CMPT 726/410 Fall 2023 Assignment 2

Your Name Here (Your Student ID #)

November 5, 2023

Due: Thursday, November 23, 2023 at 11:59 pm Pacific Time

Important: Make sure to download the zip archive associated with this assignment, which contains essential data and starter code that are required to complete the programming component of the assignment.

This assignment is designed to be substantially more challenging than the quizzes and requires thorough understanding of the course material and extensive thought, so **start early!** If you are stuck, come to office hours. Make sure to check the discussion board often for updates, clarifications, corrections and/or hints.

Partial credit will be awarded to reasonable attempts at solving each question, even if they are not completely correct.

There will be office hours dedicated to assignment-related questions. Times and Zoom links will be posted on Canvas. If you have a question that you would like to ask during office hours, make sure to post a brief summary of your question in the office hours thread on the Canvas discussion board. During office hours, questions will be answered in the order they appeared on the office hours thread. We may not be able to get to all questions, so please start early and plan ahead.

Requests for extensions will not be considered under any circumstances. Make sure you know how to submit your assignment to Crowdmark and leave sufficient time to deal with any technical difficulties, Internet outages, server downtimes or any other unanticipated events. It is possible to update your assignment after submission, so we recommend uploading a version of your assignment at least several hours before the deadline, even if it is incomplete.

Submission Instructions

Carefully follow the instructions below when submitting your assignment. Not following instructions will result in point deductions.

1. You should typeset your assignment in LaTeX using this document as a template. We recommend using Overleaf to compile your LaTeX. Include images in the section they go with and **do not** put them in an appendix.

2. At the beginning of the assignment (see above), please copy the following statement and sign your signature next to it. (macOS Preview and FoxIt PDF Reader, among others, have tools to let you sign a PDF file.) We want to make it *extra* clear so that no one inadvertently cheats.

   "I certify that all solutions are entirely in my own words and that I have not looked at another student's solutions. I have given credit to all external sources I consulted."

3. At the start of each problem—not subproblem; you only need to do this twice—please state: (1) who you received help from on the problem, and (2) who you provided help to.

4. For all theory questions, make sure to **show your work** and include all key steps in your derivations and proofs. If you apply a non-trivial theorem or fact, you must state the theorem or fact you used, either by name (e.g.: Jensen's inequality) or by writing down the general statement of the theorem or fact. You may use theorems and facts covered in lecture without proof. If you use non-trivial theorems and facts not covered in lecture, you must prove them. All conditions should be checked before applying a theorem, and if a statement follows from a more general fact, the reason why it is a special case of the general fact should be stated. The direction of logical implication must be clearly

stated, e.g.: if statement A implies statement B, statement B implies statement A, or statement A is equivalent to statement B (i.e.: statement A is true if and only if statement B is true). You should **number** the equations that you refer to in other parts and refer to them by their numbers. You must **highlight the final answer or conclusion** in your PDF submission, either by changing the font colour to red or drawing a box around it.

5. For all programming questions, starter code is provided as a Jupyter notebook. You should work out your solution in the notebook. We recommend uploading your notebook to a hosted service like Google Colab to avoid the installation of dependencies and minimize setup time. When you are done, take a screenshot of your code and output and include it in your PDF. In addition, you should download the Jupyter notebook with your solutions (`File` → `Download` → `Download .ipynb`) and also download your code (`File` → `Download` → `Download .py`). You need to **submit both** as a zip archive named "`CMPT726-410_A2_⟨Last Name⟩_⟨Student ID⟩.zip`", whose directory structure should be the same as that of the zip archive for the starter code. Do **NOT** include any data files we provided. Please include a short file named **README** listing your name and student ID. The PDF should not be included in the zip archive and should be submitted **separately**. Please make sure that your code doesn't take up inordinate amounts of time or memory. If your code cannot be executed, your solution cannot be verified.

6. Assignment submissions will be accepted through Crowdmark. The submission portal will open a week before the assignment is due. Make sure to set aside at least **one hour** to submit the assignment. Crowdmark will ask you to split the PDF by each part of each question. Make sure to upload the correct pages for each part, and double check when you are done.

Python Configuration

We recommend using Google Colab to complete the parts of this assignment that require coding. However, if you would like to set up your own Python environment on your computer, follow the instructions below.

