MAST90104: An Introduction to Statistical Learning

Week 1 Lab Solutions

1. Show that $X^T X$ is a symmetric matrix.

   Solution:
$$(X^T X)^T = X^T (X^T)^T = X^T X.$$

2. (a) Let
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
   be a nonsingular 2×2 matrix. Show by direct multiplication that
$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

   Solution:
$$\frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{bmatrix} = I.$$

   (b) Find the inverse of
$$\begin{bmatrix} 2 & 4 \\ 1 & -3 \end{bmatrix}.$$

   Solution: From above, the inverse is
$$-\frac{1}{10} \begin{bmatrix} -3 & -4 \\ -1 & 2 \end{bmatrix}.$$

3. Is
$$X = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$
   orthogonal? If not, what value of c makes the matrix cX orthogonal?

   Solution: The matrix is not orthogonal, as its columns do not form an orthonormal set (e.g. the first column has norm > 1). However they do form an *orthogonal* set, so we can just normalise each vector to produce an orthogonal matrix. This gives $c = \frac{1}{2}$.

4. (a) Find the eigenvalues, and an associated eigenvector for each eigenvalue, of the matrix
$$A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}.$$

   Solution: The characteristic equation is
$$det(A - \lambda I) = (2 - \lambda)^2 - 4 = -4\lambda + \lambda^2 = \lambda(\lambda - 4) = 0,$$
   so the eigenvalues are 0 and 4.
   For $\lambda = 4$, we solve
$$\begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$
   One such solution is
$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

For $\lambda = 0$, we solve

$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

One such solution is

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

(b) Find an orthogonal matrix P such that $P^T AP$ is diagonal.

Solution:

$$P = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

(c) Write down $P^T AP$ for the P given in part (b).

Solution:

$$P^T AP = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}.$$

5. Let

$$A = \begin{bmatrix} 1 & 4 & 3 \\ -2 & 0 & 2 \\ 4 & 4 & 0 \end{bmatrix}.$$

(a) Write down the trace of A.

Solution: $tr(A) = 1$.

(b) Are the columns of A linearly independent? Justify your answer.

Solution: The columns of A are not linearly independent: the second column is the sum of the first and third.

(c) Find the rank of A.

Solution: The first and third columns of A are not multiples of each other, so $r(A) = 2$.

6. Show that if X is of full rank, then

$$I - X(X^T X)^{-1} X^T$$

is an idempotent matrix.

Solution: In general, if A is idempotent then so is $I - A$, since

$$(I - A)(I - A) = I - A - A + A^2 = I - A - A + A = I - A.$$

To see that $X^T X$ is invertible, we know that $r(X^T X) = r(X) = k$ since X is of full rank. To see that $A = X(X^T X)^{-1} X^T$ is idempotent we just multiply it by itself:

$$[X(X^T X)^{-1} X^T][X(X^T X)^{-1} X^T] = X(X^T X)^{-1}(X^T X)(X^T X)^{-1} X^T = X(X^T X)^{-1} X^T.$$

7. Prove that a (real) symmetric matrix A is positive semidefinite if and only if all of its eigenvalues are non-negative, and positive definite if and only if all of its eigenvalues are strictly positive.

Solution: First note that since A is symmetric we can find an orthogonal P such that $P^T AP = \Lambda$, where Λ is a diagonal matrix formed from the eigenvalues.

($\Leftarrow$) Let $\lambda_1, \ldots, \lambda_n \geq 0$ be the eigenvalues of A. For any $\mathbf{x}$ we have, for $\mathbf{z} = P^T \mathbf{x} = (z_1, \ldots, z_n)^T$,

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T P\Lambda P^T \mathbf{x} = \mathbf{z}^T \Lambda \mathbf{z} = \sum_{i=1}^{n} z_i^2 \lambda_i \geq 0.$$

Thus A is positive semidefinite as required.

