STAT 341: Assignment 2

DUE: Friday February 14 by 11:59pm EST

NOTES

Your assignment must be submitted by the due date listed at the top of this document, and it must be submitted electronically in .pdf format via Crowdmark. This means that your responses for different question parts should begin on separate pages of your .pdf file. Note that your .pdf solution file must have been generated by R Markdown. Additionally:

- For mathematical questions: your solutions must be produced by LaTeX (from within R Markdown). Neither screenshots nor scanned/photographed handwritten solutions will be accepted – these will receive zero points.
- For computational questions: R code should always be included in your solution (via code chunks in R Markdown). If code is required and you provide none, you will received zero points.
- For interpretation questions: plain text (within R Markdown) is required. Text responses embedded as comments within code chunks will not be accepted.

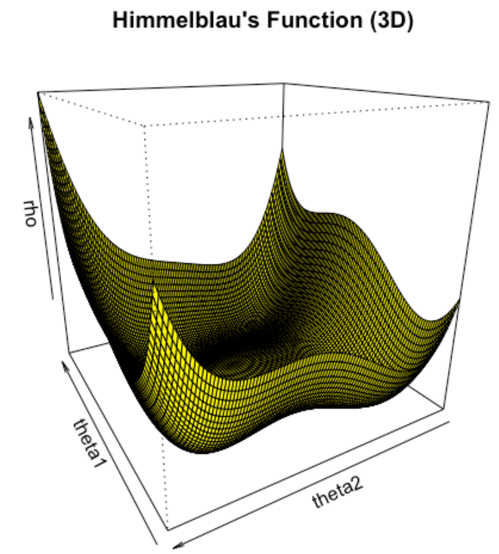
Organization and comprehensibility is part of a full solution. Consequently, points will be deducted for solutions that are not organized and incomprehensible.

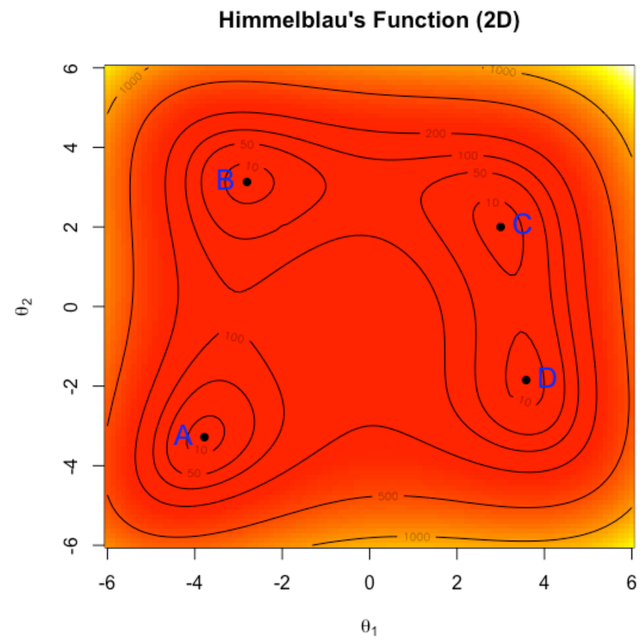
QUESTION 1: Himmelblau's Function [18 points]

Himmelblau's Function is one of many non-convex test functions used for evaluating the performance of optimization methods. For $\boldsymbol{\theta} \in \mathbb{R}^2$ the function is defined as follows

$$\rho(\boldsymbol{\theta}) = (\theta_1^2 + \theta_2 - 11)^2 + (\theta_1 + \theta_2^2 - 7)^2$$

The figures below depict the function (as a 3-dimensional surface and with 2-dimensional countours) for $\theta_1 \in [-6, 6]$ and $\theta_2 \in [-6, 6]$.





As can be seen in the figures, this function as 4 local minima labelled (in the contour plot) $A = (-3.779310, -3.283186)$, $B = (-2.805118, 3.131312)$, $C = (3, 2)$, $D = (3.584428, -1.848126)$. Himmelblau's Function attains a value of zero at each of these four points.

