Math 7243 Machine Learning - Fall 2022

Instructor: He Wang

Midterm

Student Name: _____ _____**/50**

Rules and Instructions for Exams:

1. Unless otherwise specified, to receive full credits you must show **all** necessary work. The grading is based on your work shown. Only a final result from computer will receive zero point.

2. You need to finish the exam yourself. Any discussions with the other people will be considered as **academic dishonesty**. **Cheating, Unauthorized Collaboration, and Facilitating Academic Dishonesty are not allowed.** You can read a description of each here http://www.northeastern.edu/osccr/academic-integrity-policy/

3. This is an open exam. You are allowed to look at textbooks, and use a computer.

4. You are **not** allowed to discuss with any other people.

5. You are **not** allowed to ask questions on any internet platform.

6. For programming questions, if there is no specific instruction, you can only use **numpy**, **matplotlib** library. You should **not** use any build in function from Scikit-learn or StatsModels libraries.

7. **Submit your codes for all questions if you used python.** For your **Python coding and graphing submissions**, you can either paste the images of the key codes along your solutions in pdf file, or submit a separate .**html** file.

1. (10 points) Calculate the **gradient** and **Hessian matrix** of the following functions and find the **argmin**$_\theta$ of each function. Here the norm $||\ \ ||$ is the standard l_2-norm. You can use any results in the lecture notes.

(1) Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix, and $\vec{b} \in \mathbb{R}^n$. Suppose $A + 3I$ is positive definitive.

Let $J(\vec{\theta}) = \vec{\theta}^T A\vec{\theta} - 2\vec{b}^T\vec{\theta} + 3\vec{\theta}^T\vec{\theta} + ||A\vec{b}||$.

(2) Let $J(\vec{\theta}) = \theta_1^2 + 4\theta_1\theta_2 + 5\theta_2^2 - 4\theta_1 - 6\theta_2 + 10$

2. (10 points) In this question, you may use Python (with only numpy library) to solve the matrix equation. Consider the following data points

x_1	x_2	y
1	2.1	2.3
2	4.1	4.3
3	5.9	6.3
4	8.2	7.8
5	9.9	9.8

a). Fit a linear model $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$ to this dataset when the loss is RSS= $||X\vec{\theta} - \vec{y}||^2$. You should report the best fit function and the RSS cost value. Plot the data and model.

b). Fit a linear function to this dataset when the loss is the Ridge Loss $J(\theta) = ||X\vec{\theta} - \vec{y}||^2 + \lambda(\theta_1^2 + \theta_2^2)$ with $\lambda = 1$ and with $\lambda = 10$. You should report the best fit function and the **RSS** cost value. Plot the data and model. (Hint: Do not put penalty on θ_0. You don't have to standardize the data, but you need to centralize the data.)

3. (10 points) Consider the data

$x^{(i)}$	0	0.2	0.4	0.6	0.8	1	1.2	1.4
$y^{(i)}$	5.1	6.4	6.1	8.2	9.5	8.6	12	14.8

The data file $\{\vec{x}^{(i)}, y^{(i)}\}$ for $i = 1, 2, ..., n = 8$ is drawn (with noise) from

$$f(x) = \theta_0 + \theta_1 x + \theta_2 e^x$$

(1) Find a **closed formula** for parameters $\vec{\theta}$ to minimize the RSS loss

$$J(\vec{\theta}) = \sum_{i=1}^{n} \left(y^{(i)} - f(x^{(i)})\right)^2$$

(2) Using formula in (1), **find the function** $f(x)$ fitting the data.

(3) Calculate the cost for your fitting in (2).

4. (10 points) Consider the categorical learning problem consisting of a data set with two labels.

Label 1: (contains 5 data points)

X_1	-0.6	0	0.2	-0.8	-0.3
X_2	-2	-2	-1	-1	0

Label 0: (contains 6 data points)

X_1	0.5	1	1.5	1.5	1	0.5
X_2	2.5	1	1.5	2.5	3	1.5

Answer the following logistics regression questions.

(1) Use **gradient descent** to find the **logistic regression** model

$$p(Y = 1|\vec{x}) = \frac{1}{1 + e^{-\theta^T \vec{x}}}$$

and the boundary. (Plot the data and boundary, only use numpy and Matplotlib.) Use initial value $\vec{\theta}_0 = \vec{0}$, learning rate $\alpha = 0.02$, and 1000 iterations,

(2) Find the probability $P(y = 1|\vec{x}_t)$ for a test point $\vec{x}_t = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ for your logistic model in (1). What is the predicted label for $\vec{x}_t$?

(3) Find **quadratic** Logistics Regression method for this question and obtain an quadratic boundary. (Hint: this means to use new features: X_1, X_2, X_1^2, X_1X_2, X_2^2.)

5. (10 points) Consider the categorical learning problem consisting of a data set with two labels:

Label 1: (contains 5 data points)

X_1	-0.6	0	0.2	-0.8	-0.3
X_2	-2	-2	-1	-1	0

Label 2: (contains 6 data points)

X_1	0.5	1	1.5	1.5	1	0.5
X_2	2.5	1	1.5	2.5	3	1.5

(1) For each label above, the data follow a multivariate normal distribution Normal(μ_i, Σ) where the covariance Σ is the same for both labels. Fit a pair of LDA functions to the labels by computing the covariances Σ, means μ_i, and proportion ϕ of data. You may use Python (with only numpy library)

(a) You should report the values for ϕ, μ_i and Σ.

(b) Give the **formula for the line** forming the decision boundary and plot the **graph**. (Write done the formula used in your calculation.)

(2) Find the probability $P(y = 1|\vec{x})$ for a test point $\vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ for the LDA model.

(3) (2 bonus points) Find the quadratic boundary using the QDA method and plot the graph. (You can use Sympy lib to simplify the formula.)