CSE343/CSE543/ECE363/ECE563: Machine Learning
Winter 2025

Assignment-2 (30 points)

Release: $28^{\text{th}}$ Jan, 2024 (Tuesday) **Submission: 12:50pm $3^{\text{rd}}$ Feb, 2025 (Monday)**

---

# Instructions

- **Institute Plagiarism Policy Applicable.** Both programming and theoretical questions will be subjected to strict plagiarism check.
- This assignment should be attempted individually. All questions are compulsory.
- **Theory [T]**: For theory questions, only hand-written solutions are acceptable. Attempt each question on a different sheet & staple them together (for the ease of checking). Do not start a new question at the back of the previous one. Do not forget to mention page number (bottom centre) and your credentials (bottom right) on each sheet. It must be submitted in *Assignment submission box* kept infront of **SBILab (B611), R&D Building**. Skipping of any steps in the questions will lead to the deduction of marks.
- **Programming [P]**: For programming questions, the use of python programming language is allowed only. You must submit a single *.ipynb* file named as *A2_RollNo.ipynb*. Make sure the submission is self-complete & replicable i.e., you are able to reproduce your results with the submitted files only. Use random seed wherever applicable to retain reproducability.
- **File Submission**: Submit a *.zip* named A1_RollNo.zip (e.g., *A1_2022112.zip*) file containing the report and code files.
- **Submission Policy**: Turn-in your submission as early as possible to avoid late submissions. In case of multiple sibmissions, the latest submission will be evaluated. **Late submissions will not be evaluated and hence will be awarded zero marks**.
- **Clarifications**: Symbols have their usual meaning. Assume the missing information & mention it in the report. You are allowed to use any machine learning library until exclusively mentioned in the question that it is supposed to be done from scratch. You can always use python libraries such as numpy, pandas, matplotlib, scikit-learn, seaborn, librosa, albumentations, and scipy unless specified otherwise. Use Google Classroom for any queries. In order to keep it fair for all, no email queries will be entertained. You may attend office/TA hours for personal resolutions. Start your assignment early. No queries will be answered in Google Classroom comments during the last time.
- **Compliance**: The questions in this assignment are structured to meet the Course Outcomes CO1, CO2, CO3, and CO4, as described in the course directory.
- There could be multiple ways to approach a question. Please justify your answers mathematically in the theory questions, and via commented text in the programming questions appropriately. Questions without justification will get zero marks.

---

1. **[T ∥ CO1] Linear Algebra** **(2 points)**
For the below-given data matrix $X_{(m\times n)}$, where $m$ is the number of samples and n is the number of features, compute the correlation matrix $A$ and find the eigenvalues and eigenvectors for the computed correlation matrix $A$. Also, prove that the covariance matrix is always positive semi-definite.

$$X_{(4\times 3)} = \begin{pmatrix} 0.25 & 10 & 83 \\ 0.80 & 15 & 57 \\ 0.72 & 53 & 90 \\ 0.95 & 25 & 64 \end{pmatrix} \tag{1}$$

2. **[T ∥ CO1 CO2 CO4] Linear Regression** **( 10 points)**

(a) For the below-given data, fit a linear regression model, find the $\theta_0$, $\theta_1$ and $\theta_2$ and then predict the missing value of the data. (2 points)

| $X_1$ | 2 | 4 | 9 | 12 | 18 | 23 | 26 |
|---|---|---|---|---|---|---|---|
| $X_2$ | 6 | 12 | 28 | 31 | 34 | 30 | 50 |
| $Y$ | 23 | 34 | 25 | 50 | 60 | 100 | ? |

Table 1: Table of $X_1$, $X_2$ (Independent Variable) and Y (Dependent Variable)

(b) Recall that in lectures, we solved the Maximum Likelihood Estimator(MLE) problem for linear regression where the error terms followed the Gaussian Distribution, $\epsilon \sim N(0, \sigma^2)$. Now, solve the MLE problem for following distributions - (4 points)

i. when, $\epsilon \sim Laplace(a, b)$

ii. when, $\epsilon \sim Exponential(\lambda)$

(c) Derive weight update equation for steepest gradient method for linear regression using mean absolute error (MAE) loss function. (2 points)