($\Rightarrow$) Suppose that A is positive semidefinite. Let $\mathbf{x}_i$ be its normalised i-th eigenvector, then

$$0 \leq \mathbf{x}_i^T A\mathbf{x}_i = \lambda_i \mathbf{x}_i^T \mathbf{x}_i = \lambda_i.$$

So the eigenvalues of A are non-negative as required.

8. Prove that for any matrix A
$$r(A) = r(A^T) = r(A^T A).$$
You may use the fact that pre- or post-multiplying by a non-singular matrix does not change the rank.

Solution: Suppose that A is of dimension $m \times n$ and $r(A) = k$. Then the column space of A has dimension k; let $\mathbf{c}_1, \dots, \mathbf{c}_k$ be a basis for this column space. Let $C = [\mathbf{c}_1|\cdots|\mathbf{c}_k]$. Now every column of A can be expressed as a linear combination of columns of C; this means that there is a $k \times n$ matrix R such that $A = CR$.

But now each row of A is expressed as a linear combination of the k rows of R; hence the row space of A has dimension at most k. In other words, $r(A^T) \leq r(A)$. As we can apply this argument equally to A^T, we have $r(A^T) = r(A)$.

For the second part of the problem, we first note that
$$A\mathbf{x} = \mathbf{0} \text{ if and only if } A^T A\mathbf{x} = \mathbf{0}.$$
The only if part is obvious. For the if part we have
$$A^T A\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x}^T A^T A\mathbf{x} = \mathbf{0} \Rightarrow \|A\mathbf{x}\|^2 = \mathbf{0} \Rightarrow A\mathbf{x} = \mathbf{0}.$$
Let $B = [\mathbf{b}_1|\cdots|\mathbf{b}_n]$, then $B\mathbf{x} = \mathbf{0}$ is equivalent to $x_1\mathbf{b}_1 + \cdots x_n\mathbf{b}_n = \mathbf{0}$. So since $A\mathbf{x} = \mathbf{0}$ if and only if $A^T A\mathbf{x} = \mathbf{0}$, we see that a linear combination of the columns of A is zero precisely when the same linear combination of columns of $A^T A$ is zero. Thus they have the same number of linearly independent columns and so by definition $r(A^T A) = r(A)$.

R exercises

The following are taken from Chapter 2 of spuRs (Introduction to Scientific Programming and Simulation Using R).

1. Give R assignment statements that set the variable z to
   (a) x^{a^b}
   (b) $(x^a)^b$
   (c) $3x^3 + 2x^2 + 6x + 1$ (try to minimise the number of operations required)
   (d) the second-to-last digit of x before the decimal point (hint: use `floor(x)` and/or `%%`)
   (e) $z + 1$

   Solution:

```
> x <- 123
> a <- 1.1
> b <- 1.2
> # a
> (z <- x^(a^b))

[1] 220.3624

> (z <- x^a^b)

[1] 220.3624

> # b
> (z <- (x^a)^b)

[1] 573.6867

> # c
> (z <- 3*x^3 + 2*x^2 + 6*x + 1) #8 operations
```

```
[1] 5613598

> (z <- (3*x + 2)*(x^2 + 2) - 3) #6 operations

[1] 5613598

> (z <- sum((x^(3:0))*c(3, 2, 6, 1))) #vectorised

[1] 5613598

> # d
> y <- abs(x)
> (z <- (y %% 100 - y %% 10)/10)

[1] 2

> (z <- floor(y/10 - floor(y/100)*10))

[1] 2

> # e
> (z <- z + 1)

[1] 3
```