(a) [4 points] Write `rho` and `gradient` functions for Himmelblau's Function which take a single vector-valued input `theta`. Note that you may use without proof or derivation the fact that

$$\frac{\partial \rho}{\partial \theta_1} = 4\theta_1^3 + 4\theta_1\theta_2 - 42\theta_1 + 2\theta_2^2 - 14$$

and

$$\frac{\partial \rho}{\partial \theta_2} = 2\theta_1^2 + 4\theta_1\theta_2 + 4\theta_2^3 - 26\theta_2 - 22$$

(b) [5 points] In this question you will explore the surface of the Himmelblau Function using gradient descent. In particular you will consider 5 different starting values and explore the impact of changing one's starting location. Using the `gradientDescent` function (from class) together with the `gridLineSearch` and `testConvergence` functions (from class) as well as the `rho` and `gradient` functions from part (a), find the solution to

$$\underset{\boldsymbol{\theta} \in \mathbb{R}^2}{\operatorname{argmin}}\, \rho(\boldsymbol{\theta})$$

for each of the following five starting values. In each case state which minima you've converged to (A, B, C, or D) and be sure to include the output from the `gradientDescent` function.

i. $\widehat{\boldsymbol{\theta}}_0 = (0, 0)$

ii. $\widehat{\boldsymbol{\theta}}_0 = (0, 3)$

iii. $\widehat{\boldsymbol{\theta}}_0 = (3, 0)$

iv. $\widehat{\boldsymbol{\theta}}_0 = (0, -3)$

v. $\widehat{\boldsymbol{\theta}}_0 = (-3, 0)$

(c) [5 points] Recreate the contour plot shown above. You may find the functions `outer`, `image`, and `contour` useful for this task. Include on this plot `green` triangles at each of the starting points specified in (b) as well as `green` line `segments` connecting these starting points with their respective points of convergence.

(d) [2 points] Based on your investigations in parts (b) and (c) explain the importance of the starting value when performing non-convex optimization (when locating a global optimum is desired).

(e) [2 points] Using the `gradientDescent` function (from class) together with the `fixedLineSearch` and `testConvergence` functions (from class) as well as the `rho` and `gradient` functions from part (a), attempt to find the solution to

$$\underset{\boldsymbol{\theta} \in \mathbb{R}^2}{\operatorname{argmin}}\, \rho(\boldsymbol{\theta})$$

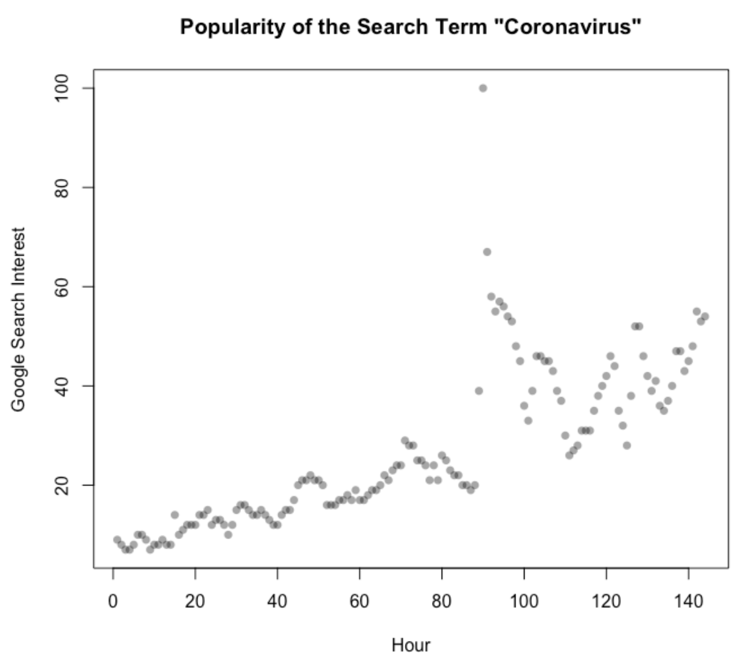
starting from $\widehat{\boldsymbol{\theta}} = (0, 0)$. Can you find input settings that lead to convergence? If so state them. Comment on the value of algorithmically choosing a step size versus using a fixed step size.

QUESTION 2: Google Trends "coronavirus" [25 points]

The 2019 Novel Coronavirus (2019-nCoV) is a recent strain of the coronavirus that impairs the respiratory systems of its host and that was first identified in Wuhan, Hubei Province, China. The first cases were identified in December of 2019 and since then thousands of cases have been confirmed around the world. Unsurpisingly, amidst this outbreak citizens of the world have been turning to the internet for more information.

In this question we will analyze Google Trends data which illustrates the interest among Canadians in the search term "coronavirus". The data available to you is stored in the `coronavirus.csv` file and summarized and visualized below. For each of the $N = 144$ hours in the time frame 1:00am January 21 – 1:00am January 28 we have recordings on the following variates:

Variate	Description
`TimeStamp`	A character string recorded as `YYYY-MM-DDTHH` indicating hours (i.e., `2020-01-22T05` corresponds to 5:00-6:00am on January 22nd)
`h`	A numeric variate indicating hours (i.e., `h=1` corresponds to 1:00-2:00am on January 21st and so on).
`interest`	A numeric variate that represents search interest. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular, etc.



