LUMS
A Not-for-Profit University

Department of Electrical Engineering
School of Science and Engineering

EE514/CS535 Machine Learning

ASSIGNMENT 2

Due Date: 11 am, Thursday, April 10, 2025.
Format: 6 problems, for a total of 100 marks
Instructions:

- You are allowed to collaborate with your peers but copying your colleague's solution is strictly prohibited. This is not a group assignment. Each student must submit his/her own assignment.
- Solve the assignment on blank A4 sheets and staple them before submitting.
- Submit in-class or in the dropbox labeled EE-514 outside the instructor's office.
- Write your name and roll no. on the first page.
- Feel free to contact the instructor or the teaching assistants if you have any concerns.

- You represent the most competent individuals in the country, do not let plagiarism come in between your learning. In case any instance of plagiarism is detected, the disciplinary case will be dealt with according to the university's rules and regulations.

Problem 1 (15 marks)

The objective of logistic regression is to find a decision boundary that separates two or more distinct classes. Here, we focus on binary classification. Given data features $x \in \mathbb{R}^{d+1}$ (with a 1 appended in the first position for the intercept term) and corresponding labels $y \in \{0, 1\}$, we aim to learn a parameter vector $\theta \in \mathbb{R}^{d+1}$ such that:

$$z = \theta^T x,$$

$$\sigma(z) = \frac{1}{1 + e^{-z}},$$

where $\sigma(z)$ represents the predicted probability that $y = 1$. The goal is to ensure that $\sigma(z)$ is close to 1 when $y = 1$ and close to 0 when $y = 0$.

To achieve this, we optimize the binary cross-entropy loss function with Elastic-Net regularization:

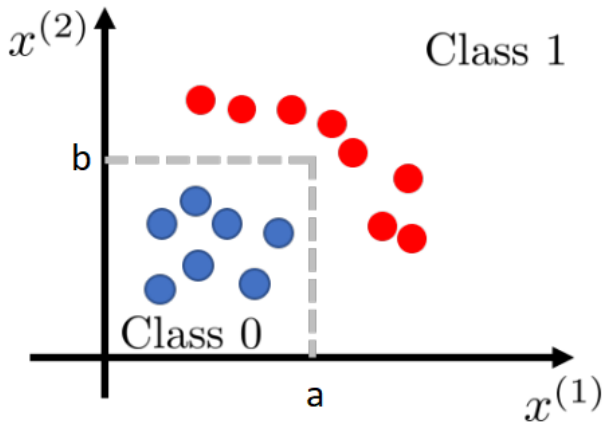
$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(\sigma(z_i)) + (1 - y_i) \log(1 - \sigma(z_i)) \right] + \lambda_1 \|\theta\|_1 + \frac{\lambda_2}{2} \|\theta\|_2^2,$$

where N is the total number of samples, and λ_1 and λ_2 are regularization parameters.

Since the minimizer of the cross-entropy loss with Elastic-Net regularization does not have a closed-form solution, it must be optimized iteratively using gradient-based methods. Derive the gradient of the regularized loss function with respect to θ.

Problem 2 (15 marks)

Design a classifier comprised of two layers of perceptron with the dashed line indicated in the figure as its (approximate) decision boundary. You must draw a perceptron indication of inputs, output, weights, and biases and provide a brief explanation of your design. Hint: You might want to start with a perceptron that determines on which side of the dashed boundary the data lies.



Problem 3 (20 marks)

You have been given a small dataset of customer reviews for different products. The reviews are classified into three categories: positive, negative, or neutral.

Dataset	Sentiment	Text
Training	Positive	excellent service and fast delivery
Training	Positive	great product quality and packaging
Training	Neutral	average experience but acceptable quality
Training	Negative	poor customer support and delayed shipping
Training	Negative	terrible product with missing parts
Testing	?	excellent product with quick delivery

You need to develop a **multinomial Naive Bayes classifier** for this problem by following the steps below.

(a) [**2 marks**] A list of stop words is given to you.
Stop words = [as, of, at, by, and, the, an, but, is, with]

Apply preprocessing to the training and test data by removing stop words and showing the documents after preprocessing.

(b) [**3 marks**] You need to work with the preprocessed documents from now onwards. Construct vocabulary from the data and state its size.

(c) [**5 marks**] Construct prior probabilities based on the training data.

(d) [**6 marks**] Compute the likelihoods of all the words in the training data using Laplace add-one smoothing.

