Tashfeen, Ahmad
CSCI 6213: Data Science Fundamentals

Homework 1

Please show all work for full credit. You're only allowed to use the standard Python 3 libraries.

Question 1. Watch Mary's Room: A philosophical thought experiment – Eleanor Nelsen on YouTube. Do you think there is something about human learning that is different than machine learning?

Question 2. Rate your skills out of ten in the following subjects,

1) Probability and Statistics
2) Calculus
3) Linear Algebra
4) Python

Question 3. Of the sub-plots (a), (b), (c) and (d) in figure 1, state which ones have no correlation, negative correlation, position correlation and non-linear correlation.

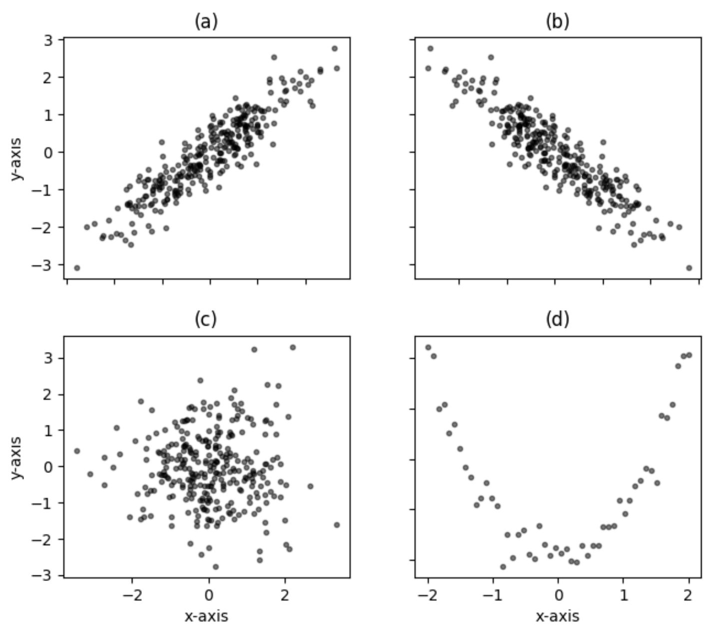

FIGURE 1. Different types of correlations.

Question 4. Assume that the probability of Alice going to class everyday given she got a good grade is $\frac{5}{6}$ and the probability of her getting a good grade is $\frac{2}{5}$ while the probability of her going to class everyday is $\frac{1}{3}$. What is the probability that Alice gets a good grade given she went to class everyday?
Hint: Thomas Bayes.

Question 5. We define the natural numbers as,

$$\mathbb{N} = \{1, 2, 3, 4, 5, ...\}$$

Let x be the arithmetic mean of the first $2^{33} - 1$ natural numbers. What is $\log_2(x)$?
Hint: Carl Friedrich Gauss.

Question 6. Let `arr = [x % 2 for x in range((2**100)-1)]` be a Python list using Python list comprehension. What is the statistical mode of the list `arr`? Justify your answer.

Question 7. Define a function $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \frac{\sin(x^2)}{y}$$

Calculate the following,

- The partial derivative with respect to y, $\dfrac{\partial f}{\partial y}$
- The partial derivative with respect to x, $\dfrac{\partial f}{\partial x}$
- The gradient vector $\nabla f(x, y)$.

Question 8. Let $\vec{v_1} = [e, \pi, \sqrt{2}]$ and $\vec{v_2} = [1, 2, 0]$. What is the dot product $\vec{v_1} \cdot \vec{v_2}$? Please do not give an approximation.

Question 9. We define a matrix of n rows and p columns with real values as $\mathbf{A} \in \mathbb{R}^{n \times p}$. We can think of $\mathbf{A}$ as having n row vectors with the $i^{\text{th}} \leq n$ row vector being $\text{row}(\mathbf{A})_i$. Similarly, we can think of $\mathbf{A}$ as having p column vectors with the $j^{\text{th}} \leq p$ column vector being $\text{col}(\mathbf{X})_j$.

