Assignment 11

Applied Machine Learning

1. [50 pts] In this assignment, we will use Apriori analysis to find phrases, or interesting patterns in a novel.

   Note that you are free to use any Apriori analytics and algorithm library in this assignment.

   Use the `nltk` library corpus `gutenberg` API and load the novel 'carroll-alice.txt' which is the Alice in Wonderland by L. Carroll. There are 1703 sentences in the novel which can be represented as 1703 transactions. Use any means to parse/extract words and save in CSV format to be read by Weka framework similar to the Apriori Analysis module.

   Hint: Removing stop words and symbols using regular expressions can be helpful:

```
from nltk.corpus import gutenberg, stopwords
Stop_words = stopwords.words('english')
Sentences = gutenberg.sents('carroll-alice.txt')
TermsSentences = []
for terms in Sentences:
    terms = [w for w in terms if w not in Stop_words]
    terms = [w for w in terms if re.search(r'^[a-zA-Z]{2}', w) is not None]
```

   If Weka picked, use FPGrowth and start with default parameters. Reduce `lowerBoundMinSupport` to reach to a sweet point for the support and avoid exploding the number of rules generated.

   Report interesting patterns.

   (Example: Some of the frequently occurring phrases are Mock Turtle, White Rabbit, etc.)

2. [50 pts] In the lecture module, the class `NeuralNetMLP` is a single hidden layer neural network implementation. Make the necessary modifications to upgrade it to a 2 hidden layer network. Run it on the MNIST dataset and report its performance.

   (Hint: Raschka, Chapter 12)