

CS 547 / IE 534 Deep Learning, Fall 2019

Homework 2

Author: Vardhan Dongre (vdongre2)

Implemented and trained a CNN from scratch in Python for the MNIST dataset. The neural network was trained on the Training Set using stochastic gradient descent.

This implementation achieved an accuracy of **94.12%** on the Test Set.

Implementation:

The code can be understood in the following subheads:

1. **Data:** MNIST database is a database of handwritten digits which contains 60,000 training images and 10,000 testing images. The images are pre-processed to fit in 28X28 pixel box. In this code the training images are stored in `x_train` and `x_test` and the corresponding labels are in `y_train` and `y_test` respectively.
2. **Model:** The model/Learning Algorithm adopted is a Convolutional Neural Network. The network consists convolution layer with multiple channels. In this model the Filter/Kernel is 3X3 and the number of kernels (Channels) is 5. The weights and biases for output layer are randomly initialized, normalized and stored in a dictionary `{C, b}`. The ReLU function is used which adds a layer of non-linearity. Forward propagation is performed on randomly sampled data points in training set. The convolution operation extracts feature maps from the image (28X28) (No padding was used) using a Filter (3X3, 5 Channels). Backward propagation is implemented to calculate the gradients of parameters which are required for computing the step that is used to optimize the model. Stochastic Gradient Descent has been implemented for optimization where we are using single data samples for computing the update direction
3. **Training:** This model is trained for 10 epochs using a Learning rate schedule with 0.01 as the beginning learning rate. The Learning rate is defined using a piecewise learning rate schedule which decreases from 0.01 to 0.00001. Predictions of the model are compared with the actual labels for the training data and the accuracy of the model is calculated. It was observed that with each iteration the accuracy of the prediction increased, and model achieved an accuracy of 96% on the training set after 8 epochs.
4. **Testing:** The model is tested on the Test set and an accuracy of 94.12% is obtained. (Objective was to achieve an accuracy of at-least 94 % on the Test set)