MAST90104: A first course in Statistical Learning

Week 3 Tutorial/Lab Solutions

Practical questions

1. Consider the function $y = f(x)$ defined by

x	≤ 0	$\in (0, 1]$	> 1
$f(x)$	$-x^3$	x^2	$\sqrt{x}$

Supposing that you are given x, write an R expression for y using `if` statements.
Add your expression for y to the following program, then run it to plot the function f.

```
# input
x.values <- seq(-2, 2, by = 0.1)

# for each x calculate y
n <- length(x.values)
y.values <- rep(0, n)
for (i in 1:n) {
   x <- x.values[i]
   # your expression for y goes here
   y.values[i] <- y
}

# output
plot(x.values, y.values, type = "l")
```

Solution: replace the comment `# your expression for y goes here` with the following lines

```
if (x <= 0) {
   y <- -x^3
} else if (x <= 1) {
   y <- x^2
} else {
   y <- sqrt(x)
}
```

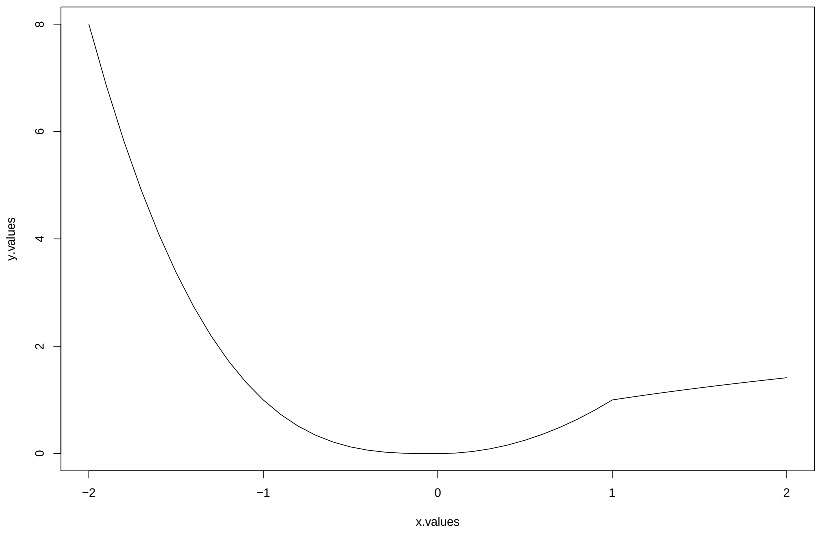

Figure 1: Lab Question 1

The function clearly has no derivative at 1, but it does have a derivative (equal to 0) at 0. It does not have a second derivative at 0 however.

2. Let $h(x,n) = 1 + x + x^2 + \cdots + x^n = \sum_{i=0}^{n} x^i$. Write an R program to calculate $h(x,n)$ using a `for` loop.

```
x <- 0.8
n <- 10

#using a for loop
h <- 0
for(i in 0:n) h <- h + x^i
print(h)

## [1] 4.570503
```

You should use the computer to calculate the formula rather than doing it yourself.

3. The function $h(x,n)$ from Exercise 2 is the finite sum of a geometric sequence. It has the following explicit formula, for $x \neq 1$,

$$h(x,n) = \frac{1 - x^{n+1}}{1 - x}.$$

Test your program from Exercise 2 against this formula using the following values

x	n	$h(x,n)$
0.3	55	1.428571
6.6	8	4243335.538178

You should use the computer to calculate the formula rather than doing it yourself.

```
x <- 0.8
# using the formula
if (x == 1) {
  h <- n + 1 } else {
  h <- (1 - x^(n+1))/(1 - x) }
print(h)
## [1] 4.570503
```

```
x <- 0.3
n <- 55
# using the formula
if (x == 1) {
   h <- n + 1
} else {
   h <- (1 - x^(n+1))/(1 - x) }
print(h)
## [1] 1.428571
x <- 6.6
n <- 8
# using the formula
if (x == 1) {
   h <- n + 1
} else {
   h <- (1 - x^(n+1))/(1 - x) }
print(h,digits=14)
## [1] 4243335.5381786
```

4. First write a program that achieves the same result as in Exercise 2 but using a `while` loop. Then write a program that does this using vector operations (and no loops).

