CSE 574

Machine Learning

Prof: Dr. S Srihari

Project 1

Jruvika Bhimani

UBitName: jruvikam

UBitNumber: 50207613

Task:

1. Compute for each variable ((CS Score, Research Overhead, Admin Base Pay, Tuition)) its sample mean, variance and standard deviation. Related variables: mu1, mu2, mu3, mu4, var1, var2, var3, var4, sigma1, sigma2, sigma3, sigma4

   Mean:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x(i)$$

CS Score	mu1	3.214
Research Overhead	mu2	53.386
Administrator Base Pay ($)	mu3	469178.816
Tuition (Out of State) ($)	mu4	29711.959

Variance:

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}[x(i) - \mu]^2$$

CS Score	var1	0.457
Research Overhead	var2	12.850
Administrator Base Pay ($)	var3	14189720820.903
Tuition (Out of State) ($)	var4	31367695.790

Standard Deviation:

CS Score	sigma1	0.669
Research Overhead	sigma2	3.548
Administrator Base Pay ($)	sigma3	117898.832
Tuition (Out of State) ($)	sigma4	5543.243

2. Compute for each pair of variables their covariance and correlation. Show the results in the form of covariance and correlation matrices. Also make a plot of the pairwise data showing the label associated with each data point. Which are the most correlated and least correlated variable pair? Related variables: covarianceMat, correlationMat

Covariance Matrix:

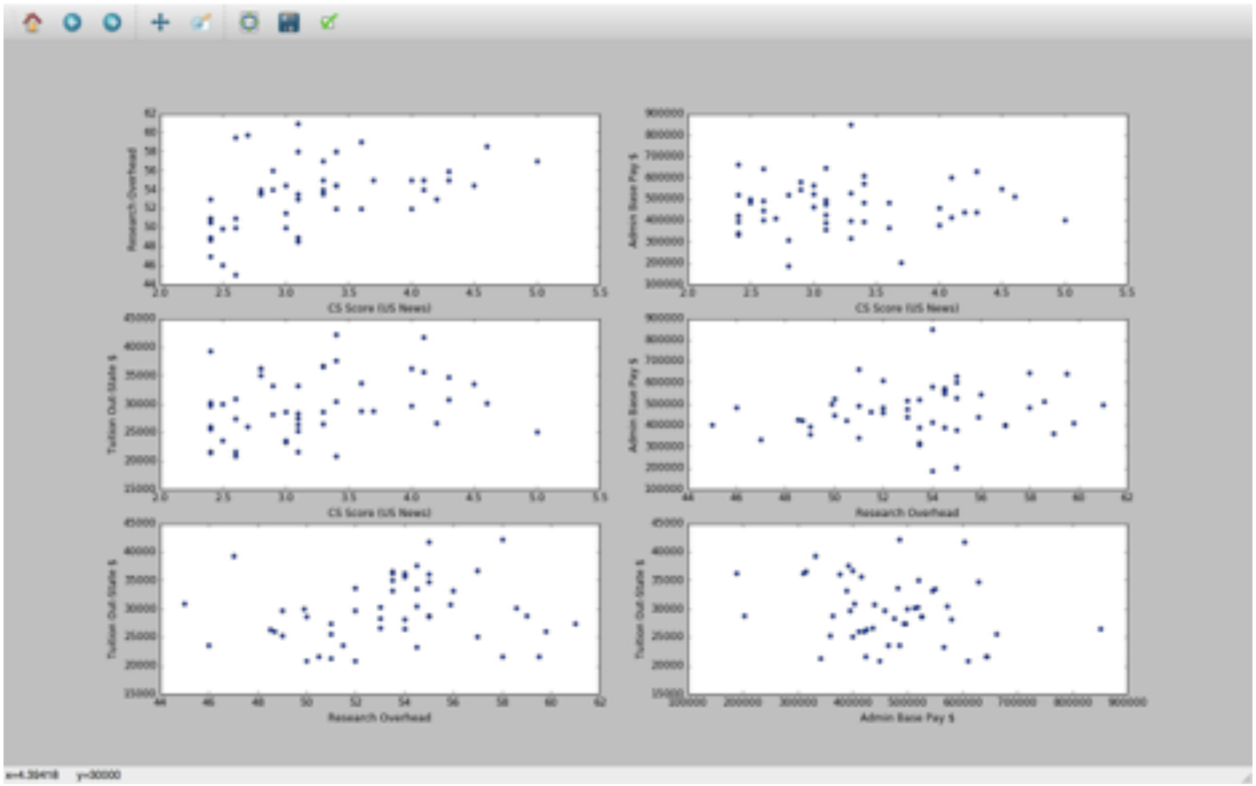
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1d} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2d} \\ \sigma_{13} & \sigma_3^2 & \sigma_3^2 & \cdots & \sigma_{3d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1d} & \sigma_{2d} & \sigma_{3d} & \cdots & \sigma_d^2 \end{bmatrix}$$

