ECEN 758 Data Mining and Analysis Project
Fall 2023

Due November 22, 2023, 11:59 PM CST

Project Description

In this project, you will develop a method to classify articles of clothing using FashionMNIST. You can implement a classifier yourselves or use a package/library. You can use any packages that come as a default option with Anaconda. If you use an artificial neural network, please use Pytorch. I need to be able to run your implementation on my machine or online (*e.g.*, Google Colab). So, be sure to get approval from me for any special packages! If I cannot run the code, you will lose points. Submit your final report as a PDF to Canvas: `https://canvas.tamu.edu/`. Share your code via an online link (*i.e.*, Github, Google Colab) in PDF, file upload, or other approved format (please ask Dr. Peeples if unsure about method to share code).

You **must** work in a group of 3 individuals. The signup sheet for groups is here (please find a group by October 30th).

If you have any questions, please reach out to Dr. Peeples.

Project Report

You should write a report that includes the sections listed below. Your report should follow the IEEE conference format (single spaced, double column). Focus your report on your training and validation strategies for the project, any unique implementations. Templates for the IEEE conference format can be found here.

The maximum number of pages for the report is 5 (including references). If there are any pages beyond page 5, they will discarded and not read or graded. It should be written with correct English grammar and spelling. Be precise - use pseudo-code and/or equations to be precise. Include visuals from code in final report that support your analysis.

Abstract A summary description of the contents of the report and your findings.
Introduction Overview of your experiment and a literature review. For the literature review, include any references to any relevant papers for your experiment. So, whatever you decide to do, search the ACM and IEEE (or other) literature for relevant papers to read and refer to.

Method Describe and outline any special implementation details for your project. A reader should be able to recreate your implementation and experiments from your project report.

Experiments Carefully describe your experiments with the training and validation datasets. Include a description of what the goal each individual experiment is and what your findings are. This is the bulk of what you will be graded on - if your experimental design is not sound or your experiments do not make sense, you will lose points.

Conclusion Describe any conclusions or things you learned from the project and/or future work. Your conclusions must follow from what you did. Do not copy something out of a paper or say something that has no experimental support in the Experiments section.

References Listing of all references in IEEE bibliography format.

Extra Credit: Website/Blog Post

It is important to make your work accessible for others. Your group can earn an additional 5 points on the project by creating a project website or blog post. An example project website is shown here. The website or blog post should have a link to your code and summarize the work completed in your report with supporting text and visuals. Include the link to website or blog post in final report.

Deliverables for project:

1. Final Report

2. Code

3. (Optional) Website/Blog post

Grading Rubric

1. Data Preparation (10 points)

    (a) Data cleansing and transformation (5 points)

    (b) Data splitting (*i.e.*, training, validation, and test splits) (5 points)

2. Exploratory Data Analysis (EDA) (10 points)

    (a) Descriptive statistics (*e.g.*, class distributions, data statistics) (5 points)

    (b) Data visualization (*e.g.*, plot examples from classes, dimensionality reduction approach to show data) (5 points)

3. Model Selection (20 points)

(a) Algorithm selection: Justification for choosing specific data mining algorithm (*e.g.*, decision trees, KNN, neural networks, etc.). (5 points)

(b) Model building: Developing and tuning selected model (*e.g.*, hyperparameter tuning) (10 points)

(c) Model evaluation: Proper evaluation metrics (e.g., accuracy, precision, recall, F1-score) and qualitative analysis (*e.g.*, confusion matrices) for classification on test set. (5 points)

4. Code and Software (10 points)

(a) Code quality (*i.e.*, well-documented, readable, and structured code) (5 points)

(b) Ability to run code (5 points)

5. Collaboration (10 points)

(a) Teamwork: Each individual will provide a ranking for their team members: excellent, very good, satisfactory, needs improvement, and unsatisfactory. Justification will be needed for rankings less than satisfactory. This will be submitted individually on Canvas. (5 points per team member)

6. Interpretability (10 point)

(a) Model interpretablity: Demonstrating an understanding of how the model makes predictions (*e.g.*, feature importance). (5 points)

(b) Business insights: Discussion of the practical implications of this work for a clothing company. (5 points)

7. Project Report (30 points)

(a) This component will be graded based on the requirements outlined above.

8. (Optional) Accessible Work (5 points)

(a) Website or blog post to share work completed on project.