Vancouver Summer Program
Package G (Linguistics)
Assignment 3 - Mongolian Syllable Stress
Due: Thursday August 10

For this homework, you will design a neural network that learns the syllable stress pattern in Mongolian. Due to the large number of problems with Pybrain in class, you will ***not*** have to create a full network. You will only need to create a set of inputs and output. This can be done without Pybrain.

The language

Mongolian is the national language of Mongolia, spoken by roughly 10 million people, mainly in eastern Asia.

Vowels in Mongolian can be either "short" or "long", which refers to the actual duration of the vowel. Short vowels last about 60-80 milliseconds, while long vowels last 120-180 milliseconds. Long vowels will be indicated by writing a vowel twice. This distinction is important, and there are pairs of words which differ only in vowel length. For example:

sulter "emblem"
suulter "tail"

Vowel length is also important for the syllable stress pattern of Mongolian. For this assignment, we will use a simplified version of the stress rule:
If there are no long vowels, then the first syllable of the word is stressed. Otherwise, the right-most long vowel is stressed.

(a) xa.da "mountain"
(b) ga.luu "goose"
(c) xon.dii.ryy.len "to separate"

In example (a), there are no long vowels, so the first syllable is stressed. In example (b) there is a long vowel, so it gets stressed. In example (c) there are two long vowels, and the second one (the right-most one) gets stressed.

The assignment

For this homework, you will design a neural network that takes a Mongolian word as input, and it outputs the correct stress pattern. *You do not actually need to create the full network.*

There is a list of Mongolian words available in the Assignment #3 folder on Connect. The words have syllable breaks marked with a period. Stress is marked with an apostrophe at the beginning of the stressed syllable. You need to write a Python script that transforms the words in this list into inputs and their corresponding outputs for a neural network.

This means you will need to decide on the number of input nodes and output nodes (don't worry about hidden nodes). You will also need to decide on system for representing Mongolian words as activation values (numbers) for the network.

What you need to submit

There are three files to submit:

1. A Python script that transforms Mongolian words into activation values for the network.

2. A file that contains Mongolian words with their input and output representations. This file should be generated by your Python script, not typed by hand.

3. A text document that describes what you did. You should discuss your system for representing Mongolian words, and how many input/output nodes your network would need. You don't need to write up a full essay. Short sentences and point-form are acceptable, as long as you are clear.

Tips

This is a pattern based on syllables, so your network doesn't need information about every consonant and vowel.

Your network does need information about vowel length (but not which specific vowel).

Not all Mongolian words are the same length, but all inputs to a network must have the same length. You'll need some way of dealing with short words, and you should discuss this in the text document you submit.

Look at the common/proper/acronym network we built in class. That may help.

Again, you do not need to actually build, train, and test a network. You only need to create the inputs and outputs for a hypothetical network.