EAS 595: Fundamentals of Artificial Intelligence (Spring 2020)

Dr. David Doerman, Mihir Chauhan
University at Buffalo, The State University of New York
Buffalo, New York 14260
Contact: mihirhem@buffalo.edu

April 20, 2020

1 Task

The task of this homework is to implement convolutional neural network for the task of classification.

The classification task will be that of recognizing an image and identify it as one of ten classes. You are required to train the classifiers using MNIST handwritten digits and USPS image data. Following are the three tasks to be performed using Keras library on Python:

1. Build a CNN model A which is trained on USPS train dataset and test CNN model A with USPS test dataset.
2. Build a CNN model B which is trained on MNIST train dataset and test CNN model B with USPS test dataset.
3. Build a CNN model C which is trained on MNIST train and USPS train dataset and test CNN model C with USPS test dataset.

You can use any CNN architecture for building Model A, B and C. Compare and comment on the test results obtained by CNN Model A, B and C as shown in the evaluation section. Deadline for the homework is 11.59PM on April 29th.

2 Datasets

2.1 MNIST Handwritten Digits Dataset

The MNIST database of handwritten digits has 70,000 images of handwrtitten digits. 60,000 training samples and 10,000 test samples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image of size 28x28.

2.2 USPS Dataset

The USPS database of handwritten digits has 21,500 images of handwrtitten digits. 20,000 training samples and 1500 test samples.

Download USPS dataset from here: `https://drive.google.com/file/d/1w6zxPR-uJiR0LVPwACvE17NHnx_eIFgx/view?usp=sharing`

3 Plan of Work

1. **Extract feature values and labels from the data:** MNIST and USPS dataset is downloaded and processed into a Numpy array that contains the feature vectors and a Numpy array that contains the labels and the training data.

2. **Pre-process USPS image data:** Reshape USPS dataset to fixed size 28x28 images.

3. **Data Partitioning:** The MNIST and USPS datasets are originally partitioned into training set and testing set. You will use this partition and train your model on the training set in respective Models.

4. **Model A Training** Using a group of hyperparameters for CNN train the model A on USPS train dataset using Keras.

5. **Model B Training** Using a group of hyperparameters for CNN train the model B on MNIST train dataset using Keras.

6. **Model C Training** Using a group of hyperparameters for CNN train the model C on MNIST train and USPS train dataset using Keras.

7. **Tune hyper-parameters:** For steps 3, 4 and 5: Change your hyper-parameters and repeat the step. Try to find what values those hyperparameters should take so as to give better performance on the testing set.

8. **Test Model A, B and C on the testing set:** For steps 3, 4 and 5: After tuning the hyper-parameters, fix your hyper-parameters and model parameter and test model A, B and C performance on USPS test dataset.

4 Evaluation

1. For each Model evaluate solution on the test set using classification accuracy:

$$Accuracy = \frac{N_{correct}}{N} \tag{1}$$

   Where where $N_{correct}$ is the number of corrected classified data samples, and N is the total number of samples of the validation set.

2. Construct a confusion matrix for each Model and observe the relative strengths and weaknesses.

5 Deliverable

You only need to submit the code. After finishing the homework, you may be asked to demonstrate it to the graders, particularly if your results and reasoning in your code are not clear enough. The code should be in Python only. Please provide necessary comments in the code. The name of the python notebook should be main.ipynb. Also we request all of you to kindly convert the .ipynb file to .py and include main.py file as well.

Submit the Python code and report on UBlearns as well as Autolab as a zip file named hw3.zip
Autolab link: https://autograder.cse.buffalo.edu/