(a) [2 points] As is demonstrated in the plot above, there is one hour in which Google searches for "coronavirus" in Canada increased dramatically. Determine the `TimeStamp` associated with this largest `interest` value. What is significant about this hour? (Hint: think current events).

(b) In this question you will fit the simple linear regression model

$$y_u = \alpha + \beta x_u + r_u, \quad u \in \mathcal{P}$$

using the least squares objective function

$$\rho(\boldsymbol{\theta}; \mathcal{P}) = \sum_{u \in \mathcal{P}} r_u^2$$

where $\boldsymbol{\theta} = (\alpha, \beta)^T$ and $r_u = y_u - \alpha - \beta x_u$.

i. [2 points] By taking appropriate derivatives, determine the 2×1 gradient vector $\boldsymbol{g} = \nabla \rho(\boldsymbol{\theta}; \mathcal{P})$. Show your work.

ii. [4 points] Write *factory functions* `createLSRho(x,y)` and `createLSGradient(x,y)` which take in as input only the data and which return as output the least squares objective function and the corresponding gradient function, respectively.

iii. [2 point] Using the `optim` function with the `rho` and `gradient` functions created by your factory functions from part ii., find $\widehat{\boldsymbol{\theta}} = \left(\widehat{\alpha}, \widehat{\beta}\right)$, the solution to

$$\underset{\boldsymbol{\theta} \in \mathbb{R}^2}{\operatorname{argmin}}\, \rho(\boldsymbol{\theta}; \mathcal{P})$$

Start the optimization at $\widehat{\boldsymbol{\theta}}_0 = (0, 0)$. For full points be sure to include the output from the `optim` function.

(c) In this question you will fit the simple linear regression model

$$y_u = \alpha + \beta x_u + r_u, \quad u \in \mathcal{P}$$

using the Huber objective function:

$$\rho(\boldsymbol{\theta}; \mathcal{P}) = \sum_{u \in \mathcal{P}} \rho_k(r_u)$$

where $\boldsymbol{\theta} = (\alpha, \beta)^T$, $r_u = y_u - \alpha - \beta x_u$ and

$$\rho_k(r) = \begin{cases} \frac{1}{2} r^2 & \text{for } |r| \le k \\ k\,|r| - \frac{1}{2} k^2 & \text{for } |r| > k \end{cases}$$

i. [2 points] By taking appropriate derivatives, determine the 2×1 gradient vector $\boldsymbol{g} = \nabla \rho(\boldsymbol{\theta}; \mathcal{P})$. Show your work.

ii. [4 points] Write *factory functions* `createHuberRho(x,y,k)` and `createHuberGradient(x,y,k)` which take in as input only the data and the threshold k, and which return as output the Huber objective function and the corresponding gradient function, respectively. Note that you may use the `huber.fn` and `huber.fn.prime` functions from class as necessary.

iii. [2 points] Using the `optim` function with the `rho` and `gradient` functions created by your factory functions from part ii., find $\widehat{\boldsymbol{\theta}} = \left(\widehat{\alpha}, \widehat{\beta}\right)$, the solution to

$$\underset{\boldsymbol{\theta} \in \mathbb{R}^2}{\operatorname{argmin}}\, \rho(\boldsymbol{\theta}; \mathcal{P})$$

while letting $k = 6.268$ and $\widehat{\boldsymbol{\theta}}_0 = (0, 0)$. For full points be sure to include the output from the `optim` function.

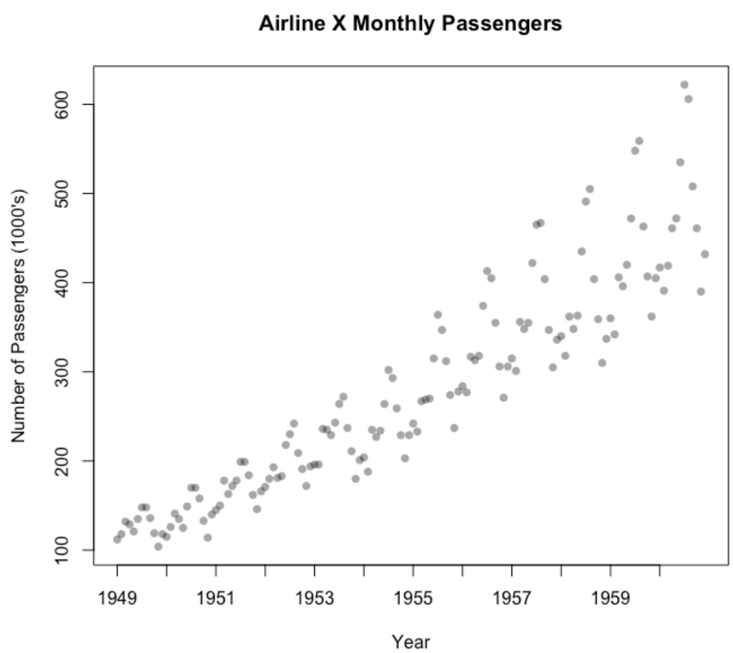
(d) [5 points] Recreate the scatter plot from above and overlay the least squares line and the robust regression line, distinguished by a legend.

