Indian Institute of Technology Jodhpur
CSL2010: Introduction to Machine Learning
Lab 3&4, Due Date: Sep 7, 2025, Max Marks: 70+30 for Viva

1. (35 points) Consider the attached dataset $\{\mathbf{x}_i, y_i\}_{i=1}^{n}$ where each data point $\mathbf{x}_i \in \mathbb{R}^8$ has eight features and the corresponding label $y_i \in \mathbb{R}$ is a continues variable. Partition the dataset into training (70%), validation (15%) and testing (15%) sets. Assume that T, V, and O represent the training set, validation set and test set, respectively. Let meachine learning model be defined as $h_\theta(\mathbf{x}_i) = \theta_0 + x_{i1}\theta_1 + x_{i2}\theta_2 + \cdots + x_{i8}\theta_8$ to predict the labels $\hat{y}_i = h_\theta(\mathbf{x}_i)$ where $\theta = \begin{bmatrix}\theta_0 & \theta_1 & \cdots & \theta_8\end{bmatrix}^\top$ represents the learnable parameters of the machine learning model h_θ. Let $\ell(\theta) = \frac{1}{|T|}\sum_{i \in T} (h_\theta(\mathbf{x}_i) - y_i)^2$ represents the loss function to find the optimal learnable parameters θ by minimizing it, i.e., $\theta^* = \arg\min_\theta \ell(\theta)$. Now, do the following.

   (a) Implement the closed form solution that we have discussed in the class to find the optimal learn able parameters. Let us call these parameters as θ^*_1. Use the training dataset T to find the optimal parameters. Find the training error $E_T = \frac{1}{|T|}\sum_{i \in T}\left(h_{\theta^*_1}(\mathbf{x}_i) - y_i\right)^2$, validation error $E_V = \frac{1}{|V|}\sum_{i \in V}\left(h_{\theta^*_1}(\mathbf{x}_i) - y_i\right)^2$, and the test error $E_O = \frac{1}{|O|}\sum_{i \in O}\left(h_{\theta^*_1}(\mathbf{x}_i) - y_i\right)^2$.

   (b) Implement the gradient descent algorithm to find the optimal learnable parameters. Let us call these parameters as θ^*_2. Use the training dataset T to find the optimal parameters. Find the training error E_T and validation error E_V at each iteration of the gradient descent algorithm and plot these errors as a function of the iterations. Fine-tune the learning rate that gives optimal performance. Compare the training, validation, and test errors at the end of the convergence of the gradient descent algorithm with those obtained in Part (a). Also, plot the magnitude of the gradient of the loss function as a function of the iterations of the gradient descent algorithm.

2. (35 points) Consider the attached dataset $\{x_i, y_i\}^n_{i=1}$ where each data point $x_i \in R^2$ has features and the corresponding label $y_i \in \{0, 1\}$ is a binary variable. Partition the dataset into training (70%), validation (15%) and testing (15%) sets. Assume that T , V, and O represent the training set, validation set and test set, respectively. Let the binary class classification model be defined as $h_\theta(x_i) = \frac{1}{1+e^{-\theta_0 - x_{i1}\theta_1 - x_{i2}\theta_2}}$ to predict the labels $\hat{y}_i = h_\theta(x_i)$ where $\theta = \begin{bmatrix}\theta_0 & \theta_1 & \theta_2\end{bmatrix}^\top$ represents the learnable parameters of the machine learning model h_θ. Let $\ell(\theta) = \frac{1}{|T|}\sum_{i\in T} y_i \log(h_\theta(x_i)) + (1 - y_i) \log(1 - h_\theta(x_i))$ represents the loss functions to find the optimal learnable parameters θ by maximizing it, i.e., $\theta^* = \arg\max_\theta \ell(\theta)$. Now, do the following.

   (a) Implement the gradient descent algorithm to find the optimal learnable parameters. Use the training dataset T to find the optimal parameters. Find the training and validation accuracy and error at each iteration of the gradient descent algorithm and plot these as a function of the iterations.

   (b) Find confusion matrix, precision, recall and F-1 score on the training, testing and validation datasets and plot precision vs. recall curve. Vary the probability threshold from 0 to 1 with a step size of 0.05.

   (c) Plot the fitted linear boundary $\theta^*_0 + x_1\theta^*_1 + x_2\theta^*_2 = 0$ that separates the two classes. Also overlay the test, train and validation data points (with different colors) on this plot.