

README

Brief Description:

The **Analysis** directory contains the following files and folders:

- **1) Analysis With Preprocessed Data of Dimension 8** : This folder contains the following graphs using which we analyse the model with standardized data of dimension 8:
 - **Accuracy_versus_Models_with_Increasing_Complexity_for_Preprocessed_Data_of_dimension_8.png**: This graph plots the Accuracy versus the Models of Increasing Complexity for all learning rates. For different model specifications, the accuracy attained is recorded and plotted on this graph.
 - **Accuracy_versus_Models_with_Learning_Rates_for_Preprocessed_Data_of_dimension_8.png**: This graph plots the Accuracy versus the Learning Rates for all models. For different learning rates, the accuracy attained is recorded and plotted on this graph.
- **2) Analysis With Reduced Dimensional Data after PCA**: This folder contains the following graphs obtained after performing the Test_A as described in the assignment:
 - **Accuracy_versus_Models_with_Increasing_Complexity_for_Reduced_Dimensional_Data_after_PCA.png**: This graph plots the Accuracy versus the Models of Increasing Complexity for all learning rates after PCA has been performed on the dataset. For different model specifications, the accuracy attained is recorded and plotted on this graph.
 - **Accuracy_versus_Models_with_Learning_Rates_for_Reduced_Dimensional_Data_after_PCA.png**: This graph plots the Accuracy versus the Learning Rates for all models after PCA has been performed on the dataset. For different learning rates, the accuracy attained is recorded and plotted on this graph.
- **Diabetes_Attributes_Correlation_Heatmap.png** : This image is the correlation matrix obtained on the features of the Diabetes dataset.
- **Plot_of_Reduced_Dimensional_Data_in_a_2D_Plane.png** : This image is the 2D plot of all the data samples in the dataset whose data points are colored differently based on class values.
- **Report.pdf**: Contains a detailed analysis report of our Multi Layered Perceptron Models using different hyperparameters such as `n_epochs`, `learning_rate`, `batch_learning` etc. and a section of extra analysis that we have performed to have better understanding of our models.

The **Project** directory contains the following files and folders:

- **Analysis** : Analysis is a folder which gets populated when `analysis.py` is run
- **diabetes.csv** : Contains the data used for training the Multi Layered Perceptron Models. The data is comprised of 768 samples with 8 continuous valued attributes and 1 binary valued target label.
- **requirements.txt** : Contains all the necessary dependencies and their versions
- **utility.py** : Contains all the helper functions such as `eval_model` etc.
- **MLP_models.py** : Contains the implementation of all the Multi Layered Perceptron Models when the number of input nodes is 8 and output node is 1. [BEFORE PCA]
- **r_MLP_models.py** : Contains the implementation of all the Multi Layered Perceptron Models when the number of input nodes is 2 and output node is 1. [AFTER PCA]
- **analysis.py** : Contains the python code we implemented to perform analysis on the Data and Multi Layered Perceptron Models.

Directions to use the code

1. Download the **Project** directory in into your local machine.
2. Ensure all the necessary dependencies with required version and latest version of Python3 are available (verify with `requirements.txt`)

```
pip3 install -r requirements.txt
```

3. Run the `analysis.py` file.
 - Initially, it will output some information regarding the project such as author etc.

- Next, it will **prompt you to enter the value of n_epochs**, which is the number of epochs for which all the Multi Layered Perceptron Models will be trained. The value of n_epochs has to be greater than or equal to 0. Please enter an integer value (n_epochs>=0). [We suggest the value of 500]
- Now, all the models are built with the above user given n_epochs as a parameter for all learning rates and all the plots are automatically generated and saved in their respective folders. (More details in the Report.pdf)
- After that, Principal Component Analysis is performed on the dataset and its dimension is reduced to 2.
- Again, all the models are built with the above user given n_epochs as a parameter for all learning rates and all the plots are automatically generated and saved in their respective folders. (More details in the Report.pdf)

Final Remarks

The dataset is taken from the following link. <https://www.kaggle.com/mathchi/diabetes-data-set>