MA5832-Assessment 1

Weighting: 20% *Total marks: 60*

Due date: Week 2-Sunday, 17^{th} May 2020, 11:59pm AEST

Overview

This assessment aims to assess students' understanding on the topics covered in week 2. It addresses the following learning outcome(s):

- understanding the roles of linear algebra, probability theory and optimisation in the realm of machines learning;
- understanding algorithms underpinning various optimisation methods;
- applying and implementing concepts in linear algebra, probability theory and optimisation using R.

Submission

You will need to submit the following:

- A PDF file clearly shows the assignment question, the associated answers, any relevant R outputs, analyses and discussions.
- **Rmarkdown** script file to reproduce your work.
- Format of PDF file: single line space, 12 pt Calibri (body) style.
- The task cover sheet.

You have up to three attempts to submit your assessment, and only the last submission will be graded.

A word on plagiarism:

Plagiarism is the act of using another's words, works or ideas from any source as one's own. Plagiarism has no place in a University. Student work containing plagiarised material will be subject to formal university processes.

1 Implementation in R

Question 1

- Let us assume that at any given time, the number of people waiting in a queue follows a Poisson distribution with mean of 3. What is the probability that 8 or more than 8 people are waiting in the queue? Show your working. **(5 marks)**

- At the James Cook University, approximately 15 % of students are international. Consider a group of 20 students in MA5832. What is the probability of having 5 or less than 5 international students in the MA5832? State underlying assumptions if they are needed and show your working. **(5 marks)**

Question 2 **(10 marks)**

In this question, we consider the `marketing` dataset from `datarium` package in R.

```
install.packages("datarium")
library(datarium)
data(marketing)
```

The data contains 200 observations and 4 variables. The response variable is sales, denoted as Y. The explanatory variables—measured in thousands of dollars—are advertising budget spent on youtube, newspapers and facebook, respectively, which are denoted as X_1, X_2 and X_3, respectively.

To model the impact of the three media on sales, a researcher uses the following multiple linear regression:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon, \tag{1}$$

$$= \mathbf{X}\boldsymbol{\beta} + \epsilon \tag{2}$$

where Y, X_1, X_2 and X_3 is a vector of $n \times 1$ (n is the number of observations in the dataset), $\mathbf{X} = (\mathbf{1}_n, X_1, X_2, X_3)$ and $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3]'$.

The parameter $\boldsymbol{\beta}$ in Equation (2) and its standard deviation ($s.d()$) can be estimated by using the function `lm()` in R. Alternatively, $\boldsymbol{\beta}$ can be estimated by minimising the following loss function - mean squared errors:

$$\mathcal{L} = \frac{1}{n}(Y - \mathbf{X}\boldsymbol{\beta})'(Y - \mathbf{X}\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2 \tag{3}$$

It is well-known that the optimal solution of $\boldsymbol{\beta}$ in Equation (3), denoted as $\hat{\boldsymbol{\beta}}$, has the

following form:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y \tag{4}$$

and its standard deviation is

$$s.d(\hat{\boldsymbol{\beta}}) = \sqrt{s^2(\mathbf{X}'\mathbf{X})^{-1}} \text{ where } s^2 = \frac{1}{n-4}\sum_{i=1}^{n}(Y_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})^2, i = 1, 2, \ldots, n. \tag{5}$$

Your tasks are to:

(a) write a program in `R` to estimate $\boldsymbol{\beta}$ and its standard deviation using equations (4) and (5).

(b) compare the results obtained in Question 2(a) with those obtained from the function `lm`().

2 Optimisation

Question 3

Another approach to estimate $\boldsymbol{\beta}$ in Equation (3) is to use Classical Gradient Descent.

(a) Write down a step-by-step procedure of Classical Gradient Descent to estimate $\boldsymbol{\beta}$ in Equation (3) **(10 marks)**

(b) Write an R code to implement the Classical Gradient Descent procedure provided in Question 3(a). **(10 marks)**

(c) Compare the estimate of $\boldsymbol{\beta}$ obtained in Question 3(b) with that obtained in Question 2(a). Provide some explanations if the results are similar or different. **(5 marks)**

Question 4 **(15 marks)**

Compare the optimisation algorithms of Classical Gradient Descent, Stochastic Gradient Descent and Newton's methods to see advantages and disadvantages of each algorithm.