STAT 3006 Assignment 1

Due date: 5:00 pm on 9 February

(30%)**Q**1: Please use the bisection method to find all zero points of the following function,

$$f(x) = x^3 + 7.8x^2 - 28.33x - 39.27.$$

(30%)**Q**2 (Poisson regression): We collected $n = 15$ independent count observations $\{y_i : i = 1, \ldots, 15\}$ and their corresponding covariates $\{x_i : i = 1, \ldots, 15\}$. Assume the relationship between y_i and x_i (for $i = 1, \ldots, 15$) is $y_i \sim Poisson(\lambda_i)$ and $log(\lambda_i) = \alpha + \beta x_i$. Please 1) write down the likelihood function $L(\alpha, \beta|\mathbf{x}, \mathbf{y})$ of the Poisson regression model; 2) derive the Newton method for maxmizing $L(\alpha, \beta|\mathbf{x}, \mathbf{y})$; 3) implement the Newton method using R to get MLE of (α, β).

x_i	-0.30	0.32	0.41	0.62	-0.21	0.31	0.41	0.81
y_i	2	10	11	22	0	6	9	34
x_i	0.50	-0.21	-0.20	0.70	0.10	0.13	0.69	-
y_i	5	0	1	21	3	2	29	-

(10%)**Q**3 (Logistic regression): We collected $n = 15$ independent binary observations $\{y_i : i = 1, \ldots, 15\}$ and their corresponding covariates $\{x_i : i = 1, \ldots, 15\}$. Assume the relationship between y_i and x_i (for $i = 1, \ldots, 15$) is $y_i \sim Bernoulli(p_i)$ and $logit(p_i) = \alpha + \beta x_i$, where $logit(t) = \log \frac{t}{1-t}$. Please 1) write down the likelihood function $L(\alpha, \beta|\mathbf{x}, \mathbf{y})$ of the logistic regression model; 2) derive the Newton method for maxmizing $L(\alpha, \beta|\mathbf{x}, \mathbf{y})$; 3) implement the Newton method using R to get MLE of (α, β).

x_i	-0.30	0.32	0.41	0.62	-0.21	0.31	0.41	0.81
y_i	0	1	1	1	0	0	0	1
x_i	0.50	-0.21	-0.20	0.70	0.10	0.13	0.69	-
y_i	0	0	0	1	1	1	0	-

(30%)**Q**4 (EM algorithm): The heights of $n = 8000$ students are drawn from a school. Assume the height largely depends on the gender. We denote the height of student i by Y_i, and the gender of student i by Z_i. $\{Y_i : 1 \leq i \leq n\}$ are observed, but $\{Z_i : 1 \leq i \leq n\}$ are unknown. Our model can be formulated as follows. First, $Pr(Z_i = 1) = \pi$ and $Pr(Z_i = 2) = 1 - \pi$, where $Z_i = 1$ indicates student i is female, $Z_i = 2$ indicates student i is male, and π can be

interpreted as the proportion of female students in all the students. Second, given $Z_i = 1$, Y_i is assumed to be from a normal distribution $N(\mu_1, \sigma_1^2)$; given $Z_i = 2$, Y_i is assumed to be from a normal distribution $N(\mu_2, \sigma_2^2)$. Based on these notations and information, please 1) write down the complete-data likelihood function $L(\pi, \mu_1, \mu_2, \sigma_1, \sigma_2 | \mathbf{Y}, \mathbf{Z})$; 2) derive E step and M step to find MLE of $(\pi, \mu_1, \mu_2, \sigma_1, \sigma_2)$; 3) use R to implement your EM algorithm, give MLE of $(\pi, \mu_1, \mu_2, \sigma_1, \sigma_2)$, and distinguish the first ten students' gender. (The data set $\{Y_i : 1 \leq i \leq n\}$ is "*heights_data.txt*".)

Requirements: your answer must contain two parts. The first part is a paper report which includes your derivation and answers for each problem. The second part is a file which includes all your R code to implement your algorithms. Please, by the due date, 1) put your paper report in the assignment box "STAT3006" besides LSB 125 and 2) submit your R code file in the elearning system. You must finish both of the two parts to get a grade. Otherwise, your homework will be regarded as missing. Details of requirements are in the table below.

-	in the paper report	in the R code file
Q1	all zero points	R code for implementing the bisection method
Q2	likelihood function derivation procedure for Newton algorithm MLE of (α, β)	R code for implementing Newton method
Q3	likelihood function derivation procedure for Newton algorithm MLE of (α, β)	R code for implementing Newton method
Q4	complete data likelihood function L derivation procedure for E step and M step MLE of $(\pi, \mu_1, \mu_2, \sigma_1, \sigma_2)$ The first ten students' genders you learned	R code for implementing EM algorithm