Programming Assignment 3: Association Analysis and Deep learning (10 points)

Shen-Shyang Ho (Dr.)

October 27, 2023

- In this assignment, you will be using the dataset assigned to you in Assignment 1 for a Deep Learning Task and a newly assigned dataset for an association analysis task. For the new dataset, use the number assigned to you earlier in the semester.
- For the association analysis task, you will use Mlxtend (http://rasbt.github.io/mlxtend/)
- For the deep learning task, you will use Tensorflow (https://www.tensorflow.org/) via Keras (https://keras.io/) or PyTorch (https://pytorch.org/)

1. **[Association Rule Generation from Transaction Data]**

   (a) Download transaction dataset to your local drive.

      i. **Go to the following Google Drive link (Students must be logged in to their Rowan accounts)**: https://drive.google.com/drive/folders/1LuFEbgq3IvisEXT1jOZ-H4jWeqzqEH3m?usp=sharing

   (b) Download the 'Grocery_Items_{DATASET_NUMBER}.csv' file from the Google Drive Link. DATASET_NUMBER is the number assigned to you earlier in the semester.

   (c) Using minimum support = 0.01 and minimum confidence threshold = 0.1, what are the association rules you can extract from your dataset? (0.5 point)
   (see http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/)

   (d) Use minimum support values (msv): 0.001, 0.005, 0.01, 0.05 and minimum confidence threshold (mct): 0.05, 0.075, 0.1. For each pair (msv, mct), find the number of association rules extracted from the dataset. Construct a heatmap using Seaborn data visualization library (https://seaborn.pydata.org/generated/seaborn.heatmap.html) to show the count results such that the x-axis is msv and the y-axis is mct. (2.5 points)

   (e) List the association rule(s) (i.e., one or more rules depending on your dataset) that have the highest confidence for minimum support = 0.005. What is that confidence value? (1 point)

2. **[Image Classification using CNN]** Construct a 4-class classification model using a convolutional neural network with the following simple architecture (2 point)

   i 1 Convolutional Layer with 8 3×3 filters.

   ii 1 max pooling with 2×2 pool size

   iii **Flatten** the Tensor

   iv 1 hidden layer with 16 nodes for fully connected neural network

   v Output layer has 4 nodes (since 4 classes) using 'softmax' activation function.

   (Use 'Relu' for all layers except the output layer.) for 20 epochs using 'adam' optimizer and 'categorical cross entropy' loss function. If your machine is too slow, you can reduce to 5 epochs. You can perform more epochs (> 20) if you want to. For validation split, you will use 20%. For batch size, you can pick a size that will not slow down the training process on your machine. (see https://keras.io/examples/vision/mnist_convnet/)

   - Plot a graph to show the learning curves (i.e., x-axis: number of epochs; y-axis: training and validation accuracy - 2 curves) (1 points)
   - Perform **ONE** of the following experiment below ((a), (b) or (c)) based on the last digit of your Rowan Banner ID (1 point):
     (a) Train the CNN using 2 other filter sizes: 5×5 and 7×7 for the convolution layer (i) with all other parameters unchanged
     (b) Train the CNN using 2 other number of filters: 4 and 16 for the convolution layer (i) with all other parameters unchanged
     (c) Train the CNN using 2 other number of nodes in the hidden layer (iv): 8 and 32 with all other parameters unchanged

     If the last digit is $\{0, 1, 2, 3\}$, do (a). If the last digit is $\{4, 5, 6\}$, do (b). If the last digit is $\{7, 8, 9\}$, do (c). State your Rowan Banner ID in your submission so that we know which experiment you are doing.
   - Plot the learning curves (i.e., x-axis: number of epochs; y-axis: training and validation accuracy - 2 curves) for the classification models using the above 2 different parameter values (1 points)
   - Describe and discuss what you observe by comparing the performance of the first model and the other two models you constructed in (a), (b) or (c) (depending on which one you did). Are there model overfit or underfit or just right? (1 point)