EE6132 : Advanced Topics in Signal Processing

Programming Assignment 1: MNIST Classification using Multilayer Neural Networks

Kranthi Kumar ee18d004@smail.iitm.ac.in

Due: Sep 12th, Wednesday, 11:55pm

September 11, 2018

Instructions

1. You may do the assignment in MATLAB, Python.
2. For any questions, please schedule a time with TAs before deadline according to their convenience. Please use moodle discussion threads for posting your doubts and also check it before mailing to TAs, if the same question has been asked earlier.
3. Submit a single zip file in the moodle named as *PA1_Rollno.zip* containing report and folders containing corresponding codes.
4. Read the problem fully to understand the whole procedure.
5. Do not copy any part from any source including your friends, seniors or the internet.
6. Late submissions will be evaluated for reduced marks and for each day after the deadline we will reduce the weightage by 10%.

The aim of this assignment is to experiment with Multilayer Feedforward Neural Network (MLP) with Backpropagation (BP) learning. Please write your own code for BP algorithm and MLP training and testing. Do not use any libraries or copy directly any code from any resource coding BP gives you greater insight into the practical issues.

Dataset : The aim is to code a complete handwritten digit recognizer and test it on the MNIST digit dataset. (Download the dataset and its description from: http://yann.lecun.com/exdb/mnist/). The MNIST database of handwritten digits, available from this page, has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image.

MLP Architecture and Training : Input to the network is a digit image (28×28) flattened as 784 dimensional vector x_i. Input is followed by layers $h_1(1000)$, $h_2(500)$, $h_3(250)$. The number of units are mentioned beside the corresponding hidden layer. Use sigmoid activation for all these hidden layers. The output unit consists of 10 units for each digit. Let the output activation be linear. Since you want to do classification, use softmax to get probabilities of digit belonging to 10 classes. This will be a 10 dimensional vector $\hat{y}_i$. Now use cross entropy loss between $\hat{y}_i$ and y_i, where y_i is the one hot representation of ground truth label.

Use a minibatch size of 64 while training. Run the training for 8000 iterations. For parameter initialization using random gaussian with standard deviation close to zero (say 0.08). Use *SGD* for training with momentum based acceleration.

Submissions:

1. Use $5-fold$ cross validation[1] for your experiments. You need to submit your training progress as a plot showing train loss and test loss convergence over iterations *for any one fold*. Test loss need not be calculated for every iteration, instead you can calculate it for every 200 iterations or so. Observe the convergence with varying learning rates.

2. Report the confusion matrix that shows the kind of errors that your classifier makes. In this problem, your confusion matrix is a 10×10 matrix, where the rows represent the true label of a test sample and the columns represent the predicted labels of the classifier. Report the average error rate as well as standard deviation of the error rate *for each fold* along with other metrics such as Precision, Recall/Sensitivity, and $F1-score$.

3. Now change the activation function to *relu* instead of *sigmoid* and use $5-fold$ cross validation for your experiments. Submit the training plot *for any one fold*. Report the average error rate as well as standard deviation of the error rate *for each fold* along with other metrics such as Precision, Recall/Sensitivity, and $F1-score$. Compare the results with the sigmoid activation case.

4. Try one variation of data-augmentation using *de-skewing*, *adding noise*, etc. and also try one variation with other regularization terms such as L_1-*Regularization*, L_2-*regularization/weight-decay*, *tangentprop*, etc. and compare your results from (2 and 3) above.

5. In all the cases after training, choose a fixed set of 20 images from test set and show the network's top 3 guesses for each of them.

6. Instead of using the flattened image as input, use any of the off-the-shelf feature extractor[2] to get the feature representation for images. Due credit will be given for choosing non-trivial feature extraction.

   Choose your own architecture (number of hidden layers, number of nodes in each layer, activations, loss function, regularization etc).

   Your recognizer needs to read the image data, extract features from it and use a multilayer feedforward neural network classifier to recognize any test image.

   Repeat steps 2 & 3 and comment on the results.

7. Compare the results from (6) with classification using *k-NN* (Euclidean distance based) and *SVM*[3] classifier. Present an analysis and discussion of your results.

---

[1] Refer *Section 5.3.1 Cross-Validation* in Deep Learning Book

[2] SIFT, SURF, HOG, MSER etc. Open-source implementations are available online for all of these. For example, If you are using MATLAB, some of these are availble as in-built functions or you can use VLFeat library. If you are using Python, you can use OpenCV.

[3] Python - use LIBSVM or scikit; MATLAB - use VLFeat