(d) Let $\mathbf{x}$ be a $d$-dimensional random vector distributed according to some arbitrary distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{S}$. Let $y = \mathbf{w}^T\mathbf{x} + b$. Let us suppose that we have $N$ i.i.d. samples from the probability density function (pdf) $p(y) = Ce^{-(y-\mu_y)^2}$, where $C$ is a scalar constant. Find MLE estimates of $\mathbf{w}$ such that $\mu_y^2 = \sigma_y^2$ is satisfied. $\mathbf{w}$ should be expressed in terms of $N$, $\boldsymbol{\mu}$, $\mathbf{S}$, $b$, and $y$. (2 points)

3. **[T ∥ CO1 CO2] Logistic Regression** **( 2 points)**

(a) Derive the update rule of gradient descent algorithm for logistic regression (i) using the tanh activation function for the binary classification task. (2 point)

4. **[T ∥ CO1 CO2] Ridge Regression** **( 4 point)**

(a) Recall the ridge estimator for $\lambda > 0$ ,

$$\hat{\theta}_\lambda := \arg\min_\theta \|X\theta - y\|^2 + \lambda\|\theta\|^2, \tag{2}$$

Let

$$X = U\Sigma V^T = \sum_i \sigma_i u_i v_i^T \tag{3}$$

i. Show that (2 point)

$$\hat{\theta}_\lambda = \sum_{i=1}^{d} \frac{\sigma_i}{\sigma_i^2 + \lambda} v_i \langle u_i, y\rangle. \tag{4}$$

ii. Show that (2 point)

$$\|\hat{\theta}_\lambda\|^2 = \sum_{i:\sigma_i>0} \left(\frac{\sigma_i}{\sigma_i^2 + \lambda}\right)^2 \langle u_i, y\rangle^2 \tag{5}$$

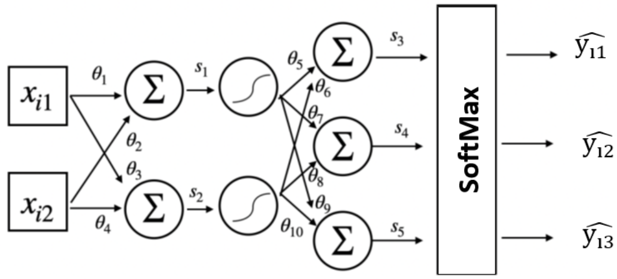


Figure 1: Question 5 $a$

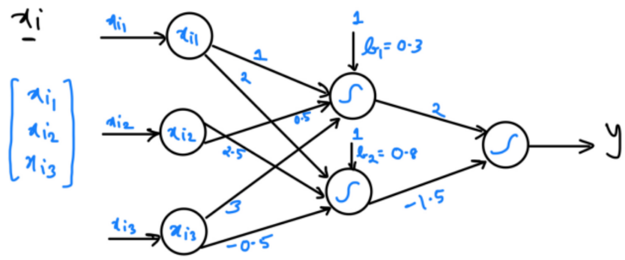


Figure 2: Question 5 $b$

5. **[T ‖ CO1 CO2 CO4] Neural Networks** **( 12 points)**

(a) You are given a Neural Network for a 3-class classification problem in Figure 1. You have to compute the weight update rules for $\theta_1$, $\theta_3$ and $\theta_5$ in terms of $x_{ij}$ (Input), $\hat{y_{ij}}$ (Predicted Probability), $y_{ij}$ (Groundtruth Label). In the hidden layers, the activation function used is sigmoid. You can choose the loss function based on your assumption. (4 points)

(b) Given a neural network in Figure 2 for binary classification. Consider a batch size of 2, learning rate of 0.01 for all the learnable parameters. Given two input samples, $\mathbf{x}_1 = [0.6, 0.7, 1.3]^T$ belonging to class 1 and $\mathbf{x}_2 = [-1.8, 0.1, 0.2]^T$ belonging to class 0. The initial weights of the network are shown in the Figure 2.

i. Forward propagate the samples $\mathbf{x}_1$ and $\mathbf{x}_2$ to the output and compute the value of the output at each node of the network. ( 2 points)

ii. Compute Mean Squared Error (MSE) and the average Binary Cross Entropy (BCE) loss. Write an inference by comparing the two losses. ( 2 points)

iii. Back-propagate the gradient and update all the learnable parameters. ( 4 points)