covarianceMat:
[[4.575e-01   1.106e+00   3.880e+03   1.058e+03]
[1.106e+00   1.285e+01   7.028e+04   2.806e+03]
[3.880e+03   7.028e+04   1.419e+10 -1.637e+08]
[1.058e+03   2.806e+03 -1.637e+08   3.137e+07]]

Correlation Coefficients:

correlationMat:
[[1.   0.456 0.048 0.279]
[0.456 1.   0.165 0.14]
[0.048 0.165 1.   -0.245]
[0.279 0.14 -0.245 1.   ]]

Pairwise Plot:



3. Assuming that each variable is normally distributed and that they are independent of each other, determine the log-likelihood of the data (Use the means and variances computed earlier to determine the likelihood of each data value.) Related variables: logLikelihood

   logLikelihood:

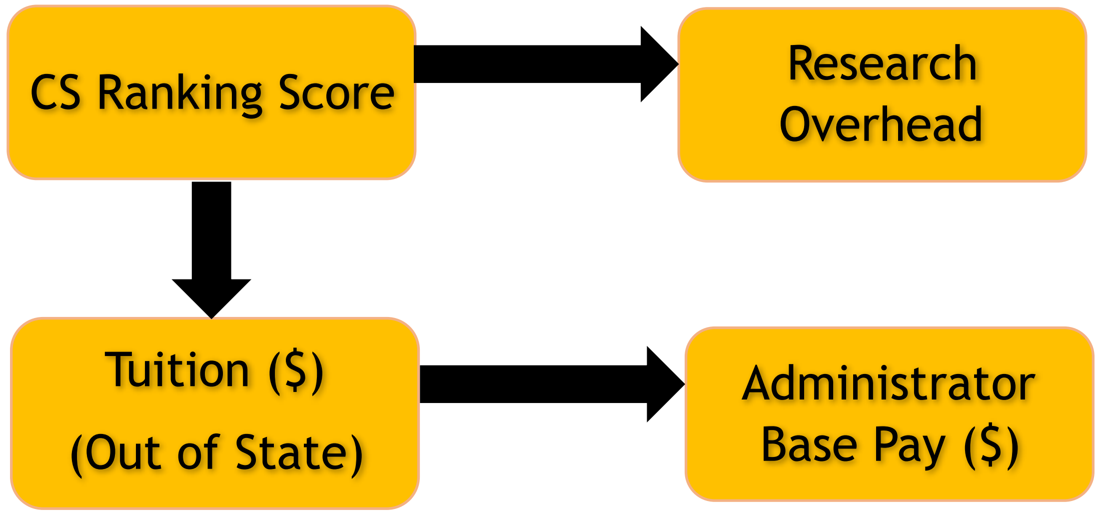
   $$\mathbf{L}(\mathbf{x}_1, ..\mathbf{x}_N) = \sum_{i=1}^{N} log\ p(\mathbf{x}_i)$$

   logLikelihood:
   -1315.099

4. Using the correlation values construct a Bayesian network which results in a higher loglikelihood than in Related variables: BNgraph, BNlogLikelihood

   The Bayesian network was created using the Brute-Force technique which led to the best possible logLikelihood for the network and was greater than the one derived by mathematical computation.

Bayesian Network:



BNGraph:

[[0 1 0 1]

[0 0 0 0]