   If it doesn't already, make sure your program works for the case $x = 1$.

   Solution:

```
# using a while loop
x <- 0.8
n<-10
h <- 0
i <- 0
while (i <= n) {
   h <- h + x^i
   i <- i + 1 }
print(h)
## [1] 4.570503
# vectorised
(h <- sum(x^(0:n)))
## [1] 4.570503
```

5. We return to the house price example in class.

Price ($\times$ \$10k)	Age (years)	Area ($\times 100m^2$)
50	1	1
40	5	1
52	5	2
47	10	2
65	20	3

   Recall that we want to predict a house price based on its age and area. The linear model is of the form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

   Create a *data frame* in R with this data. Fit the linear model using function `lm()` in R. **Solution:**

```
   > y <- c(50,40,52,47,65)
   > x1 <- c(1,5,5,10,20)
   > x2 <- c(1,1,2,2,3)
   > houseprice <- data.frame(price =y, age = x1, area = x2)
   > houseprice
   price age area
   1    50   1    1
   2    40   5    1
   3    52   5    2
```

```
4    47  10    2
5    65  20    3
> m1 <- lm(price~ age + area, data = houseprice)
> summary(m1)

Call:
lm(formula = price ~ age + area, data = houseprice)

Residuals:
1      2      3      4      5
6.409 -2.832 -1.551 -5.602 3.576

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.0626   10.5066   3.147   0.0879 .
age          -0.1897    1.1106  -0.171   0.8801
area         10.7182    9.7272   1.102   0.3854
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.916 on 2 degrees of freedom
Multiple R-squared: 0.7142, Adjusted R-squared: 0.4285
F-statistic: 2.499 on 2 and 2 DF, p-value: 0.2858

# another way
> m2 <- lm(y~ x1 + x2)
> m2

Call:
lm(formula = y ~ x1 + x2)

Coefficients:
(Intercept)         x1          x2
33.0626      -0.1897     10.7182
```

Workshop questions

1. Let X be a 10×5 matrix of full rank and let $H = X(X^TX)^{-1}X^T$.

   Find $tr(H)$ and $r(H)$.

   Solution: $tr(H) = tr(X(X^TX)^{-1}X^T) = tr(X^TX(X^TX)^{-1}) = tr(I_5) = 5$. Since H is symmetric and idempotent, $r(H) = tr(H) = 5$.

2. Let

$$A = \begin{bmatrix} 3 & 1 & 8 \\ 1 & 0 & 1 \\ 2 & 1 & -4 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

   Let $z = \mathbf{y}^TA\mathbf{y}$. Write out z in full, then find $\frac{\partial z}{\partial \mathbf{y}}$ directly *and* using the matrix formula.

   Solution: $z = 3y_1^2 - 4y_3^2 + 2y_1y_2 + 10y_1y_3 + 2y_2y_3$.

$$\frac{\partial z}{\partial \mathbf{y}} = \mathbf{y}^T(\mathbf{A} + \mathbf{A}^T) = \begin{bmatrix} 6y_1 + 2y_2 + 10y_3 \\ 2y_1 + 2y_3 \\ 10y_1 + 2y_2 - 8y_3 \end{bmatrix}^T.$$

3. Let $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^T$ be a random vector with mean $\boldsymbol{\mu} = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix}^T$, and assume that $\operatorname{Var} y_i = 4$ and $\operatorname{Cov}(y_i, y_j) = 0$.

(a) Write down $\operatorname{Var}\mathbf{y}$.

Solution:

$$\operatorname{Var}\mathbf{y} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

(b) Let

$$A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \\ -1 & 6 & 1 \end{bmatrix}$$

and find $\operatorname{Var} A\mathbf{y}$ and $\mathbb{E}[\mathbf{y}^T A\mathbf{y}]$.

