greatlearning
Power Ahead

AIML | MODULE PROJECT

greatlearning
Power Ahead

5 Takeaways

1. AIML module projects are designed to have a detailed hands on to integrate theoretical knowledge with actual practical implementations.

2. AIML module projects are designed to enable you as a learner to work on realtime industry scenarios, problems and datasets.

3. AIML module projects are designed to enable you simulating the designed solution using AIML techniques onto python technology platform.

4. AIML module projects are designed to be scored using a predefined rubric based system.

5. AIML module projects are designed to enhance your learning above and beyond. Hence, it might require you to experiment, research, self learn and implement.

AIML | MODULE PROJECT

greatlearning
Power Ahead

NEURAL NETWORKS

PART I

AIML module project part I consists of industry based problem which can be solved using neural networks as a regressor.

PART III

AIML module project Part III consists of implementing a clickable GUI which can automate Part I & II of this project.

PART IV

AIML module project consists of building an image classifier using a neural network classifier.

PART II

AIML module project part II consists of industry based problem which can be solved using neural networks as a classifier.

TOTAL SCORE 60

PART ONE | PROJECT BASED | TOTAL SCORE | 10

- **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:** The need is to build a regressor which can use these parameters to determine the signal strength or quality [as number].

Steps and tasks: [Total Score: 10 points]

1. Import data.
2. Data analysis & visualisation
    - Perform relevant and detailed statistical analysis on the data.
    - Perform relevant and detailed uni, bi and multi variate analysis.

    Hint: Use your best analytical approach. Even you can mix match columns to create new columns which can be used for better analysis. Create your own features if required. Be highly experimental and analytical here to find relevant hidden patterns.
3. Design, train, tune and test a neural network regressor.

    Hint: Use best approach to refine and tune the data or the model. Be highly experimental here.
4. Pickle the model for future use.

PART TWO | PROJECT BASED | TOTAL SCORE | 10

- **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 deep 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:** The need is to build a classifier which can use these parameters to determine the signal strength or quality [as number].

Steps and tasks: [Total Score: 10 points]

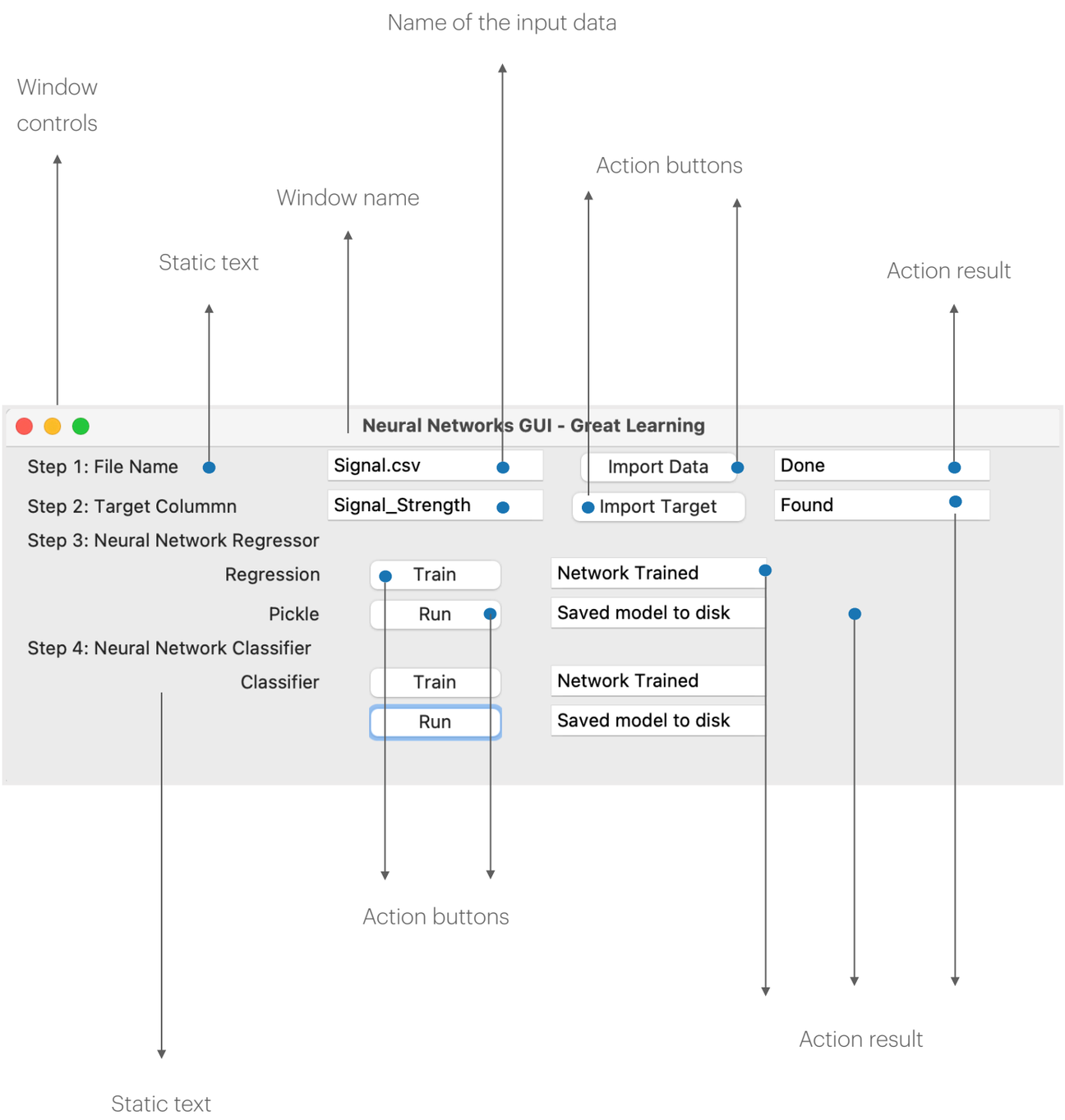
1. Import data
2. Data analysis & visualisation
    - Perform relevant and detailed statistical analysis on the data.
    - Perform relevant and detailed uni, bi and multi variate analysis.