(e) [2 points] Does the outlier identified in part (a) appear to have a large influence on the least squares line? State YES or NO and provide a brief justification.

QUESTION 3: Modeling Airline Ticket Sales [27 points]

The `AirPassengers` dataset is a classic dataset used in times series and forecasting. For Airline X (the actual airline is kept confidential due to privacy concerns) this dataset contains $N = 144$ observations corresponding to their monthly number of passengers (in thousands) from January 1949 to December 1960. The data is available in the `airpassengers.json` file and summarized and visualized below.

Variate	Description
`month`	A numeric variate indicating month. Note that `month=1` corresponds to January 1949, `month=2` corresponds to February 1949 and so on.
`passengers`	A numeric variate indicating the number of passengers (in 1000s) on Airline X flights in a given month.



(a) In this question you will fit the simple linear regression model

$$y_u = \alpha + \beta x_u + r_u, \quad u \in \mathcal{P}$$

via the Newton-Raphson Method in conjunction with the least squares objective function

$$\rho(\boldsymbol{\theta}; \mathcal{P}) = \sum_{u \in \mathcal{P}} r_u^2$$

where $\boldsymbol{\theta} = (\alpha, \beta)^T$ and $r_u = y_u - \alpha - \beta x_u$.

i. [4 points] Determine the vector $\boldsymbol{\psi}(\boldsymbol{\theta}; \mathcal{P})$ and matrix $\boldsymbol{\psi}'(\boldsymbol{\theta}; \mathcal{P})$ required to apply the Newton-Raphson method. Show your work.

ii. [4 points] Write *factory functions* `createLSpsi(x,y)` and `createLSpsiPrime(x,y)` which take in as input only the data and which return as output the `psi` and `psiPrime` functions associated with the least squares objective function (for the linear model).

iii. [2 points] Using the `NewtonRaphson` function (from class) together with the `testConvergence` function (from class) as well as `psi` and `psiPrime` functions created by your factory functions from part ii., find $\widehat{\boldsymbol{\theta}} = \left(\widehat{\alpha}, \widehat{\beta}\right)$, the solution to $\boldsymbol{\psi}(\boldsymbol{\theta}; \mathcal{P}) = \mathbf{0}$. Start the optimization at $\widehat{\boldsymbol{\theta}}_0 = (0, 0)$. For full points be sure to include the output from the `NewtonRaphson` function.

(b) In this question you will fit the linear regression model

$$y_u = \alpha + \beta x_u + \gamma x_u^2 + r_u, \quad u \in \mathcal{P}$$

via the Newton-Raphson Method in conjunction with the least squares objective function

$$\rho(\boldsymbol{\theta}; \mathcal{P}) = \sum_{u \in \mathcal{P}} r_u^2$$

where $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^T$ and $r_u = y_u - \alpha - \beta x_u - \gamma x_u^2$.

i. [4 points] Determine the vector $\boldsymbol{\psi}(\boldsymbol{\theta}; \mathcal{P})$ and matrix $\boldsymbol{\psi}'(\boldsymbol{\theta}; \mathcal{P})$ required to apply the Newton-Raphson method. Show your work.

ii. [4 points] Write *factory functions* `createLSQpsi(x,y)` and `createLSQpsiPrimew(x,y)` which take in as input only the data and which return as output the `psi` and `psiPrime` functions associated with the least squares objective function (for the quadratic model).

iii. [2 points] Using the `NewtonRaphson` function (from class) together with the `testConvergence` function (from class) as well as `psi` and `psiPrime` functions created by your factory functions from part ii., find $\widehat{\boldsymbol{\theta}} = \left(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}\right)$, the solution to $\boldsymbol{\psi}(\boldsymbol{\theta}; \mathcal{P}) = \mathbf{0}$. Start the optimization at $\widehat{\boldsymbol{\theta}}_0 = (0, 0, 0)$. For full points be sure to include the output from the `NewtonRaphson` function.

(c) [5 points] Recreate the scatter plot from above and overlay the least squares line and the least squares quadratic curve, distinguished by a legend.

(d) [2 points] Based on the plot in part (c), which model best represents the relationship in the population – the line or the quadratic curve? Briefly justify your response.