1. Ensure you have Python 3 installed. If you don't, we recommend miniconda as a package manager. To ensure you're running Python 3, open a terminal in your operating system and execute the following command: `python --version` **Do not proceed until you're running some (recent) version of Python 3.**

2. Install `scikit-learn` and `scipy`:
   `conda install -y -c conda-forge scikit-learn scipy matplotlib`

3. To check whether your Python environment is configured properly for this homework, ensure the following Python script executes without error. Pay attention to errors raised when attempting to import any dependencies. Resolve such errors by manually installing the required dependency (e.g. execute `conda install -c conda-forge numpy` for import errors relating to the numpy package).

```
import sys
if sys.version_info[0] < 3:
    raise Exception("Python 3 not detected.")

import numpy as np
import matplotlib.pyplot as plt
from scipy import io
```

4. Please use only the packages listed above.

Collaboration and Academic Integrity

While collaboration is encouraged, all of your submitted work must be your own. Concretely, that means you may discuss general approaches to problems with other students, but you must write your solutions on your own. For more information, see the course syllabus under "Academic Integrity." If you got help from or gave help to other students, please note their names at start of every question. Additionally, sign the declaration at the bottom of this page.

Warning: Copying others' solutions, seeking help from others not in this course, posting questions online or entering questions into automated question answering systems are considered cheating. Consequences are severe and could lead to suspension or expulsion. If you become aware of such instances, you must report them here: https://forms.gle/9fw3oMLyhD1A81qy5.

Page Numbers

Declaration

I certify that all solutions are entirely in my own words and that I have not looked at solutions other than my own. I have given credit to all external sources I consulted.

Your Name Here (Your Student ID #)

Question 1 Parameter Estimation [35 points]

In the following parts, consider a random variable X that follows a distribution parameterized by a parameter θ, whose probability density is:

$$p(x|\theta) = \sqrt{\frac{2}{\pi}} \frac{x^2}{\theta^3} \exp\left(\frac{-x^2}{2\theta^2}\right) \qquad \theta > 0, x > 0$$

List your collaborators on this question below.

Part 1(a) [5 points]

Write down the likelihood function $\mathcal{L}(\theta; \mathcal{D})$ of the parameter θ given the training data $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$, where $x_1, x_2 \ldots, x_N$ are assumed to have been drawn independently of each other from identical distributions. Show your work.

Part 1(b) [5 points]

Write down the log-likelihood function $\log \mathcal{L}(\theta; \mathcal{D})$ **of the parameter** θ **given the training data** $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$**, where** $x_1, x_2 \ldots, x_N$ **are assumed to have been drawn independently of each other from identical distributions.** Move the logs in as much as possible and simplify. Show your work.

Part 1(c) [5 points]

Derive the expression for critical point(s) of the log-likelihood. Show your work.

Part 1(d) [5 points]

Prove that the critical point found in the previous part is a local maximum. Then using the fact that there is only one critical point, show that the critical point is the global maximum. Show your work and justify each step that is not a simple algebraic manipulation.

Part 1(e) [5 points]

Given a prior over θ, i.e., $\theta \sim \mathcal{N}(0, 1)$, write down the objective function for finding the MAP estimate of the parameter θ. Simplify the objective function to a form that can be differentiated to find the critical point(s) in closed form easily. Justify each step that is not a simple algebraic manipulation.

Part 1(f) [5 points]

Differentiate the objective function in the previous part to find the expression for the critical point. Then prove it is a local maximum and a global maximum. Show your work and justify each step that is not a simple algebraic manipulation.

Part 1(g) [5 points]

Consider an arbitrary probability distribution $p(x|\theta)$ parameterized by a scalar parameter $\theta \in [a, b]$. **Given that the maximum likelihood estimate (MLE) and maximum a posteriori (MAP) estimate of θ are the same, can you find the prior distribution of θ?** If so, state the name of the distribution and its parameters, and derive it. If not, explain why.

Question 2 Linear Regression [20 points]

List your collaborators on this question below.

Part 2(a) [5 points]

Consider a vector $\vec{v} \in \mathbb{R}^n$, where $n \geq 2$. Given the first element of $\vec{v}$ and the differences between adjacent elements of $\vec{v}$, our goal is to reconstruct the vector.

