

Recommended Exercise 1 in Statistical Linear Models, Spring 2021

alexaoh

11.01.2021

Problem 1 Simple matrix calculations

Solve the problems by hand and by use of R when possible.

a)

Construct A as a matrix in R.

```
A <- matrix(c(9, -2, -2, 6), nrow = 2)
A
```

```
#>      [,1] [,2]
#> [1,]    9  -2
#> [2,]   -2   6
```

b)

A is symmetric, which can be seen in R via

```
t(A) == A
```

```
#>      [,1] [,2]
#> [1,] TRUE TRUE
#> [2,] TRUE TRUE
```

c)

Show that A is positive definite.

From the definition of positive definiteness we get

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 9 & -2 \\ -2 & 6 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 9a^2 - 4ab + 6b^2 = 9\left(a^2 - \frac{4}{9}ab + \frac{4}{81}b^2\right) + \frac{50}{9}b^2 = 9\left(a - \frac{2}{9}b\right)^2 + \frac{50}{9}b^2 > 0.$$

Hence, A is positive definite.

d)

Find the eigenvalues and eigenvectors of A .

```
evalues <- eigen(A)
```

```
# Names in list given from eigen().
names(evalues)
```

```
#> [1] "values" "vectors"
```

```

# Eigenvalues.
evaluations$values

#> [1] 10 5

# Eigenvectors.
evaluations$eigenvectors

#>      [,1]      [,2]
#> [1,] -0.8944272 -0.4472136
#> [2,]  0.4472136 -0.8944272

# Check if the eigenvectors are normal.
t(evaluations$eigenvectors[, 1]) %*% evaluations$eigenvectors[, 1]

#>      [,1]
#> [1,] 1

t(evaluations$eigenvectors[, 2]) %*% evaluations$eigenvectors[, 2]

#>      [,1]
#> [1,] 1

# Since t(vector)*vector = 1 for both vectors, they are normal.

```

e)

Find an orthogonal diagonalization of A .

```

lambda <- diag(evaluations$values)
P <- evaluations$eigenvectors

# Orthogonal diagonalization.
ortho.diag <- P %*% lambda %*% t(P)
ortho.diag

#>      [,1] [,2]
#> [1,] 9 -2
#> [2,] -2 6

ortho.diag == A

#>      [,1] [,2]
#> [1,] TRUE TRUE
#> [2,] TRUE TRUE

```

By hand it can be done in the same way. Construct a matrix Λ which contains the eigenvalues of A on its diagonal. Furthermore, a matrix P is constructed by using the normalized eigenvectors as columns, in the same respective order as in Λ . The product $A = P\Lambda P^T$ is the orthogonal diagonalization of A .

f)

Find A^{-1} .

```

A.inverse <- solve(A)
A.inverse

#>      [,1] [,2]
#> [1,] 0.12 0.04
#> [2,] 0.04 0.18

```

By hand it can be done by use of standard Gaussian elimination.

g)

Find the eigenvalues and eigenvectors of A^{-1} . Is there a relationship between the eigenvalues and eigenvectors of A and A^{-1} ?

```
evalues.inverse <- eigen(A.inverse)
```

```
# Eigenvalues of A.  
evalues$values
```

```
#> [1] 10 5
```

```
# Eigenvalues of A inverse.  
evalues.inverse$values
```

```
#> [1] 0.2 0.1
```

```
# Eigenvectors of A.  
evalues$vectors
```

```
#>           [,1]      [,2]  
#> [1,] -0.8944272 -0.4472136  
#> [2,]  0.4472136 -0.8944272
```

```
# Eigenvectors of A inverse.  
evalues.inverse$vectors
```

```
#>           [,1]      [,2]  
#> [1,] 0.4472136 -0.8944272  
#> [2,] 0.8944272  0.4472136
```

It is apparent that the eigenvalues of A and A^{-1} are reciprocals. Furthermore, the eigenvectors have the following relationship

- The eigenvector of A 's eigenvalue 10 is identical to the eigenvector of A^{-1} 's eigenvalue $\frac{1}{10}$ (the reciprocal eigenvalue).
- The eigenvector of A 's eigenvalue 5 is -1 times the eigenvector of A^{-1} 's eigenvalue $\frac{1}{5}$ (the reciprocal eigenvalue).

In general theory this makes sense because an eigenvalue/-vector pair is defined as $A\mathbf{v} = \lambda\mathbf{v}$ for an eigenvector $\mathbf{v} \neq 0$. This expression is equivalent to $A^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v}$ (by multiplication of the inverse matrix A^{-1} and division by λ). This means that each eigenvector \mathbf{v} of A corresponding to the eigenvalue λ is an eigenvector of A^{-1} corresponding to the reciprocal eigenvalue $\frac{1}{\lambda}$. As a sidenote, the negative sign on the second eigenvector calculated via R is probably a result of the way the eigenvectors have been calculated in R. Also, the different order of the eigenvalues/-vectors is a coincidence.

h)

Why can A be a covariance matrix?

