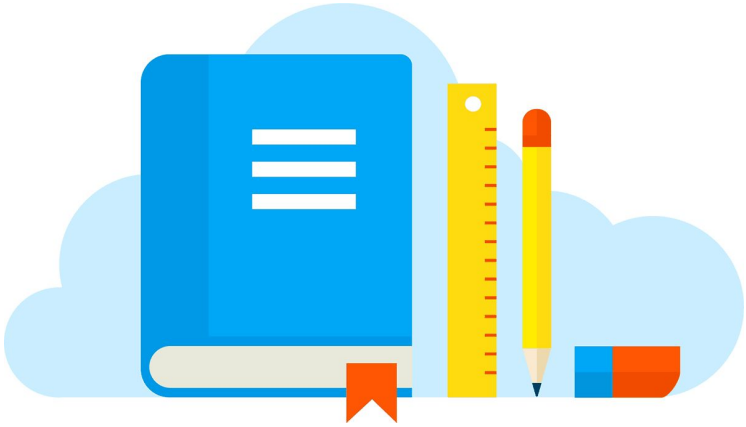


Assignment 2

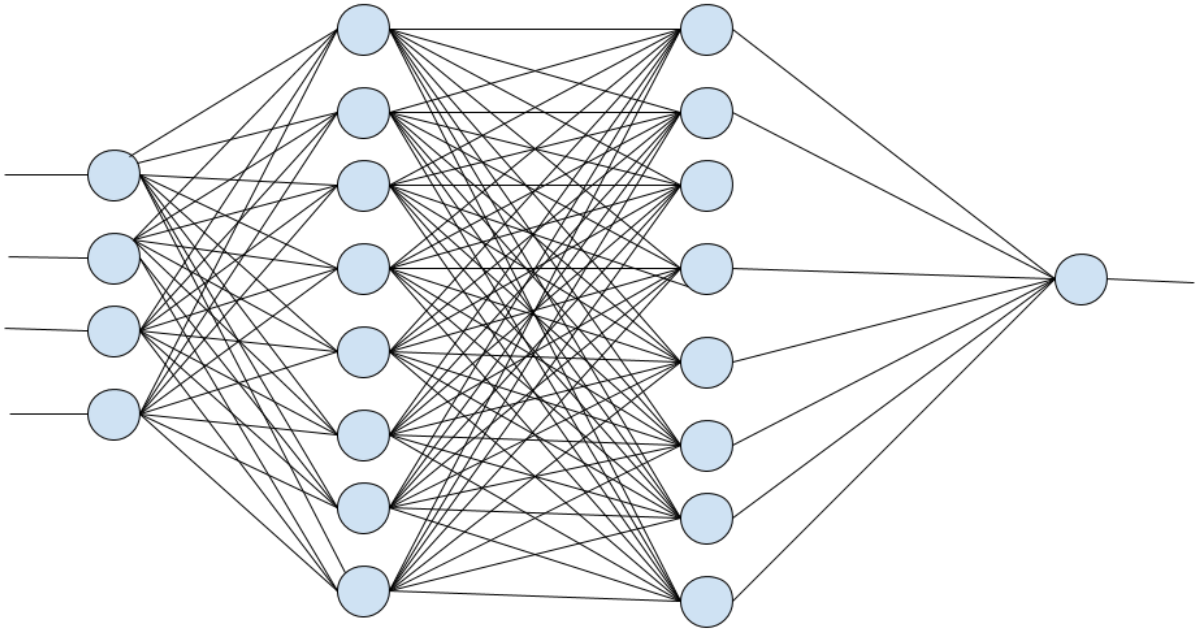
Date: 16/04/2017

Feature Engineering

- Kept the numerical features as it is already present.
- For non-numerical features, I assigned them some numbers based on my observation:
    - Work Class: 1 if private, 0 otherwise.
    - Education: 1 for 11th, 2 for Bachelors, 3 for HS-grad, 4 for some-college, 5 for masters, 6 for doctorate, 0 otherwise.
    - Marital-status: 1 for married-civ-spouse, 0 otherwise
    - Occupation: 1 for prof-speciality, 0 otherwise
    - Relationship: 2 for husband, 1 for not-in-family, 0 otherwise
    - Race: 1 for white, 0 otherwise
    - Sex: 1 for male, 0 otherwise
    - Native-country: 1 for United-States, 0 otherwise.
- Normalised each feature by subtracting the minimum value and then dividing by the interval so that all features have value in the range 0 to 1.

Neural-network

- Used sigmoid function for classification in each neuron.
- Added two hidden layers each with twice the number of neurons as input layer(which consists of given features). Output layer with single neuron. Here for depicting in diagram I have considered 4 features. Actually there are 15 features.



- There are 3 weights matrices. Weights0 store weights between input and first hidden layer. Weights1 store weights between first and second hidden layers. Weights2 store weights between second hidden layer and output layer.
- These weights are trained using forward and backward propagation.
- Finally these weight matrices along with normalization factors are stored after training the neural network to the file weight.txt.

- The file test_net.py uses these values of weights to predict the output for given test data-set and stores the predictions in the file predictions.csv.

Comparison of results-obtained

- Used readymade library from **sklearn** to get predictions for results obtained using SVM, SGD and decision trees.

- As far as Kaggle score is concerned:
    - My neural network: Public score: 0.85460 Private score: 0.85197.
    - Standard SVM: Public score: 0.72242 Private score: 0.73304.
    - Standard SGD: Public score: 0.70725 Private score: 0.71516.
    - Standard decision tree: Public score: 0.74704 Private score: 0.75214.

- Hence best results are obtained for neural network than other 3 standard classification techniques.