MANE 4962: Machine Learning for Engineering
Homework 3
Points:100

This homework is designed to test your understanding of clustering, classification, regression, and model-based learning processes. Use a Jupyter notebook to solve individual problems. You can submit a pdf printouts of the notebooks. Please include your name and RIN in the PDF. Please keep your notebooks and Python codes organized in your Github repository.

1. **(15 pts.)** Image segmentation is a process to highlight useful regions in images. Use the skimage.io module to load the following image. Afterwards, segment the image into multiple useful regions using the k-means clustering method. The segmented image should highlight, for example, the dashboard, the driver's arms, cars ahead etc., by grouping similar pixels together. You do not need to split the data into train and test set for this problem.

2. **(40 pts.)** Implement your own linear machine learning model optimized with **mini-batch gradient descent method** to predict the price of a house in a city with population of 160, 000. Train the model to fit the housing prices dataset found on LMS. Vary the batch size from 1,5,10, and 20. Plot the objective function, J for each batch size. You do not need to split the data into train and test set for this problem. What happens when you use batch size equal to one?

3. **(20 pts.)** Use the Scikit-learn breast cancer Wisconsin dataset and a logistic regression model to classify breast cancers. You must use a cross-validation method to progressively reduce the number of features and find the best two features to perform the classification. Evaluate the model using various classification metrics and report your findings. Use a 70%-30% split.

4. **(25 pts.)** Construct a neural network with a single hidden layer containing two neurons using Tensorflow. Use ReLU as activation function. Optimize the network with stochastic gradient descent method. Choose mean squared error to calculate the loss. Fit the housing prices dataset found on LMS using the network. Use the trained neural network model to predict the price of a house in a city with population of 165, 000. Calculate a useful regression metric. Plot the training and validation losses. Use a 70%-30% split for the training and validation dataset. The architecture of the neural network and the optimizer are fixed for this problem. Therefore, you need to choose a suitable learning rate and number of epochs to minimize the loss. Explain the trends you found in the plots for training and validation losses.