A can be a covariance matrix since $(a_{12}) = (a_{21})$, since the covariance operator has the property $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$, $i \neq j$. This means that it is symmetric. Furthermore, A is positive definite. We know that a matrix can be a covariance matrix if it is symmetric positive semi-definite, which is the case for A . As noted in suggested solutions, later in the course we will see that for all positive definite symmetric matrices there exists a multivariate normal distribution having that covariance matrix.

i)

Assume that A is the covariance matrix of a random vector. Find the correlation matrix. Check the computations with `cov2cor`.

The correlation matrix is given by the following calculation in R

```
R <- diag(1, nrow = 2)
rho <- function(i, j){
  return (A[i, j]/sqrt(A[i, i]*A[j, j]))
}

R[1, 2] <- rho(1, 2)
R[2, 1] <- rho(2, 1)

# Correlation matrix calculated manually.
R

#>           [,1]      [,2]
#> [1,]  1.0000000 -0.2721655
#> [2,] -0.2721655  1.0000000

# Correlation matrix calculated with cov2cor.
cov2cor(A)

#>           [,1]      [,2]
#> [1,]  1.0000000 -0.2721655
#> [2,] -0.2721655  1.0000000

# Perhaps a smoother calculation given below (from LF).
varvec <- diag(A) # Variances.
varvec

#> [1] 9 6

invs <- diag(1/sqrt(varvec)) # Inverse standard deviations.
invs

#>           [,1]      [,2]
#> [1,] 0.3333333 0.0000000
#> [2,] 0.0000000 0.4082483

invs %*% A %*% invs

#>           [,1]      [,2]
#> [1,]  1.0000000 -0.2721655
#> [2,] -0.2721655  1.0000000
```

It is apparent that the two matrices are identical. The same formula for $\rho_{ij} = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Cov}(X_i)\text{Cov}(X_j)}}$ can be used to find the correlation matrix R by hand, which has ones on the entire diagonal (naturally, since each value in the stochastic vector correlates to the fullest with itself).

j)

Find the expectation and covariance matrices of the given expressions.

```
E.X <- c(3,1)
Cov.X <- A

# First expression.
```

```

B <- matrix(c(1,1,1,2), nrow = 2)
B%*%E.X # Expectation.

#>      [,1]
#> [1,]    4
#> [2,]    5

B%*%Cov.X%*%t(B) # Covariance.

#>      [,1] [,2]
#> [1,]   11   15
#> [2,]   15   25

# Second expression.
C <- matrix(c(1,2), nrow=1)
C%*%E.X # Expectation.

#>      [,1]
#> [1,]    5

C%*%Cov.X%*%t(C) # Covariance.

#>      [,1]
#> [1,]   25

# Third expression (block matrix).
c(E.X, 3*E.X) # Expectation.

#> [1] 3 1 9 3

cbind(rbind(Cov.X, 3*Cov.X), rbind(3*Cov.X, 9*Cov.X)) # Covariance.

#>      [,1] [,2] [,3] [,4]
#> [1,]    9  -2  27  -6
#> [2,]   -2    6  -6  18
#> [3,]   27  -6  81 -18
#> [4,]   -6   18 -18  54

```

The last task was to find the expectation and covariance of the block matrix

$$\begin{pmatrix} \mathbf{X} \\ 3\mathbf{X} \end{pmatrix}.$$

Finding the expectation is pretty obvious, since expectation is linear. The expectation is given by

$$\begin{pmatrix} 3 \\ 1 \\ 9 \\ 1 \end{pmatrix}.$$

However the covariance matrix could perhaps be explained. This matrix should encapsulate all four combinations of correlation between the two blocks in the block matrix, which means that the covariance matrix should take the form

$$\begin{pmatrix} \text{Cov}(\mathbf{X}) & \text{Cov}(\mathbf{X}, 3\mathbf{X}) \\ \text{Cov}(3\mathbf{X}, \mathbf{X}) & \text{Cov}(3\mathbf{X}) \end{pmatrix},$$

which explains the covariance matrix found via R.

Problem 2 Mean and covariance of linear combinations

Find the mean and covariance of AX (below).

```
E.X <- c(1,1,1)
Cov.X <- diag(3)

D <- matrix(c(2/3, -1/3, -1/3, -1/3, 2/3, -1/3, -1/3, -1/3, 2/3), nrow=3)

# Mean.
D%*%E.X

#>      [,1]
#> [1,]    0
#> [2,]    0
#> [3,]    0

# Covariance.
D%*%Cov.X%*%t(D)

#>      [,1]      [,2]      [,3]
#> [1,]  0.6666667 -0.3333333 -0.3333333
#> [2,] -0.3333333  0.6666667 -0.3333333
#> [3,] -0.3333333 -0.3333333  0.6666667
```

Problem 3 Covariance formula

Show that $\text{Cov}(\mathbf{V}, \mathbf{W}) = E(\mathbf{V}\mathbf{W}^T) - (E\mathbf{V})(E\mathbf{W})^T$.

