Math of AI (2023)

Assignment II

Shiraz University

Due date: 22 Dec (2023)

Part One:

1- **Given the instance-feature matrix $X_{N\times D}$, and a normal direction $\vec{W}_D$ on which instances are projected to computed the new feature $\vec{f}_N$. Compute the variance of $\vec{f}$.**

2- **Given grades of a student in D exams as a vector $\vec{g}_D$, it is desired to assign a positive weight (w) to each exam in order to maximize the final grade $\sum g_i \,.\, w_i$. However, it is desired to limit the weights such that $\sum w_i^2 = 1$ (the weight vector should be normal). What is the best weight vector?**

3- **Having N instances with D features in the matrix $X_{N\times D}$, it is desired to find a representative D −vector $\vec{r}$ in order to minimize degree q^{th} of norm of distances between instances and the representative. Distance is also computed based on the norm of degree p (i.e., $\|\vec{x}_n - \vec{r}\|_p$).**

   a. Model this minimization problem in the form of linear algebra.

   b. Find the best solution for $q = p = 1$ and $q = \infty$, $p = 2$.

   c. Determine q and p for which, r_i is Mode of feature $\vec{X}_i$.

4- **A dice is tossed N times and the sides 1, 2, …, & 6 are observed N_1, N_2, … , N_6 times, respectively. If the probability of observing i^{th} side is p_i, you know that probability of this set of observations is**

$$\prod_{i=1}^{6} p_i^{N_i}$$

It is desired to model the problem of finding maximum probability in the form of linear algebra with both objective function and constraint(s).

Assume that lg ($\vec{a}$) = [lg a_1, lg a_2, …, lg $a_{|\vec{a}|}$]^T.

5- **Having a graph with adjacency matrix M where $m_{i,j}$ is the probability of moving from city i to j. What does M^2 mean? What is Column-Row interpretation of M^2?**

6- **Analyze the covariance matrix from viewpoints of elementwise, column-wise and row-wise matrix production. Hint: $\sum = X^T X$ where, X is the demeaned instance-feature matrix.**

Part Two:

In this part, you will create a Python program to minimize different error norms in the context of linear regression. You must generate a set of data points consisting of x and their corresponding y values. The predicted value, $\hat{y}$, is defined as $\hat{y} = x * \alpha + \beta$, where α and β are constants.

Your task is to minimize several error norms, which include:

- **L0 Norm**
- **L1 Norm**
- **L2 Norm**
- **Infinity Norm**

The signed error is calculated using the formula:

Signed Error = $\hat{y} - y$.

Submission Guidelines:

- The answers to the exercises of part one must be in PDF file.
- Allowed programming language: Python
- Your implementation reports should be in a PDF file including an explanation of your approach, key points of your implementation, and report of your final results.
- You should upload your submissions at Quera. All of the files should be saved in a ZIP file named in this format: “Lastname-SudentNumber.zip”.
  Ex.: “Zamani-40230401.zip”
