Neural **Networks**

TOTAL **SCORE** 60

Part A - 30 Marks

- **DOMAIN:** Electronics and Telecommunication
- **CONTEXT:** A communications equipment manufacturing company has a product which is responsible for emitting informative signals. Company wants to build a machine learning model which can help the company to predict the equipment's signal quality using various parameters.
- **DATA DESCRIPTION:** The data set contains information on various signal tests performed:
  1. Parameters: Various measurable signal parameters.
  2. Signal_Quality: Final signal strength or quality
- **PROJECT OBJECTIVE:** To build a classifier which can use the given parameters to determine the signal strength or quality.

Steps and tasks: **[Total Score: 30 Marks]**

1. Data import and Understanding [10 Marks]
   A. Read the 'Signals.csv' as DataFrame and import required libraries. [2 Marks]
   B. Check for missing values and print percentage for each attribute. [2 Marks]
   C. Check for presence of duplicate records in the dataset and impute with appropriate method. [2 Marks]
   D. Visualise distribution of the target variable. [2 Marks]
   E. Share insights from the initial data analysis (at least 2). [2 Marks]

2. Data preprocessing [7 Marks]
   A. Split the data into X & Y. [1 Marks]
   B. Split the data into train & test with 70:30 proportion.[1 Marks]
   C. Print shape of all the 4 variables and verify if train and test data is in sync. [1 Marks]
   D. Normalise the train and test data with appropriate method. [2 Marks]
   E. Transform Labels into format acceptable by Neural Network [2 Marks]

3. Model Training & Evaluation using Neural Network [13 Marks]
   A. Design a Neural Network to train a classifier. [3 Marks]
   B. Train the classifier using previously designed Architecture [2 Marks]
   C. Plot 2 separate visuals. [3 Marks]
      i. Training Loss and Validation Loss
      ii. Training Accuracy and Validation Accuracy
   D. Design new architecture/update existing architecture in attempt to improve the performance of the model. [2 Marks]
   E. Plot visuals as in Q3.C and share insights about difference observed in both the models. [3 Marks]

Part B - 30 Marks

- **DOMAIN:** Autonomous Vehicles

- **CONTEXT:** A Recognising multi-digit numbers in photographs captured at street level is an important component of modern-day map making. A classic example of a corpus of such street-level photographs is Google's Street View imagery composed of hundreds of millions of geo-located 360-degree panoramic images.

  The ability to automatically transcribe an address number from a geo-located patch of pixels and associate the transcribed number with a known street address helps pinpoint, with a high degree of accuracy, the location of the building it represents. More broadly, recognising numbers in photographs is a problem of interest to the optical character recognition community.

  While OCR on constrained domains like document processing is well studied, arbitrary multi-character text recognition in photographs is still highly challenging. This difficulty arises due to the wide variability in the visual appearance of text in the wild on account of a large range of fonts, colours, styles, orientations, and character arrangements.

  The recognition problem is further complicated by environmental factors such as lighting, shadows, specularity, and occlusions as well as by image acquisition factors such as resolution, motion, and focus blurs. In this project, we will use the dataset with images centred around a single digit (many of the images do contain some distractors at the sides). Although we are taking a sample of the data which is simpler, it is more complex than MNIST because of the distractors.

- **DATA DESCRIPTION:** The SVHN is a real-world image dataset for developing machine learning and object recognition algorithms with the minimal requirement on data formatting but comes from a significantly harder, unsolved, real-world problem (recognising digits and numbers in natural scene images). SVHN is obtained from house numbers in Google Street View images.

Where the labels for each of this image are the prominent number in that image i.e. 2,6,7 and 4 respectively.

The dataset has been provided in the form of h5py files. You can read about this file format here: https://docs.h5py.org/en/stable/

Acknowledgement: Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, Andrew Y. Ng Reading Digits in Natural Images with Unsupervised Feature Learning NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011. PDF

http://ufldl.stanford.edu/housenumbers as the URL for this site.

- **PROJECT OBJECTIVE:** To build a digit classifier on the SVHN (Street View Housing Number) dataset.

  ### Steps and tasks: [Total Score: 30 Marks]

  1. Data Import and Exploration [5 Marks]
     A. Read the .h5 file and assign to a variable. [2 Marks]
     B. Print all the keys from the .h5 file. [1 Marks]
     C. Split the data into X_train, X_test, Y_train, Y_test [2 Marks]
  2. Data Visualisation and preprocessing [13 Marks]
     A. Print shape of all the 4 data split into x, y, train, test to verify if x & y is in sync. [1 Marks]
     B. Visualise first 10 images in train data and print its corresponding labels. [4 Marks]
     C. Reshape all the images with appropriate shape update the data in same variable. [3 Marks]
     D. Normalise the images i.e. Normalise the pixel values. [2 Marks]
     E. Transform Labels into format acceptable by Neural Network [2 Marks]
     F. Print total Number of classes in the Dataset. [1 Marks]
  3. Model Training & Evaluation using Neural Network [12 Marks]
     A. Design a Neural Network to train a classifier. [3 Marks]
     B. Train the classifier using previously designed Architecture (Use best suitable parameters). [3 Marks]
     C. Evaluate performance of the model with appropriate metrics. [2 Marks]
     D. Plot the training loss, validation loss vs number of epochs and training accuracy, validation accuracy vs number of epochs plot and write your observations on the same. [4 Marks]