Statistical Machine Learning (SML)
Winter 2021

Assignment 4

Due Date: 14th April, 2021

Instructions:

1. You are free to use either python or MATLAB for this assignment.
2. You can use inbuilt libraries for Math, plotting, and handling the data (eg. NumPy, Pandas, Matplotlib).
3. Usage instructions for other libraries can be found in the question.
4. Only (*.py) and (*.m) files should be submitted for code.
5. Create a (*.pdf) report explaining your assumptions, approach, results, and any further detail asked in the question.
6. You should be able to replicate your results if required.

A. In this question, you will explore autoencoders.

1. [5] Generate 1000 multivariate gaussian points from $N(\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{bmatrix}$

   Take 800 samples for training and call the remaining samples test set.
   Normalize the data so that value in each dimension lies in [0,1] and a sigmoid can be used in the output layer.

2. [25] Implement an autoencoder with one input layer, one hidden layer, and one output layer. Train the model using data generated in part 1.

   Note the following:
   a. No. of nodes in the hidden layer should be one less than the input layer.
   b. Evaluation metric - MSE
   c. Backpropagation has to be derived (Pen-Paper), and implemented from scratch. The derivation also needs to be submitted and will be evaluated. Note that the equations should demonstrate update of weights. Updates of all the weights should be captured in equations.

3. [10] Report and plot the training and testing loss after each epoch.

4. [20] Perform backpropagation using autograd and compare the results with previous
   ` part.

Graded Only for Course Code 542 students. (Others can also try)

B. [40] Repeat question 1 for binary inputs (dimension = 3), everything else will remain the same.
You can use multivariate Bernoulli distribution to generate data.