Solution:

$$\begin{aligned}
\operatorname{Var} A\mathbf{y} &= AVA^T = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \\ -1 & 6 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 \\ -3 & 2 & 6 \\ 1 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \\ -1 & 6 & 1 \end{bmatrix} \begin{bmatrix} 8 & 4 & -4 \\ -12 & 8 & 24 \\ 4 & 0 & 4 \end{bmatrix} \\
&= \begin{bmatrix} 56 & -16 & -76 \\ -16 & 20 & 44 \\ -76 & 44 & 152 \end{bmatrix}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\mathbf{y}^T A\mathbf{y}] &= tr(AV) + \boldsymbol{\mu}^T A\boldsymbol{\mu} \\
&= tr\left(\begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \\ -1 & 6 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \right) + \begin{bmatrix} 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \\ -1 & 6 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \\
&= tr\left(\begin{bmatrix} 8 & -12 & 4 \\ 4 & 8 & 0 \\ -4 & 24 & 4 \end{bmatrix} \right) + \begin{bmatrix} 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} -5 \\ 7 \\ 19 \end{bmatrix} \\
&= 20 + 54 = 74.
\end{aligned}$$

4. Let $\mathbf{y} = \begin{bmatrix} y_1 & y_2 \end{bmatrix}^T$ be a normal random vector with mean and variance

$$\boldsymbol{\mu} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Let

$$A = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

(a) Find the distributions of $\mathbf{y}^T A\mathbf{y}$ and $\mathbf{y}^T B\mathbf{y}$.

Solution: A and B are both idempotent and have rank 1, so $\mathbf{y}^T A\mathbf{y}$ and $\mathbf{y}^T B\mathbf{y}$ have noncentral χ^2 distributions with 1 degree of freedom each and noncentrality parameters

$$\frac{1}{2}\begin{bmatrix} 2 & 4 \end{bmatrix} \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 9$$

and

$$\frac{1}{2}\begin{bmatrix} 2 & 4 \end{bmatrix} \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 1$$

respectively.

(b) Are $\mathbf{y}^T A\mathbf{y}$ and $\mathbf{y}^T B\mathbf{y}$ independent?

Solution: $AB = 0$, so they are independent.

(c) What is the distribution of $\mathbf{y}^T A\mathbf{y} + \mathbf{y}^T B\mathbf{y}$?

Solution: Since $\mathbf{y}^T A\mathbf{y}$ and $\mathbf{y}^T B\mathbf{y}$ are independent, by Theorem 3.5 in Lecture 3, $\mathbf{y}^T A\mathbf{y} + \mathbf{y}^T B\mathbf{y}$ has a noncentral χ^2 distribution with 2 degrees of freedom and noncentrality parameter 10.

5. Let $\mathbf{Y} \sim \text{MVN}(\mathbf{0}, \mathbf{I})$ be a $n \times 1$ random vector and let $\mathbf{A}$ be a n be a symmetric square matrix of order n (n by n symmetric matrix). Show that $\mathbf{Y}^T\mathbf{AY}$ has a (central) χ^2 distribution with k degrees of freedom if and only if $\mathbf{A}$ is idempotent and has rank k.

   Solution: Just observe that if $\mu = \mathbf{0}$ then $\lambda = \frac{1}{2}\mu^T A\mu = 0$. Then, the required result follows by Theorem 3.6 of Lecture 3.

6. Let $\mathbf{Y} \sim \text{MVN}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ be a $n \times 1$ random vector and let $\mathbf{A}$ be a n be a symmetric square matrix of order n. Show that $\frac{1}{\sigma^2}\mathbf{Y}^T\mathbf{AY}$ has a noncentral χ^2 distribution with k degrees of freedom and non-centrality parameter $\lambda = \frac{1}{2\sigma^2}\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$ if and only if $\mathbf{A}$ is idempotent and has rank k.

   Solution: From the distribution of $\mathbf{Y}$, we know that $\frac{1}{\sigma}\mathbf{Y} \sim MVN(\frac{1}{\sigma}\boldsymbol{\mu}, I)$. Therefore $(\frac{1}{\sigma}\mathbf{Y})^T A(\frac{1}{\sigma}\mathbf{Y}) = \frac{1}{\sigma^2}\mathbf{Y}^T\mathbf{AY}$ has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter

   $$\lambda = \frac{1}{2}(\frac{1}{\sigma}\boldsymbol{\mu})^T\mathbf{A}(\frac{1}{\sigma}\boldsymbol{\mu}) = \frac{1}{2\sigma^2}\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$$

   if and only if $\mathbf{A}$ is idempotent and has rank k (by Theorem 3.6 of Lecture 3).