Deadline: April 19, 2018

One Layer Neural Network

Implement 1-layer neural network recognising languages of text documents. Neural network should be able to recognise 3 or more different languages. Number of languages is unknown and should be detected based on a data in training set.

1. Prepare training set in directory 'training'. Create sub-directories with text documents. Each sub-directory should contain at least about 10 documents in one language (You may use Wikipedia articles). Each Sub-directory should be named by appropriate language code (i.e. pl, en, fr).
2. As an input of neural network count relative frequency (letterFreq/numberOfLetters) of each letter in each document. Use only 26 of ASCII characters. One input vector represents relative frequencies of 26 letters in one document.
3. Choose local representation of the output.
4. Input vector and weight vector should be normalized (weight vector after each update).
5. Instead of a discrete perceptron, you may consider other activation functions like linear function or sigmoid function (continous activation function has different learning rule than discrete). Use a maximum selector to classify language.
6. Train of neural network until the network error from the whole training set is smaller than chosen maximum error.
7. Testing data can be provided in external files or through user interface.