CEE 554: Machine Learning for Infrastructure Systems

Homework Assignment 1
Due date: 01/21, 11:59pm on Canvas

- You may either type your answers or handwrite them and submit a scanned copy. If you choose to handwrite your solutions, your writing must be clearly legible and the scan quality must be sufficient for grading. Illegible submissions will receive no credit.
- Include an *Appendix A* at the end of the assignment stating whether you used any Generative AI tools in developing your solutions. If so, specify the tool used and the prompt or commands used for each problem.
- Include an *Appendix B* at the end of the assignment containing all code used to solve each problem.

1. [Perceptron Learning Algorithm (PLA) on Bridge Condition Data] In this problem, you will implement the Perceptron Learning Algorithm (PLA) and the pocket algorithm on an infrastructure-inspired dataset derived from bridge inspection condition ratings.

   Dataset. Use the file `Bridge_Condition.txt`. Each record has the form

   $$(x_1,\ x_2,\ y),$$

   where x_1 is the *deck condition rating*, x_2 is the *superstructure condition rating*, and $y \in \{-1, +1\}$ is the class label. The first 20 records in the file form **Dataset 1**, and the full 40 records in the data set form **Dataset 2**.

   (a) **Visualize Dataset 1.**
   Create a scatter plot in the (x_1, x_2) plane using different markers/colors for $y = +1$ and $y = -1$. Is Dataset 1 appears linearly separable or not?

   (b) **Run PLA on Dataset 1.**
   Initialize $\mathbf{w}^{(0)} = \mathbf{0}$. Run PLA on Dataset 1. Process points sequentially in the order they appear in the file, looping over the dataset until convergence.

   i. Report the number of updates until convergence.
   ii. Plot the final separating line (decision boundary) on top of the scatter plot.
   iii. Plot the *in-sample misclassification rate* per iteration.

   (c) **Effect of shuffling on Dataset 1.**
   Randomly shuffle Dataset 1 and rerun PLA from $\mathbf{w}^{(0)} = \mathbf{0}$. Repeat this experiment for 10 different random shuffles.

   i. Report the number of updates to convergence for each run.
   ii. In 2–4 sentences, comment on why the number of updates depends on the order of the data points.

   (d) **Visualize dataset 2.**
   Now load the *full file* (all records). Create a scatter plot in the (x_1, x_2) plane, again using different markers/colors for $y = +1$ and $y = -1$. Based on your plot, explain whether the dataset is or is not linearly separable.

   (e) **Run PLA on the full dataset with a stopping rule.**

   i. Shuffle the full dataset once.
   ii. Run PLA starting from $\mathbf{w}^{(0)} = \mathbf{0}$ for a maximum of T updates (choose T large enough to observe behavior, e.g., $T = 1000$).
   iii. Plot the misclassification rate versus the number of iterations.

iv. Implement the pocket algorithm, which keeps the best-performing weight vector seen so far.

A. Report the best (lowest) misclassification rate achieved and the corresponding $\mathbf{w}_{\text{best}}$.

B. Plot the decision boundary for $\mathbf{w}^{(T)}$ and for $\mathbf{w}_{\text{best}}$ on the same figure (together with the scatter plot).

C. In a few sentences, compare PLA and pocket PLA in this non-separable setting.

2. [Logistic-MSE Classifier] In this problem, you will implement the linear classifier with a logistic-MSE cost function (with a linear signal) as discussed in class.

Let
$$\sigma(z) = \frac{1}{1+e^{-z}}, \qquad \rho(z) = 2\sigma(z) - 1,$$
and consider the cost function
$$C(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N}\Big(\rho(\mathbf{w}^\top \mathbf{x}_i) - y_i\Big)^2.$$

(a) **Derive the SGD update rule.**
Derive the stochastic gradient descent update for $\mathbf{w}$ using one data point $(\mathbf{x}_i, y_i)$ at a time.

(b) **Apply the logistic-MSE classifier to Dataset 1.**

i. Draw the final classifier and the training data points in one figure.

ii. Report the changes in the cost function and misclassification error rate per iteration.

iii. Show the final PLA classifier from Problem 1 and the logistic-MSE classifier side by side, and draw comparisons between their performances.

(c) **Apply the logistic-MSE classifier to Dataset 2.**

i. Draw the final classifier and the training data points in one figure.

ii. Report the changes in the cost function and error rate per iteration.

iii. Show the pocket PLA classifier from Problem 1 and the logistic-MSE classifier side by side, and draw comparisons between their performances.

Implementation notes.
You might have to play with parameters in your code, such as the max number of iterations, tolerance, and step size. Additionally, this method only provides a local minimum. Therefore, run the code multiple times, each time starting with a random weight vector $\mathbf{w}$, to obtain an acceptable classifier.