COMS 3007: Machine Learning Assigment 2016

Find a suitable dataset on which you can apply various supervised learning algorithms.
This can be either a classification problem or a regression problem.
Apply various supervised learning algorithms to your dataset.

You must **submit a document** to Sakai containing the following information:

(1) A description of your dataset: what are the attributes, what are the targets, how many datapoints do you have, and some sample datapoints from the dataset. State what you are trying to predict with the data.

(2) A description of how you structured your inputs/targets and normalised the data, and the split into training/validation/test data.

(3) A list of supervised learning algorithms you applied to the data, together with the details of each implementation and the error on the test set.
For example:
Neural network with one hidden layer with 50 nodes, $\eta = 0.3$, error = ...
Neural network with two hidden layers with 20 nodes in first layer and 15 in second, $\eta = 0.2$, error = ...
RBF network with 25 RBF nodes and random centres, $\eta = 0.25$, error = ...
RBF network with 40 RBF nodes and centres obtained by k-means, $\eta = 0.25$, error = ...
etc.

(4) A brief discussion of your results from the various algorithms. E.g., what worked best/worst and why you think this is so.

Notes:

- Your dataset must be sufficiently large and with enough attributes. The more interesting, the better... (If you are really stuck for ideas, try one of the Weka datasets.)
- You can use your own implementations of algorithms, or those on Weka, or Matlab or anything. Just state what you have used.
- The more algorithms you try, the better...
- If you have some nice visualisations/graphs, please include them.

Important:

- The closing date for submission is the end of **Friday, May 13, 2016**.
- This assignment counts 10% of your final mark.