CISC 867: Project 2 (10 points)

Data Preparation

1. In this project, you will use the Fashion-MNIST dataset using a CNN neural network architecture.
   You can use keras/tensorflow or write your own routines.

1) First, download the *data* file, load it, and
    1. Describe the data
    2. Clean the data
    3. Check the data for missing values or duplicates and carry out proper correction methods
    4. Visualize the data using proper visualization methods.
    5. Draw some of the images
    6. Carry out required correlation analysis
2) Carry out any required preprocessing operations on the data
3) Encode the labels

Training a CNN neural network

In this exercise you need to implement a LeNet-5 network to recognize the Fashion-MNIST digits.

- Modify hyperparameters to get to the best performance you can achieve.
- Evaluate the model using 5-fold cross-validation.
- In the report, provide a plot of accuracy improvement using the previously mentioned techniques. Also plot the convergence curve for LeNet-5.
- Comment on why you think LeNet-5 further improves the accuracy if any at all. And if it doesn't, why not?
- Note that this is an open-ended problem. So, there could be very good solutions.
- Try to use other two CNN models (using transfer learning) and compare the results with the full trained LeNet-5.

Use Google's Colab to save time.

Submission: Please include the following files (project_2.zip):

- project_2.pdf/docx and project_2.py/ipynb