MEEM 5990 Project 5

Total Points : 100

Deep Learning

This project deals with standard neural networks and convolutional neural networks. The objective of this project is to understand how certain hyperparameters affect the accuracy of the neural networks.

There are three parts in this project.
Part 1: NN for Auto MPG dataset
Part 2: NN for fashion-MNIST dataset
Part 3: CNN for fashion-MNIST dataset

Part 1

- Project5_Part1.ipynb is the code file for this part.
- Understand the dataset and write a short summary of the dataset. The summary should include but not limited to: number of variables, type of variables, number of data points, problem type (regression, classification etc), source of dataset, data of publication of the dataset, and other relevant details.
- You have to run the code multiple times by changing parameters like: number of layers, activation function, loss function etc. (see more in the code)
- Retain your best 3 runs. Save the code files as `Project5_Part1_run1.ipynb`, `Project5_Part1_run2.ipynb, Project5_Part1_run3.ipynb.` Run3 should be your best run, i.e. least mean squared error on test data.
  Note: You will have to run the code multiple times to arrive at the best three runs.
- Report the values in a table like this.

Run Number	Number or trainable parameters	Time required for training	Correlation Coefficient (Testdata)	Mean Squared error on test data
Run 1				
Run 2				
Run 3				

- Save the plots for Run3 (your best run). The plots include: correlation plot and Error vs epoch (learning curve).
- Report should contain
  - Data set summary
  - Above table
  - Plots for run3

    - Some deliverables of the report are mentioned as comments in the code.
- Deliverables
    - Report
    - Three code files: `Project5_Part1_run1.ipynb`, `Project5_Part1_run2.ipynb, Project5_Part1_run3.ipynb.`

Part 2

- Project5_Part2.ipynb is the code file for this part.
- Understand the dataset and write a short summary of the dataset. The summary should include but not limited to: number of variables, type of variables, number of data points, problem type (regression, classification etc), source of dataset, data of publication of the dataset, and other relevant details.
- You have to run the code multiple times by changing parameters like: number of layers, activation function, loss function etc. (see more in the code)
- Retain your best 3 runs. Save the code files as `Project5_Part2_run1.ipynb`, `Project5_Part2_run2.ipynb, Project5_Part2_run3.ipynb.` Run3 should be your best run, i.e, highest test accuracy.
  Note: You will have to run the code multiple times to arrive at the best three runs.
- Report the values in a table like this.

Run Number	Number or trainable parameters	Time required for training	Test Accuracy	Min value of true predictions
Run 1				
Run 2				
Run 3				

- Save the plots for Run3 (your best run). The plots include: Accuracy vs epoch (learning curve), Confusion matrix.
- Min value of true predictions is the least value in the diagonal of confusion matrix
- Report should contain
    - Data set summary
    - Above table
    - Plots for run3
    - Some deliverables of the report are mentioned as comments in the code.
- Deliverables
    - Firstname_lastname_project5.zip

  The zipped folder should contain
    1. Report
    2. Three code files: `Project5_Part2_run1.ipynb`, `Project5_Part2_run2.ipynb, Project5_Part2_run3.ipynb.`

Part 3

- Project5_Part3.ipynb is the code file for this part.
- Understand the dataset and write a short summary of the dataset. The summary should include but not limited to: number of variables, type of variables, number of data points, problem type (regression, classification etc), source of dataset, data of publication of the dataset, and other relevant details.
- You have to run the code multiple times by changing parameters like: number of layers, activation function, loss function etc. (see more in the code)
- Retain your best 3 runs. Save the code files as `Project5_Part3_run1.ipynb`, `Project5_Part3_run2.ipynb, Project5_Part3_run3.ipynb.` Run3 should be your best run.
  Note: You will have to run the code multiple times to arrive at the best three runs.
- Report the values in a table like this.

Run Number	Number or trainable parameters	Time required for training	Test Accuracy	Min value of true predictions
Run 1				
Run 2				
Run 3				

- Save the plots for Run3 (your best run). The plots include: Accuracy vs epoch (learning curve), Confusion matrix.
- Min value of true predictions is the least value in the diagonal of confusion matrix
- Report should contain
  - Data set summary
  - Above table
  - Plots for run3
  - Some deliverables of the report are mentioned as comments in the code.
- Deliverables
  - Report
  - Three code files: `Project5_Part3_run1.ipynb`, `Project5_Part3_run2.ipynb, Project5_Part3_run3.ipynb.`

Note: The plots should contain axis labels,titles, legend(wherever necessary) and the lines should be clear.

Deliverables

- Report (Single pdf file for all three parts)
- Codes (total 9 .ipynb files)

Point distribution

- 25 points: Part 1 (report+code)
- 20 points: Part 2 (report+code)
- 25 points: Part 3 (report+code)
- 10 points: Accuracy obtained in part 1. Points will be calculated as
$$\frac{Your\ highest\ R\ value}{Highest\ R\ value\ amongst\ all\ students} \times 10$$
- 10 points: Accuracy obtained in part 2. Points will be calculated as
$$\frac{Your\ highest\ test\ accuracy}{Highest\ test\ accuracy\ \ amongst\ all\ students} \times 10$$
- 10 points: Accuracy obtained in part 3. Points will be calculated as
$$\frac{Your\ highest\ test\ accuracy}{Highest\ test\ accuracy\ \ amongst\ all\ students} \times 10$$