Experimental Setup

1 Experimental Setup

In order to examine the effectiveness of the proposed *unsupervised bin wise pre-training* model in terms of both as a Optimizer and Regularizer, three benchmark datasets were used: (i)MNIST handwritten digits dataset (ii) Zalando's article images dataset (Fashion MNIST) and (iii) University of Bern EEG dataset.

Evaluating on different characteristic datasets would help to estimate the benefit gained and ubiquitous nature of the proposed work. To make the experimentation process simple and efficient, models were designed with minimal number of architectural elements to enforce parsimony[].For all three datasets, Cross entropy loss was employed as the objective function, Rectified Linear Units (ReLUs) was set as activation function and Adam optimizer was applied. The models were implemented in Python3 using PyTorch library. In addition, NumPy and Matplotlib were used to perform matrix operations and plotting respectively.All the experimentation were carried out in the machine with 6 GB of Primary Memory, Intel i3 dual core CPU with 1.7 GHz and 64-bit Debian operating system.

1.1 MNIST dataset

The MNIST dataset consists of handwritten digits, which are widely preferred by the researchers to evaluate their proposed models. The dataset comprises of 70,000 handwritten digit images, out of which 60,000 and 10,000 images were used for training (TR_{D1}) and testing (TE_{D1}) respectively. The training dataset (TR_{D1}) was sub-divided into training dataset (T_{D1}) and validation dataset (V_{D1}) in the ratio 80:20 (48,000:12,000). DNN was configured with an input layer consisting of 784 (28 x 28 pixels) input neurons, five hidden layers comprising 1024, 200, 20, 20, 20 neurons in each layer respectively and an output layer with 10 output neurons which corresponds the class labels of the MNIST dataset.

1.2 Zalando's article images dataset

Zalando's article images dataset consists of 60,000 training and 10,000 testing examples, with each example represents a $28x28$ grayscale image associated with a class from 10 different articles of clothing classes. Due to its high similarity with original MNIST dataset, it is popularly known as Fashion MNIST. As with MNIST dataset, the TR_{D2} was divided into T_{D2} and V_{D2} in 80:20 ratio, and DNN model was configured with hidden layers of dimension 1024, 200, 20, 20, 20 and output layer of size 10.

1.3 University of Bern EEG dataset

University of Bern-Barcelona EEG dataset comprises of EEG recordings from five epileptic patients using intracranial method[$CitationReq$]. During data acquisition the sampling rate was set either 512 or 1024 Hz and later down-sampled to 512Hz. Dataset consists of 7500 samples with equally balanced focal and non-focal types which in turn each sample consists of 10,240 data points. The reason behind for choosing a dataset of physiological signals for evaluating the proposed work is because of their extremely complex and vulnerable to noise behavior. The dominance of DNNs over

interpreting physiological signals is still not as expected and a much amount of research is devoted in this field [https://www.sciencedirect.com/science/article/pii/S0169260718301226]. As a common approach of EEG preprocessing, the signals were decomposed using Daubechies10 basis wavelet function [Scopus paper citing] and entropy features (comprising of Sample, Shannon, Permutation and Multiscale entropies) were extracted. The datset was splitted into T_{D3}, V_{D3} and TE_{D3} in 60:30:10 ratio, and DNN architecture was designed with hidden layers of size 200, 150, 100, 20 and the output layer with size 2.