00000000 0.03703704 0.04081633 0.50000000 0.19047619 0.00000000
   F  0.00000000 0.00000000 0.00000000 0.01851852 0.00000000 0.12500000 0.76190476 1.00000000

  Parameters of node EC160 (multinomial distribution)

Conditional probability table:
 
     QP
EC160          n          y
   AA 0.00000000 0.07500000
   AB 0.00000000 0.10000000
   BB 0.01388889 0.18750000
   BC 0.01388889 0.36250000
   CC 0.15277778 0.22500000
   CD 0.44444444 0.03125000
   DD 0.26388889 0.01875000
   F  0.11111111 0.00000000

  Parameters of node IT101 (multinomial distribution)
```

```
Conditional probability table:

     QP
IT101          n          y
   AA 0.00000000 0.07500000
   AB 0.00000000 0.15625000
   BB 0.04166667 0.19375000
   BC 0.02777778 0.29375000
   CC 0.13888889 0.20000000
   CD 0.30555556 0.08125000
   DD 0.31944444 0.00000000
   F  0.16666667 0.00000000

  Parameters of node IT161 (multinomial distribution)

Conditional probability table:

     IT101
IT161         AA         AB         BB         BC         CC         CD         DD          F
   AA 0.58333333 0.24000000 0.14705882 0.04081633 0.00000000 0.00000000 0.00000000 0.00000000
   AB 0.16666667 0.40000000 0.29411765 0.02040816 0.04761905 0.00000000 0.00000000 0.00000000
   BB 0.16666667 0.24000000 0.32352941 0.20408163 0.11904762 0.02857143 0.00000000 0.00000000
   BC 0.08333333 0.04000000 0.20588235 0.36734694 0.38095238 0.17142857 0.00000000 0.00000000
   CC 0.00000000 0.04000000 0.00000000 0.24489796 0.33333333 0.31428571 0.08695652 0.16666667
   CD 0.00000000 0.04000000 0.02941176 0.10204082 0.09523810 0.31428571 0.52173913 0.08333333
   DD 0.00000000 0.00000000 0.00000000 0.02040816 0.02380952 0.14285714 0.39130435 0.58333333
   F  0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.02857143 0.00000000 0.16666667

  Parameters of node MA101 (multinomial distribution)

Conditional probability table:
        AA         AB         BB         BC         CC         CD         DD          F
0.01724138 0.05603448 0.22413793 0.23275862 0.21120690 0.10344828 0.09051724 0.06465517

  Parameters of node PH100 (multinomial distribution)

Conditional probability table:

     EC100
PH100         AA         AB         BB         BC         CC         CD         DD          F
   AA 0.71428571 0.40909091 0.22857143 0.08333333 0.00000000 0.00000000 0.00000000 0.00000000
   AB 0.14285714 0.31818182 0.20000000 0.18750000 0.05555556 0.00000000 0.00000000 0.00000000
   BB 0.00000000 0.18181818 0.31428571 0.29166667 0.13888889 0.03448276 0.05000000 0.00000000
   BC 0.14285714 0.04545455 0.14285714 0.22916667 0.33333333 0.13793103 0.00000000 0.00000000
   CC 0.00000000 0.04545455 0.11428571 0.18750000 0.25000000 0.41379310 0.20000000 0.02857143
   CD 0.00000000 0.00000000 0.00000000 0.02083333 0.19444444 0.31034483 0.45000000 0.11428571
   DD 0.00000000 0.00000000 0.00000000 0.00000000 0.02777778 0.10344828 0.20000000 0.45714286
   F  0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.10000000 0.40000000

  Parameters of node PH160 (multinomial distribution)
```

```
Conditional probability table:

     QP
PH160          n          y
   AA 0.05555556 0.14375000
   AB 0.09722222 0.15625000
   BB 0.02777778 0.17500000
   BC 0.18055556 0.34375000
   CC 0.29166667 0.13750000
   CD 0.19444444 0.04375000
   DD 0.12500000 0.00000000
   F  0.02777778 0.00000000

  Parameters of node HS101 (multinomial distribution)

Conditional probability table:

     QP
HS101          n          y
   AA 0.00000000 0.26250000
   AB 0.00000000 0.21250000
   BB 0.05555556 0.22500000
   BC 0.12500000 0.16875000
   CC 0.18055556 0.08125000
   CD 0.19444444 0.03750000
   DD 0.37500000 0.01250000
   F  0.06944444 0.00000000

  Parameters of node QP (multinomial distribution)

Conditional probability table:

   EC100
QP         AA        AB        BB        BC        CC        CD        DD         F
  n 0.0000000 0.0000000 0.0000000 0.0000000 0.1388889 0.4482759 0.9500000 1.0000000
  y 1.0000000 1.0000000 1.0000000 1.0000000 0.8611111 0.5517241 0.0500000 0.0000000

>
```

3. What grade will a student get in PH100 if he earns DD in EC100, CC in IT101 and CD in MA101.