(e) [**4 marks**] Predict the sentiment of the test data and show your working.

Problem 4 (15 marks)

Consider a network of n sensors measuring the temperature Y. The sensor measurements, denoted by X_i, are corrupted by noise, with the measurement model given as:

$$X_i = Y + Z_i, \quad i = 1, \ldots, n,$$

where $Z_1, \ldots, Z_n$ are independent and non-identically distributed noise terms, with $Z_i \sim \mathcal{N}(0, \sigma_i^2)$. These noise terms are statistically independent of Y. Moreover, it is known that the true temperature follows a Gaussian distribution $Y \sim \mathcal{N}(0, \sigma_y^2)$.

(a) Determine the ML (Maximum Likelihood) estimate of Y.

(b) Determine the MAP (Maximum A Posteriori) estimate of Y.

Problem 5 (20 marks)

Consider a binary classification problem with two inputs and the following labeled dataset for training.

Label y	**Data Point** $(x^{(1)}, x^{(2)})$
1	(-2, -2)
1	(-2, 2)
1	(2, 2)
0	(1, 1)
0	(1, -1)
0	(-1, 1)

Table 2: Data points

(a) [**2 marks**] Plot the points on a 2D plane. Can we use hard SVM for this problem? Provide a brief justification to support your answer.

(b) [**3 marks**] Since the data is not linearly separable, we map the 2D feature space to 3D feature space using the mapping function $\phi(x)$ to make it linearly separable. Determine the mapping function that can enable us to use hard SVM in 3D feature space.

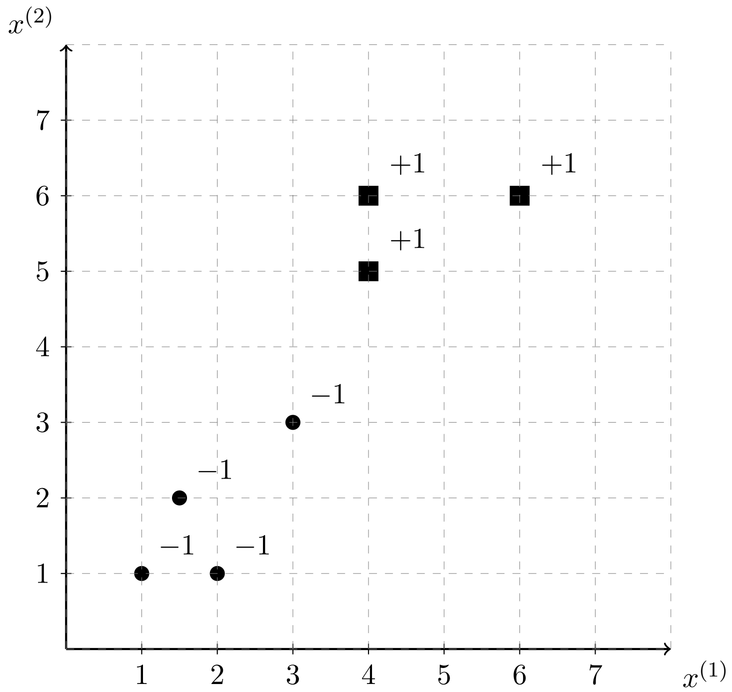
(c) [**2 marks**] We have a linear decision boundary (hard SVM) in 3D space to separate the transformed data in 3D (new feature space). Indicate this boundary as a (non-linear) decision boundary on the plot obtained in part (a).

(d) [**3 marks**] Instead of mapping the data into 3D space and using hard SVM to learn the decision boundary in 3D, we can use the kernel trick to learn a non-linear boundary you have plotted in part(c) in the original 2D feature space. Formulate a kernel function associated with the mapping function you used in part (b).

(e) [**10 marks**] Determine the predicted probabilities for two test points, $(0, 1)$ and $(2, -1)$, using Kernel Logistic Regression. Initialize all weights $\alpha = 0.1$ and perform one iteration of gradient descent to update α. Use the cross-entropy loss function to compute the gradient and update the weights accordingly, where learning rate $\eta = 0.01$. Show your working on paper.

Problem 6 (15 marks)

Consider the following 2D dataset with two classes, where each point is labeled either as $+1$ or -1. Find the weight vector and bias for the decision boundary $\mathbf{w}^T\mathbf{x} - \theta = 0$ maximizing the classification margin. Also, compute the classification margin.



— End of Assignment —