Indian Institute of Technology Kharagpur
AI61003, Autumn 2023,
Assignment due on November 15, 2023

ANSWER ALL THE QUESTIONS

1. Download the MNIST dataset (both train and test sets) and vectorize each data point. Select samples from classes 1 and 7 and assign their corresponding labels as $+1$ and -1, respectively. Create an 80-20 train-test split. Using Least Squares, train your linear model on the training set and report the classification accuracy and the confusion matrix on the test set.

2. Download the MNIST dataset (both train and test sets) and vectorize each data point. Select samples from class i, in the train set, and assign the corresponding label as $+1$ where $i \in C = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. All samples that belong to the classes $C \setminus i$ are assigned the label -1. To achieve a balanced dataset, randomly sample x number of data points from the negative class where x is the number of data points in the positive class. Using Least Squares, train 10 separate linear regression models, one for each class in C. For each sample in the test set, get a prediction from all 10 linear models and assign it the class for which it receives the highest score. Finally, report the accuracy and confusion matrix.

3. Generate two vectors, $p \in \mathbb{R}^{100}$ and $q \in \mathbb{R}^{100}$ sampled from a uniform random distribution with the range $[-10, 10]$. Generate the target variable, b, in the following manner:

$$b_i = \begin{cases} +1 & \text{if } p_i q_i > 1 \\ -1 & \text{otherwise} \end{cases}$$

   Here, b_i, p_i and q_i denote the i-th index of the corresponding vectors b, p and q, respectively. Now, generate the following basis functions,

$$f_0(p_i, q_i) = 1 \quad f_1(p_i, q_i) = p_i \quad f_2(p_i, q_i) = q_i$$
$$f_3(p_i, q_i) = p_i^2 \quad f_4(p_i, q_i) = q_i^2 \quad f_5(p_i, q_i) = p_i q_i$$

   Use the basis functions to define $A \in \mathbb{R}^{100\times 6}$ where $A_i^j = f_j(p_i, q_i)$. Here i is the row index while j is the column index. Use least squares to predict the target variable b by learning the coefficient vector $x \in \mathbb{R}^6$. Report x.

4. Generate two vectors, $p \in \mathbb{R}^{100}$ and $q \in \mathbb{R}^{100}$ sampled from a uniform random distribution with the range $[-1, 1]$. The i-th element of the target variable, b, is defined as,

$$b_i = p_i q_i + p_i^2 + q_i^2$$

   Now, generate the following basis functions,

$$f_0(p_i, q_i) = 1 \quad f_1(p_i, q_i) = p_i \quad f_2(p_i, q_i) = p_i$$
$$f_3(p_i, q_i) = p_i^2 \quad f_4(p_i, q_i) = q_i^2 \quad f_5(p_i, q_i) = p_i q_i$$

   Use the basis functions to define $A \in \mathbb{R}^{100\times 6}$ where $A_i^j = f_j(p_i, q_i)$. Here i is the row index while j is the column index. Use least squares to predict the target variable b by learning the coefficient vector $x \in \mathbb{R}^6$. Report x as well as the Mean Squared Error (MSE).

5. Generate a vector $p \in \mathbb{R}^{100}$ from a uniform random distribution with range $[0, 1]$. The i-th element of the target variable, b, is defined as,

$$b_i = 7p_i - 3p_i^2$$

where p_i is the i-th element of vector p. For $n \in \mathbb{N}$, we define basis functions $\{f_1, f_2, \ldots, f_n\}$.
For $j = 1, 2, \ldots, n$, the basis function $f_j(t)$ is defined as follows.

- $\forall\ t\ \in [0,1] \cap \left[\frac{j-1}{n}, \frac{j}{n}\right]^C,\ f_j(t) = 0.$
- On the interval $\left[\frac{j-1}{n}, \frac{j}{n}\right]$, The graph of the function $f_j(t)$ is a triangle with its vertices at $\left(\frac{j-1}{n}, 0\right)$, $\left(\frac{j}{n}, 0\right)$ and $(\delta, 1)$ where δ is the mid-point of $\frac{j-1}{n}$ and $\frac{j}{n}$.