```
> cpquery(mydata.net.fit, event = ( PH100 == "CD" ), evidence = ( EC100 == "DD" & IT101 == "CC" & MA101 == "CD" ))
[1] 0.45
>
```

4. **The last column in the data file indicates whether a student qualifies for an internship program or not. From the given data, take 70 percent data for training and build a naive Bayes classifier (considering that the grades earned in different courses are independent of each other) which takes in the student's performance and returns the qualification status with a probability. Test your classifier on the remaining 30 percent data. Repeat this experiment for 20 random selection of training and testing data. Report results about the accuracy of your classifier.**

```
> data_train <- data1[random == 1, ]
> data_test <- data1[random == 2, ]
> data_nb <- naiveBayes(QP ~ . , data = data_train)
> print(data_nb)

Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
        n         y 
0.2830189 0.7169811 

Conditional probabilities:
   EC100
Y           AA         AB         BB         BC         CC         CD         DD          F
  n 0.00000000 0.00000000 0.00000000 0.00000000 0.06666667 0.11111111 0.31111111 0.51111111
  y 0.06140351 0.16666667 0.21929825 0.28947368 0.16666667 0.08771930 0.00877193 0.00000000

   EC160
Y           AA         AB         BB         BC         CC         CD         DD          F
  n 0.00000000 0.00000000 0.02222222 0.00000000 0.17777778 0.44444444 0.22222222 0.13333333
  y 0.09649123 0.10526316 0.18421053 0.32456140 0.22807018 0.03508772 0.02631579 0.00000000

   IT101
Y           AA         AB         BB         BC         CC         CD         DD          F
  n 0.00000000 0.00000000 0.06666667 0.02222222 0.13333333 0.28888889 0.31111111 0.17777778
  y 0.08771930 0.16666667 0.19298246 0.27192982 0.20175439 0.07894737 0.00000000 0.00000000

   IT161
Y           AA         AB         BB         BC         CC         CD         DD          F
  n 0.00000000 0.02222222 0.02222222 0.04444444 0.22222222 0.28888889 0.37777778 0.02222222
  y 0.10526316 0.17543860 0.18421053 0.30701754 0.16666667 0.05263158 0.00877193 0.00000000

   MA101
Y           AA         AB         BB         BC         CC         CD         DD          F
  n 0.00000000 0.00000000 0.00000000 0.08888889 0.13333333 0.24444444 0.35555556 0.17777778
  y 0.03508772 0.10526316 0.35087719 0.26315789 0.22807018 0.01754386 0.00000000 0.00000000

   PH100
Y           AA         AB         BB         BC         CC         CD         DD          F
  n 0.00000000 0.00000000 0.02222222 0.00000000 0.11111111 0.31111111 0.33333333 0.22222222
  y 0.19298246 0.18421053 0.21052632 0.20175439 0.14035088 0.07017544 0.00000000 0.00000000

   PH160
Y           AA         AB         BB         BC         CC         CD         DD          F
  n 0.06666667 0.15555556 0.00000000 0.13333333 0.28888889 0.17777778 0.13333333 0.04444444
  y 0.14912281 0.14912281 0.14912281 0.35964912 0.15789474 0.03508772 0.00000000 0.00000000

   HS101
Y           AA         AB         BB         BC         CC         CD         DD          F
  n 0.00000000 0.00000000 0.02222222 0.13333333 0.20000000 0.15555556 0.40000000 0.08888889
  y 0.28070175 0.22807018 0.21929825 0.13157895 0.07894737 0.04385965 0.01754386 0.00000000
```

```
> pred_nb <- predict(data_nb, data_test)
> confusionMatrix(table(pred_nb, data_test$QP))
Confusion Matrix and Statistics


pred_nb  n  y
      n 24  0
      y  3 46

               Accuracy : 0.9589
                 95% CI : (0.8846, 0.9914)
    No Information Rate : 0.6301
    P-Value [Acc > NIR] : 3.09e-11

                  Kappa : 0.9098

 Mcnemar's Test P-Value : 0.2482

            Sensitivity : 0.8889
            Specificity : 1.0000
         Pos Pred Value : 1.0000
         Neg Pred Value : 0.9388
             Prevalence : 0.3699
         Detection Rate : 0.3288
   Detection Prevalence : 0.3288
      Balanced Accuracy : 0.9444

       'Positive' Class : n

> |
```

5. **Repeat 4, considering that the grades earned in different courses may be dependent.**

```
>
> data_train <- data1[random == 1, ]
> data_test <- data1[random == 2, ]
>
> tn <- tan_cl('QP', data_train)
> tn<- lp(tn,data_train, smooth =1)
> pred_nb_2 <- predict(tn, data_test)
> accuracy(pred_nb_2 , data_test $QP)
[1] 0.890411
> confusionMatrix(table(pred_nb_2, data_test $QP))
Confusion Matrix and Statistics


pred_nb_2  n  y
        n 20  1
        y  7 45

               Accuracy : 0.8904
                 95% CI : (0.7954, 0.9515)
    No Information Rate : 0.6301
    P-Value [Acc > NIR] : 5.405e-07

                  Kappa : 0.7536

 Mcnemar's Test P-Value : 0.0771

            Sensitivity : 0.7407
            Specificity : 0.9783
         Pos Pred Value : 0.9524
         Neg Pred Value : 0.8654
             Prevalence : 0.3699
         Detection Rate : 0.2740
   Detection Prevalence : 0.2877
      Balanced Accuracy : 0.8595

       'Positive' Class : n

> |
```