Recalling the definition of covariance, the proof is as follows

$$\begin{aligned}\text{Cov}(\mathbf{V}, \mathbf{W}) &= E\{(\mathbf{V} - E\mathbf{V})(\mathbf{W} - E\mathbf{W})^T\} \\ &= E\{\mathbf{V}\mathbf{W}^T - \mathbf{V}(E\mathbf{W})^T - (E\mathbf{V})\mathbf{W}^T + (E\mathbf{V})(E\mathbf{W})^T\} \\ &= E\{\mathbf{V}\mathbf{W}^T\} - (E\mathbf{V})(E\mathbf{W})^T - (E\mathbf{V})(E\mathbf{W})^T + (E\mathbf{V})(E\mathbf{W})^T \\ &= E(\mathbf{V}\mathbf{W}^T) - (E\mathbf{V})(E\mathbf{W})^T,\end{aligned}$$

provided that $E(\mathbf{X}^T) = E(\mathbf{X})^T$.

Problem 4 The square root matrix and the Mahalanobis transform

a)

Σ is positive semidefinite because, for any vector \mathbf{a} of equal length to \mathbf{X} , the following holds

$$0 \leq \text{Cov}(\mathbf{a}^T \mathbf{X}) = \mathbf{a}^T \text{Cov}(\mathbf{X}) \mathbf{a} = \mathbf{a}^T \Sigma \mathbf{a}.$$

This is a direct consequence of the fact that the variance always is nonnegative. Also, it is symmetric, since the covariance-operator is symmetric in its arguments. Hence, Σ is symmetric positive semidefinite.

b)

Since we are assuming that Σ is positive definite, we know the following

$$\begin{aligned} 0 < \mathbf{a}^T \Sigma \mathbf{a} &= \mathbf{a}^T P \Lambda P^T \mathbf{a} \\ &= \mathbf{y}^T \Lambda \mathbf{y} \quad \text{define } \mathbf{y} = P^T \mathbf{a} \\ &= \sum_i \lambda_i y_i^2, \end{aligned}$$

Since this must hold for all non-zero vectors \mathbf{a} , we are free to choose. Choose \mathbf{a} such that \mathbf{y} equals the unit vector δ_j ($\mathbf{a} = \delta_j P^T$). From the calculations above, this gives $0 < \lambda_j$. By choosing all necessary unit vectors it becomes apparent that each of the eigenvalues must be positive. Hence, all the eigenvalues of Σ are positive.

Σ has an inverse because it is non-singular, which is apparent since all eigenvalues are positive. The same goes for Σ^{-1} because $\Sigma^{-1} = (P \Lambda P^T)^{-1} = P \Lambda^{-1} P^T$ (since P is orthogonal) is non-singular as well, since this shows that the eigenvalues of Σ^{-1} are the reciprocals of the eigenvalues of Σ .

c)

Define $\Sigma^{1/2} = P \Lambda^{1/2} P^T$ and $\Sigma^{-1/2} = P \Lambda^{-1/2} P^T$. Both are symmetric, because $(\Sigma^{1/2})^T = \Sigma^{1/2}$ and $(\Sigma^{-1/2})^T = \Sigma^{-1/2}$. Furthermore

$$\begin{aligned} \Sigma^{1/2} \Sigma^{1/2} &= P \Lambda^{1/2} P^T P \Lambda^{1/2} P^T = P \Lambda^{1/2} \Lambda^{1/2} P^T = P \Lambda P^T = \Sigma, \\ \Sigma^{-1/2} \Sigma^{-1/2} &= P \Lambda^{-1/2} P^T P \Lambda^{-1/2} P^T = P \Lambda^{-1/2} \Lambda^{-1/2} P^T = P \Lambda^{-1} P^T = \Sigma^{-1}, \\ \Sigma^{1/2} \Sigma^{-1/2} &= P \Lambda^{1/2} P^T P \Lambda^{-1/2} P^T = P \Lambda^{1/2} \Lambda^{-1/2} P^T = P P^T = I. \end{aligned}$$

d)

The Mahalanobis transform is given as $\mathbf{Y} = \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$. The expectation of this transform is $E\mathbf{Y} = E\{\Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\} = \Sigma^{-1/2}E\{\mathbf{X} - \boldsymbol{\mu}\} = \Sigma^{-1/2}(E\mathbf{X} - \boldsymbol{\mu}) = \Sigma^{-1/2}(\boldsymbol{\mu} - \boldsymbol{\mu}) = \mathbf{0}$ and the covariance is $\text{Cov}(\mathbf{Y}) = \text{Cov}(\Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})) = \Sigma^{-1/2} \text{Cov}(\mathbf{X})(\Sigma^{-1/2})^T = \Sigma^{-1/2} \Sigma \Sigma^{-1/2} = \Sigma^{-1/2} \Sigma^{1/2} \Sigma^{1/2} \Sigma^{-1/2} = I$.