Report: Al-Phabet (WIDS Project)

Week 1: Essential Libraries, Data Distribution, and Numpy

Essential Libraries: Numpy and Matplotlib

Introduction:

In Week 1, we delved into foundational concepts crucial for machine learning. We started by exploring essential libraries, specifically Numpy and Matplotlib, laying the groundwork for subsequent topics.

Data Visualization using Matplotlib:

Understood data relationships and identified features using Matplotlib and seaborn libraries by plotting to visually explore and analyze data, aiding in model building.

Data Distribution:

Bernoulli, Uniform, and Normal distributions are discussed, providing insights into their applications in machine learning.

Week 2: Linear Regression and Logistic Regression

Linear Regression

Overview:

Week 2 initiates exploration into one of the fundamental concepts in machine learning—Linear Regression. The simple and multiple linear regression models are introduced, emphasizing the relationship between input variables and the target variable.

Logistic Regression

Binary Logistic Regression:

Logistic Regression is introduced as a classifier that predicts the probability of input data belonging to discrete classes. Binary logistic regression is highlighted, emphasizing the sigmoid function as the activation function.

Multinomial Logistic Regression:

Multinomial Logistic Regression, also known as Softmax Regression, is discussed. The Softmax activation function and its role in classifying input vectors into different classes are explained.

Generalized Linear Models:

On a higher level, Generalized Linear Models (GLM) are introduced, encompassing linear and logistic regression. GLMs serve as a generalization of linear models, considering various data distributions.

Week 3: Feedforward Neural Networks

FeedForward Neural Networks

Introduction:

Delved into feedforward neural networks, the simplest artificial neural network. The information flow, forward propagation, and the role of activation functions are discussed.

Activation Functions:

Different activation functions, such as sigmoid, tanh, and ReLU, are explored. The necessity of non-linear activation functions and their advantages are discussed.

Optimization Problem:

Training neural networks involves solving an optimization problem. Gradient descent is introduced as the optimization algorithm, and the concept of loss functions is discussed.

- [Loss Function and Maximum Likelihood]

- [Cross Entropy and Mean Squared Loss]

Back Propagation:

Back propagation, using the loss calculated by the loss function to update model parameters, is explained. The iterative process of forward and backward propagation is crucial for model training.

Multi-class Classification using NN:

Building on the concepts learned, the implementation of a neural network for multi-class classification using the MNIST dataset is explored. The transition to the EMNIST dataset is introduced, highlighting the similarities and necessary adaptations.

Week 4: Building a Classifier for MNIST and EMNIST

Introduction:

In Week 4 of our project, we embarked on the practical application of our knowledge by building a classifier for the MNIST dataset from scratch using only the NumPy library. This week's focus was on leveraging our understanding of neural networks and extending it to handwritten alphabets using the Extended MNIST (EMNIST) dataset.

Implementation Simplicity:

Due to the structural similarities between the MNIST and EMNIST datasets, transitioning from our MNIST model to an EMNIST model is expected to be a straightforward process. Minor adjustments are anticipated, and identified these changes as part of the learning experience.

Accuracy Score (MNIST): 90.2%

Accuracy Score (EMNIST): 88.9%

Conclusion:

Week 4 marks the practical application of our machine learning knowledge as we construct a classifier for both the MNIST and EMNIST datasets using the NumPy library. This hands-on approach enhances our understanding of neural networks and their adaptability to diverse datasets, setting the stage for more advanced topics in the subsequent weeks of the project.