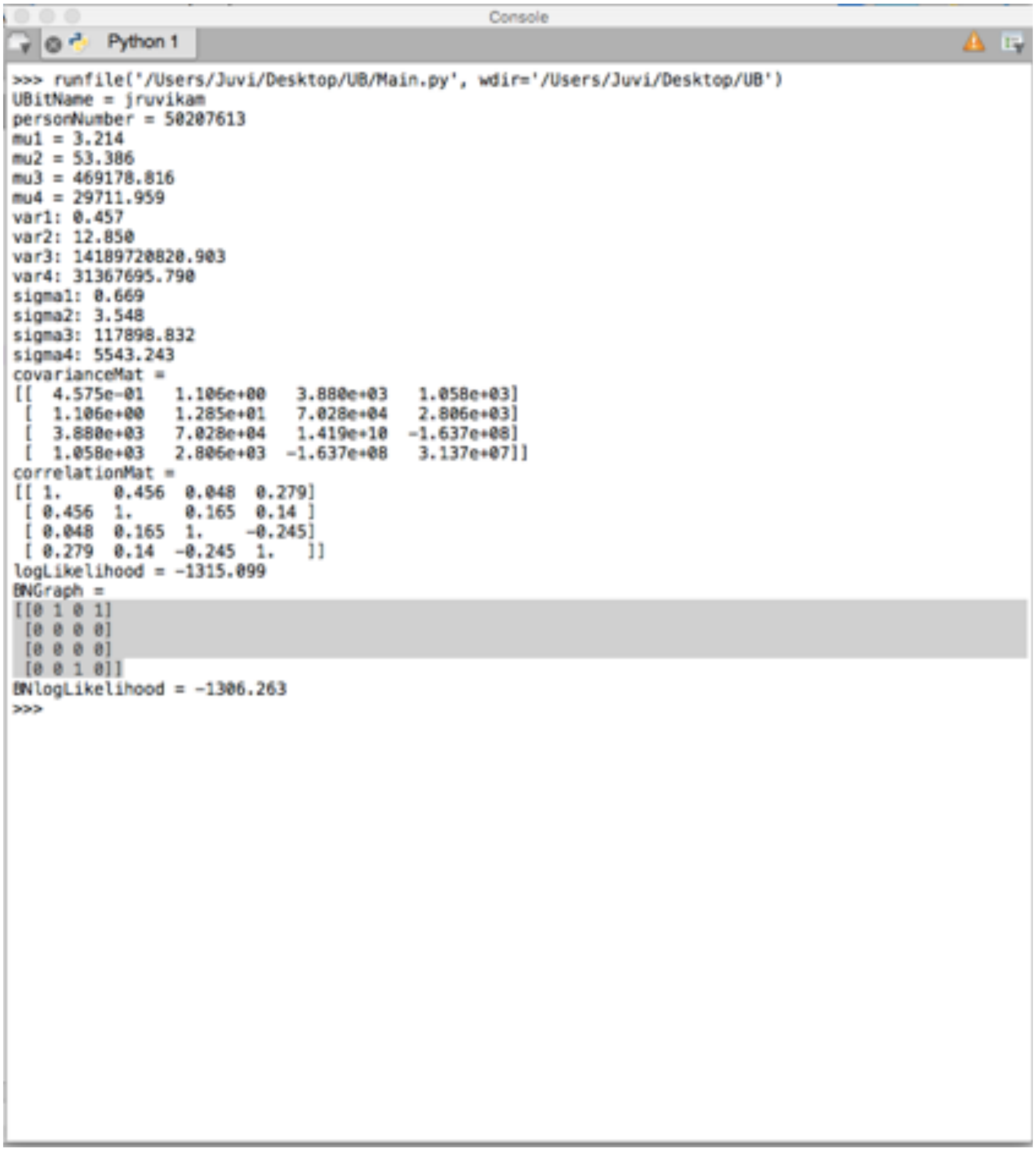
[0 0 0 0]

[0 0 1 0]]

BNlogLikelihood:

-1306.263

logLikelihood	-1315.099
BNlogLikelihood	-1306.263



Console

Python 1

```
>>> runfile('/Users/Juvi/Desktop/UB/Main.py', wdir='/Users/Juvi/Desktop/UB')
UBitName = jruvikam
personNumber = 50207613
mu1 = 3.214
mu2 = 53.386
mu3 = 469178.816
mu4 = 29711.959
var1: 0.457
var2: 12.850
var3: 14189720820.903
var4: 31367695.790
sigma1: 0.669
sigma2: 3.548
sigma3: 117898.832
sigma4: 5543.243
covarianceMat =
[[  4.575e-01   1.106e+00   3.880e+03   1.058e+03]
 [  1.106e+00   1.285e+01   7.028e+04   2.806e+03]
 [  3.880e+03   7.028e+04   1.419e+10  -1.637e+08]
 [  1.058e+03   2.806e+03  -1.637e+08   3.137e+07]]
correlationMat =
[[ 1.     0.456  0.048  0.279]
 [ 0.456  1.     0.165  0.14 ]
 [ 0.048  0.165  1.    -0.245]
 [ 0.279  0.14  -0.245  1.   ]]
logLikelihood = -1315.099
BNGraph =
[[0 1 0 1]
 [0 0 0 0]
 [0 0 0 0]
 [0 0 1 0]]
BNlogLikelihood = -1306.263
>>>
```