Learning outcomes:

Training Neural Networks on multiple data sets for

- Classification problem (Demo)
- Regression problem
- Digit recognition problem

1. A marketing department of a bank runs various marketing campaigns for cross-selling products, improving customer retention and better customer services. Let us take a marketing data sample in which the bank is interested in selling a term deposit product to its existing customers.
   Contacting all customers is costly and does not create good customer experience. So, the bank wants to build a predictive model which will identify customers who are more likely to respond to the marketing campaign.

2. The dataset "housing.csv" has information on the housing values in suburbs of Boston. Predict the median value of the owner occupied homes "OwnOcc" using neural net algorithm. The attributes description is given below:

   - Crimerate: per capita crime rate by town
   - ResiLandZone: proportion of residential land zoned for lots over 25,000 sq.ft.
   - INDUS: proportion of nonretail business acres per town
   - CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
   - NOX: nitric oxides concentration (parts per 10 million)
   - Rooms: average number of rooms per dwelling
   - AGE: proportion of owner occupied units built prior to 1940
   - Distance: weighted distances to five Boston employment centers
   - RAD: index of accessibility to radial highways
   - TAX: full value property tax rate per $10,000
   - PTRATIO: pupil teacher ratio by town
   - Blacks: 1000(Bk 0.63)^2 where Bk is the proportion of blacks by town
   - LSTAT: % lower status of the population
     OwnOcc: Median value of owner occupied homes in $1000's

Steps

a) Do necessary preprocessing steps required for neural network like converting categorical to numerical(use dummies) ,standardizing numerical attribute, split in to train and test etc..

b) Change   eval.metric parameter as mx.metric.rmse and play with the other parameters
   model <- mx.mlp (train.x, train.y, hidden_node=10, out_node=1,
                activation="tanh",out_activation="rmse", num.round=20,
                array.batch.size=100, learning.rate=0.07, momentum=0.9,

eval.metric=mx.metric.rmse)

c) Predict with test data
d) Compute error metrics (use regr.eval in DMwR package)

3. Digit Recognition problem. The dataset have images corresponding to 10 digits (0-9)
Each sample in the dataset is a greyscale image of 28x28 resolution. Consider each pixel as a variable and train the neural network to classify samples.
Note: The data shouldn't be normalized in the regular manner, we need to divide the pixel values by 255 to normalize the data instead of using range method.