Formulate an ordinary least squares (OLS) problem such that the input data A and desired outputs $\vec{y}$ are derived from the quantities we are given and the solution (i.e., $\vec{w}^* := \arg\min_{\vec{w}} \|A\vec{w} - \vec{y}\|_2^2$) can be turned into a reconstruction of the vector. **Write down what each element of A, $\vec{y}$ and $\vec{w}$ corresponds to. What are the dimensions of A, y and $\vec{w}$?** Justify your answer.

Part 2(b) [5 points]

Any grayscale image can be represented as a matrix I where $I_{i,j}$ is the intensity of the pixel on the ith row and jth column. Consider an $n \times n$ image $I \in \mathbb{R}^{n \times n}$, where $n \geq 2$. Given the values of I in the first column and the horizontal image gradients, which are the differences between values in adjacent columns, our goal is to reconstruct the full image.

More precisely, we denote the horizontal image gradient at a pixel in the ith row and jth column as $\Delta^x_{i,j}$, which is defined to be $I_{i,j+1} - I_{i,j}$. We are given $\Delta^x_{i,j}$ for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, n-1\}$ and $I_{i,1}$ for all $i \in \{1, \ldots, n\}$, and would like to find $I_{i,j}$ for all $i \in \{1, \ldots, n\}$ and $j \in \{2, \ldots, n\}$.

Formulate an ordinary least squares (OLS) problem such that the input data A and desired outputs $\vec{y}$ are derived from the quantities we are given and the solution (i.e., $\vec{w}^* := \arg\min_{\vec{w}} \|A\vec{w} - \vec{y}\|_2^2$) can be turned into a reconstruction of the image. **Write down what each element of A, $\vec{y}$ and $\vec{w}$ corresponds to. What are the dimensions of A, y and $\vec{w}$?** Justify your answer.

Part 2(c) [5 points]

Unlike in the previous part, we are no longer given the values of I in the first column. However, in addition to being given the horizontal image gradients, we are also given the vertical image gradients, which are the differences between values in adjacent rows. Our goal is still to reconstruct the full image.

More precisely, we will denote horizontal image gradients in the same way as before, and the vertical image gradient at a pixel in the ith row and jth column as $\Delta^y_{i,j}$, which is defined to be $I_{i+1,j} - I_{i,j}$. We are given $\Delta^x_{i,j}$ for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, n-1\}$, and $\Delta^y_{i,j}$ for all $i \in \{1, \dots, n-1\}$ and $j \in \{1, \dots, n\}$.

We are given $\Delta^x_{i,j}$ for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, n-1\}$ and $\Delta^y_{i,j}$ for all $i \in \{1, \dots, n-1\}$ and $j \in \{1, \dots, n\}$, and would like to find $I_{i,j}$ for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, n\}$.

Formulate an ordinary least squares (OLS) problem such that the input data A and desired outputs $\vec{y}$ are derived from the quantities we are given and the solution (i.e., $\vec{w}^* := \arg\min_{\vec{w}} \|A\vec{w} - \vec{y}\|_2^2$) can be turned into a reconstruction of the image. **Write down what each element of A, $\vec{y}$ and $\vec{w}$ corresponds to. What are the dimensions of A, y and $\vec{w}$?** Justify your answer.

Part 2(d) [5 points]

In the `x_gradients.npy` and `y_gradients.npy` files, you are given the horizontal image gradients and vertical image gradients of an unknown image. Write an implementation that constructs the input data matrix A and the vector of desired outputs $\vec{y}$ used in the OLS formulation in the previous part. Then write an implementation that solves the OLS problem using the closed-form formula for the solution. You can only use functions that implement basic linear algebra operations, such as matrix multiplication, inverses, transposes and solving a linear system of equations where the number of equations is equal to the number of unknowns. You have been provided with starter code for loading and visualizing data.

Include a screenshot of the relevant part of your code and the reconstructed image in your PDF submission.

Part 2(e) [Bonus: 3 points]

In one or two sentences comment on the memory requirement of the approach implemented for the previous part and if it can be reduced by considering the characteristics of input data matrix A.