(a) For $n = 10$ and $n = 50$, use the least squares method to predict the target variable b by learning the coefficient vector $x \in \mathbb{R}^n$ by arranging the problem in the standard form $Ax = b$. Report x as well as the Mean Squared Error (MSE).

(b) For $n = 150$ and $n = 200$, use the multiobjective least squares method to predict the target variable b by learning the coefficient vector $x \in \mathbb{R}^n$ by arranging the problem in the standard form $Ax = b$. The two objectves J_1 and J_2 are $J_1 = \frac{1}{n}\sum_{i=1}^{n}(b_i - Ax)^2$ and $J_2 = \|x\|_2^2$. Minimize $J_1 + \lambda J_2$ for 5 different randomly chosen values of λ in $(0, 0.2)$. Report x as well as the Mean Squared Error (MSE).

6. Download the *auto-regressive-data.csv* file. The target variable is the *Consumption* column. Create lag features for each sample with the number of lag variables set to 8. Doing so will generate an input-output pair (x_i, y_i) where $x_i^j = y_{i-j}$ and $y_i = y_{i+1}$. Here, x_i^j denotes the i-th sample and j-th column. Train a linear regression model on this dataset to predict the next day's consumption value.

7. Take a single sample from the MNIST dataset. Flatten the image into vector $x \in \mathbb{R}^{784}$. Create the following Gaussian blurring kernel,

$$K = \begin{bmatrix} \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \end{bmatrix}$$

Now generate $A \in \mathbb{R}^{784\times 784}$, the Toeplitz matrix of K. Finally, create the blurred image of $y = Ax$. We now have to de-blurr y. This can be done by minimizing the following cost function with respect to $\hat{x}$,

$$\|A\hat{x} - y\|^2 + \lambda\|D_h\hat{x}\|^2 + \lambda\|D_v\hat{x}\|^2 \tag{1}$$

where $\hat{x} \in \mathbb{R}^{784}$ is the flattened, estimated, de-blurred image while λ is a hyper-parameter that needs to be set according to your preference (try starting out with

$\lambda = 0.007$). Here, $D_h \in \mathbb{R}^{(27.28\times 28.28)}$ and $D_v \in \mathbb{R}^{(27.28\times 28.28)}$ where,

$$D_h = \begin{bmatrix} -I & I & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -I & I & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -I & I & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -I & I \end{bmatrix}$$

$$D_v = \begin{bmatrix} D & 0 & \cdots & 0 \\ 0 & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D \end{bmatrix}$$

$$D = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}$$

and $I \in \mathbb{R}^{28\times 28}$ while $D \in \mathbb{R}^{27\times 28}$.

8. Generate two vectors, $p \in \mathbb{R}^{100}$ and $q \in \mathbb{R}^{100}$ sampled from a uniform random distribution with the range $[-1, 1]$. The i-th element of the target variable, b, is defined as,

$$b_i = p_i q_i + p_i^2 + q_i^2 \tag{2}$$

Now, generate the following basis functions,

$$\begin{aligned} &f_0(p_i, q_i) = 1 \quad && f_1(p_i, q_i) = p_i \quad && f_2(p_i, q_i) = q_i \\ &f_3(p_i, q_i) = p_i^2 \quad && f_4(p_i, q_i) = q_i^2 \quad && f_5(p_i, q_i) = p_i q_i \end{aligned} \tag{3}$$

Use the basis functions to define $A \in \mathbb{R}^{100\times 6}$ where $A_i^j = f_j(p_i, q_i)$. Here i is the row index while j is the column index. Implement least squares using the gradient descent algorithm to predict the target variable b by learning the coefficient vector $x \in \mathbb{R}^6$. Report x as well as the Mean Squared Error (MSE).

9. **Piece-wise polynomial least squares problem** Using given data, fit the piece-wise polynomial function on interval $[0, 3]$ with the following constraint. Fit f_1 on interval $[0, 1]$, f_2 on interval $[1, 2]$ and f_3 on interval $[2, 3]$ such that,

   (a) Degree of f_1 is 2
   (b) Degree of f_2 is 3
   (c) Degree of f_3 is 2
   (d) $f_1(1) = f_2(1)$
   (e) $f_2(2) = f_3(2)$
   (f) $f_2'(2) = f_3'(2)$

***************** THE END *****************