2. Give R expressions that return the following matrices and vectors

   (a) $(1, 2, 3, 4, 5, 6, 7, 8, 7, 6, 5, 4, 3, 2, 1)$

   (b) $(1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5)$

   (c) $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$

   (d) $\begin{pmatrix} 0 & 2 & 3 \\ 0 & 5 & 0 \\ 7 & 0 & 0 \end{pmatrix}$

   Solution:

```
> # a
> c(1:8, 7:1)

 [1] 1 2 3 4 5 6 7 8 7 6 5 4 3 2 1

> # b
> rep(1:5, 1:5)

 [1] 1 2 2 3 3 3 4 4 4 4 5 5 5 5 5

> # c
> matrix(1, 3, 3) - diag(3)

     [,1] [,2] [,3]
[1,]    0    1    1
[2,]    1    0    1
[3,]    1    1    0

> # d
> matrix(c(0,0,7, 2,5,0, 3,0,0), 3, 3)
```

```
     [,1] [,2] [,3]
[1,]    0    2    3
[2,]    0    5    0
[3,]    7    0    0
```

3. Suppose `vec` is a strictly positive vector of length 2. Interpreting `vec` as the coordinates of a point in $\mathbb{R}^2$, use R to express it in polar coordinates. You will need (at least one of) the inverse trigonometric functions: `acos(x)`, `asin(x)`, and `atan(x)`.

   Solution:

```
> # assuming x > 0
> vec <- c(sqrt(3), 1)
> x <- vec[1]
> y <- vec[2]
> (R <- sqrt(x^2 + y^2))          # radial distance

[1] 2

> (theta.rad <- atan(y/x))        # angle in radians

[1] 0.5235988

> (theta.deg <- theta.rad*180/pi)  # angle in degrees

[1] 30

> # in general
> vec <- c(sqrt(3), 1)
> x <- vec[1]
> y <- vec[2]
> (R <- sqrt(x^2 + y^2))

[1] 2

> if (x > 0) {
+   (theta.rad <- atan(y/x))
+ } else {
+   (theta.rad <- atan(y/x) + pi)
+ }

[1] 0.5235988

> (theta.deg <- theta.rad*180/pi)

[1] 30
```

4. Use R to produce a vector containing all integers from 1 to 100 that are not divisible by 2, 3, or 7.

   Solution:

```
> x <- 1:100
> idx <- (x %% 2 != 0) & (x %% 3 != 0) & (x %% 7 != 0)
> x[idx]

 [1]  1  5 11 13 17 19 23 25 29 31 37 41 43 47 53 55 59 61 65 67 71 73 79 83 85
[26] 89 95 97
```

5. Suppose that `queue <- c("Steve", "Russell", "Alison", "Liam")` and that `queue` represents a supermarket queue with Steve first in line. Using R expressions update the supermarket queue as successively:

   (a) Barry arrives;

(b) Steve is served;
(c) Pam talks her way to the front with one item;
(d) Barry gets impatient and leaves;
(e) Alison gets impatient and leaves.

For the last case you should not assume that you know where in the queue Alison is standing.

Finally, using the function `which(x)`, find the position of Russell in the queue.

Note that when assigning a text string to a variable, it needs to be in quotes.

Solution:

```
> (queue <- c("S", "R", "A", "L"))

[1] "S" "R" "A" "L"

> # a
> (queue <- c(queue, "B"))

[1] "S" "R" "A" "L" "B"

> # b
> (queue <- queue[-1])

[1] "R" "A" "L" "B"

> # c
> (queue <- c("P", queue))

[1] "P" "R" "A" "L" "B"

> # d
> (queue <- queue[1:(length(queue)-1)])

[1] "P" "R" "A" "L"

> # e
> (queue <- queue[queue != "A"])

[1] "P" "R" "L"

> which(queue == "R")

[1] 2
```

6. Which of the following assignments will be successful? What will the vectors `x`, `y`, and `z` look like at each stage?

```
rm(list = ls())
x <- 1
x[3] <- 3
y <- c()
y[2] <- 2
y[3] <- y[1]
y[2] <- y[4]
z[1] <- 0
```

7. Build a 10×10 identity matrix. Then make all the non-zero elements 5. Do this latter step in at least two different ways.

Solution:

```
> Id <- diag(10)
> Id <- 5*Id # one way, using the fact that Id is the identity
> Id[Id != 0] <- 5 # another way, using vector indexing of a matrix
```