Problem 9.4

Chen Bo Calvin Zhang

01/01/2021

Firstly, let us load the data set and analyse it.

```
library("MASS")
data(painters)

X.painters = painters[, 1:4]
print(dim(X.painters))
```

```
## [1] 54  4
```

```
print(colnames(X.painters))
```

```
## [1] "Composition" "Drawing"     "Colour"      "Expression"
```

```
L.painters = as.factor(painters[, 5]) # school
print(levels(L.painters))
```

```
## [1] "A" "B" "C" "D" "E" "F" "G" "H"
```

```
print(table(L.painters))
```

```
## L.painters
##  A  B  C  D  E  F  G  H
## 10  6  6 10  7  4  7  4
```

Now we find the conditional independence graph relating the four variables. First, we need to compute the correlation matrix.

```
Sigma = cor(X.painters)
print(Sigma)
```

```
##             Composition    Drawing      Colour Expression
## Composition  1.00000000  0.4154456 -0.09758818  0.6571846
## Drawing      0.41544563  1.0000000 -0.51696052  0.5737066
## Colour      -0.09758818 -0.5169605  1.00000000 -0.1995179
## Expression   0.65718460  0.5737066 -0.19951793  1.0000000
```

Now we can compute the partial correlation matrix by computing the inverse, standardising and negating the off-diagonal elements.

```
Omega = solve(Sigma)
Rho = -cov2cor(Omega)
diag(Rho) = - diag(Rho)
print(Rho)
```

```
##             Composition     Drawing      Colour Expression
## Composition  1.00000000  0.09834096  0.08871752 0.55290355
## Drawing      0.09834096  1.00000000 -0.50582006 0.41080947
## Colour       0.08871752 -0.50582006  1.00000000 0.06584993
## Expression   0.55290355  0.41080947  0.06584993 1.00000000
```

Values close to zero indicate that the variables are conditionally independent. Values close to one (minus one) mean that the variables are positively (negatively) correlated.

Now we want to check how accurately we can predict the school for each artist. We will use LDA and K-fold cross-validation with 5 folds and 50 repetitions.

```
predfun.lda = function(train.x, train.y, test.x, test.y)
{
  lda.fit = lda(train.x, train.y)
  ynew = predict(lda.fit, test.x)$class
  acc = mean(ynew == test.y)
  return(acc)
}

library("crossval")

cv.out = crossval(predfun.lda, X.painters, L.painters, K=5, B=50, verbose=FALSE)
print(c(cv.out$stat, cv.out$stat.se))
```

```
## [1] 0.383121523 0.008276648
```

The accuracy in the prediction is not very high. We can try and predict only artists in school D ("Venetian school").

```
L.venetian = factor(ifelse(L.painters=="D", "Venetian", "other"))
cv.out = crossval(predfun.lda, X.painters, L.venetian, K=5, B=50, verbose=FALSE)
print(c(cv.out$stat, cv.out$stat.se))
```

```
## [1] 0.870218182 0.005672645
```

Here we have a higher accuracy.