Then matrix multiplication between $\mathbf{A} \in \mathbb{R}^{n \times p}$ and $\mathbf{B} \in \mathbb{R}^{p \times m}$ is defined as,

$$\mathbf{AB} = \begin{bmatrix} \text{row}(\mathbf{A})_1 \,\text{col}(\mathbf{X})_1 & \text{row}(\mathbf{A})_1 \,\text{col}(\mathbf{X})_2 & \cdots & \text{row}(\mathbf{A})_1 \,\text{col}(\mathbf{X})_m \\ \text{row}(\mathbf{A})_2 \,\text{col}(\mathbf{X})_1 & \text{row}(\mathbf{A})_2 \,\text{col}(\mathbf{X})_2 & \cdots & \text{row}(\mathbf{A})_2 \,\text{col}(\mathbf{X})_m \\ \vdots & \vdots & \ddots & \vdots \\ \text{row}(\mathbf{A})_n \,\text{col}(\mathbf{X})_1 & \text{row}(\mathbf{A})_n \,\text{col}(\mathbf{X})_2 & \cdots & \text{row}(\mathbf{A})_n \,\text{col}(\mathbf{X})_m \end{bmatrix}$$

where $\text{row}(\mathbf{A})_i \,\text{col}(\mathbf{X})_j \in \mathbb{R}$ is a dot or inner product.

Further, the transpose of $\mathbf{A} \in \mathbb{R}^{n \times p}$ is written as $\mathbf{A}^T \in \mathbb{R}^{p \times n}$ and is defined as,

$$\text{If } \mathbf{A}^T \text{ is the transpose of } \mathbf{A} \text{ then } \text{row}(\mathbf{A})_i = \text{col}(\mathbf{A}^T)_j.$$

(a) What must be true of the number of columns of $\mathbf{A}$ and number of rows of $\mathbf{B}$ for $\mathbf{AB}$ to be defined?
(b) What information do the number of rows of $\mathbf{A}$ and number of columns of $\mathbf{B}$ give you about the dimensions of the product $\mathbf{AB}$?
(c) For the two matrices bellow, give the product $\mathbf{AB}$.

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 4 & 2 & 2 \end{bmatrix}$$

(d) Prove that the matrix product is not a symmetric relation, i. e., $\mathbf{AB} \neq \mathbf{BA}$. Hint: can you construct a small counter example?
(e) Prove that $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$.
(f) For $\mathbf{X} \in \mathbb{R}^{n \times p}, \beta \in \mathbb{R}^p, y \in \mathbb{R}^n$ prove that,

$$\beta^T \mathbf{X}^T y = y^T \mathbf{X} \beta$$

(g) Let $f(\beta) = \beta^T \beta$. Prove that the derivative $\frac{\mathrm{d}f}{\mathrm{d}\beta} = 2\beta$.

Question 10. For $x \in \mathbb{R}$ we have,

$$f(x) = \frac{x^4}{4} - \frac{x^3}{3}$$

Give,

$$\min_x f(x) = \min_x \left(\frac{x^4}{4} - \frac{x^3}{3} \right)$$

Plot f to double check your answer.

Question 11. Let $x \sim \mathcal{N}(0,1)$ be a normally distributed random variable with mean 0 and standard deviation 1. What is the likelihood that $x = \sqrt{2}$. Give an exact answer.

Question 12. The code snippet in listing 1 reads the plain text of the novel *The Cosmic Computer* by the famous American science fiction writer Henry Beam Piper over the internet and saves it to a string variable `text`.

```python
from urllib.request import urlopen as get

url = 'https://www.gutenberg.org/files/20727/20727.txt'
with get(url) as response:
    text = response.read().decode('utf-8')

print(text)
```

LISTING 1. A Python program to download the text of the *The Cosmic Computer* by Piper.

Write a Python program that prints out the ten most used words in the novel that have more than 5 letters. State your findings. Put your code in a file called `frequency.py`.

Question 13. Watch Laziness in Python - Computerphile on YouTube.
Implement the function `fibonacci(n)` in the code listing 2.

```python
def fibonacci(n):
    # implement me

for t in fibonacci(50):
    print(t)
```

LISTING 2. A Python 3 program to print out the first $n \geq 0$ Fibonacci numbers.

Run your program for $n = 50$. What are the last five numbers printed? Put your code in a file called `fibonacci.py`.

Submission Instructions

1) Submit a PDF that answers all the questions.
2) Submit Python files, e. g., `frequency.py` and `fibonacci.py`.

Oklahoma City University, Petree College of Arts & Sciences, Computer Science