CSE 151B/251B Deep Learning
UC San Diego, Spring 2025

Homework 1

Name: Yosub Lim UID: A18657974

1 Supervised Learning

Problem A: Feature Representation

Solution A:

Each document is represented as a binary BoW feature vector, using a predefined dictionary of five keywords. Each position in the vector corresponds to whether that word appears in the document. A value of 1 means the word is present in the document, a value of 0 means the word is absent.

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Problem B: Logistic Regression

Solution B:

$$\frac{\partial L}{\partial w_j} = -\sum_{i=1}^{N} \left[y^{(i)} \cdot \frac{1}{f(x^{(i)})} \cdot \frac{\partial f(x^{(i)})}{\partial w_j} + (1 - y^{(i)}) \cdot \frac{1}{1 - f(x^{(i)})} \cdot \left(-\frac{\partial f(x^{(i)})}{\partial w_j} \right) \right]$$

$$= \sum_{i=1}^{N} \left[\left(f(x^{(i)}) - y^{(i)} \right) \cdot \frac{\partial w^\top x^{(i)}}{\partial w_j} \right]$$

$$= \sum_{i=1}^{N} \left(f(x^{(i)}) - y^{(i)}\right) x_j^{(i)}$$

2 Multi-Layer Perceptron

Problem A: Perceptron

i. Implementation of Perceptron

Solution A.i:

Implemented in notebook.

ii. Linear Separability

Solution A.ii:

In 2D cases, the smallest dataset that is not linearly separable contains 4 points. If three points form a triangle and are labeled as one class, and the fourth point lies inside the triangle and is labeled as the opposite class, no straight line an separate the two classes. This makes the dataset not linearly separable.

In 3D case, the smallest dataset that is not linearly separable contains 5 points. For example, four points forming the vertices of a tetrahedron and labeled as one class, with a fifth point placed inside the tetrahedron and labeled as the opposite class, results in a configuration that cannot be separated by any plane.

The smallest non-linearly separable dataset in N-dimensional space is of size is N + 2.

iii. Non-linearly Separable Dataset

Solution A.iii:

If a dataset is not linearly separable, the Perceptron Learning Algorithm will not converge. The Perceptron Learning Algorithm updates the weight vector whenever it encounters a misclassified point. If no linear boundary exists that can perfectly separate the data, there will always be at least one misclassified point in every iteration. As a result, the algorithm will continue to update the weights indefinitely, without ever reaching a solution that classifies all points correctly.

Problem B: Multilayer Perceptron

i. MNIST Classification Implementation

Solution B.i:

Implemented in notebook.

ii. Parameter Counts

Solution B.ii:

From input layer to hidden layer, there are 784 *inputs* $\times$ 500 *hidden units* = $392{,}000$ *weights, plus* 500 *biases. From the hidden layer to output layer, there are* 500 *hidden units* $\times$ 10 *outputs* = $5{,}000$ *weights, plus* 10 *biases. In total, there are 397,510 parameters.*

Compared to logistic regression, which only has 784 $\times 10 + 10 = 7{,}850$ *parameters,*

MLP has significantly more due to the additional hidden layer.