    Hint: Use your best analytical approach. Even you can mix match columns to create new columns which can be used for better analysis. Create your own features if required. Be highly experimental and analytical here to find hidden patterns.
3. Design, train, tune and test a neural network classifier..

    Hint: Use best approach to refine and tune the data or the model. Be highly experimental here.
4. Pickle the model for future use.

PART THREE | PROJECT BASED | TOTAL SCORE | 10

- **CONTEXT:** Develop a clickable GUI [desk application or web service application] which can automate Part I & II of this project.



The above screenshot shows a sample GUI for reference purpose

PART FOUR | PROJECT BASED | TOTAL SCORE | 30

- **DOMAIN**: Autonomous Vehicles
- **BUSINESS 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: http://docs.h5py.org/en/stable/high/dataset.html
  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 when necessary

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

  Steps and tasks: [Total Score: 30 points]

  1. Import the data.
  2. Data pre-processing and visualisation.
  3. Design, train, tune and test a neural network image classifier.
     Hint: Use best approach to refine and tune the data or the model. Be highly experimental here to get the best accuracy out of the model.
  4. 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.

LEARNING OUTCOME

Hands on designing, training, testing and tuning a handmade a deep learning neural network.

Implementing a neural network as a classifier and a regressors on the same data.

Implementing a neural network image classifier.

Working with images and concepts of computer vision on how a model can be trained using images as a data source.

Implementing a working GUI

Hands on storing the best model and its weights for future usage avoiding the need to retraining the model for future.

" Put yourself in the shoes of an actual "

DATA SCIENTIST

THAT's YOU

Assume that you are working at the company which has received the above problem statement from internal/external client. Finding the best solution for the problem statement will enhance the business/ operations for your organisation/project. You are responsible for the complete delivery. Put your best analytical thinking hat to squeeze the raw data into relevant insights and later into an AIML working model.

PLEASE NOTE

Designing a data driven decision product typically traces the following process:

1. Data and insights:
   Warehouse the relevant data. Clean and validate the data as per the the functional requirements of the problem statement. Capture and validate all possible insights from the data as per the the functional requirements of the problem statement. Please remember there will be numerous ways to achieve this. Sticking to relevance is of utmost importance. Pre-process the data which can be used for relevant AIML model.
2. AIML training:
   Use the data to train and test a relevant AIML model. Tune the model to achieve the best possible learnings out of the data. This is an iterative process where your knowledge on the above data can help to debug and improvise. Different AIML models react differently and perform depending on quality of the data. Baseline your best performing model and store the learnings for future usage.
3. AIML end product:
   Design a trigger or user interface for the business to use the designed AIML model for future usage. Maintain, support and keep the model/ product updated by continuous improvement/training. These are generally triggered by time, business or change in data.

IMPORTANT POINTERS

Project should be submitted as a single ".html" and ".ipynb" file. Follow the below best practices where your submission should be:

- ".html" and ".ipynb" files should be an exact match.
- Pre-run codes with all outputs intact.
- Error free & machine independent i.e. run on any machine without adding any extra code.
- Well commented for clarity on code designed, assumptions made, approach taken, insights found and results obtained.

Project should be submitted on or before the deadline given by the program office.

Project submission should be an original work from you as a learner. If any percentage of plagiarism found in the submission, the project will not be evaluated and no score will be given.

greatlearning
Power Ahead

HAPPY LEARNING