Of all machine learning techniques, neural networks are probably best known, at least in name, by the general public. Foundational work on the implementation of artificial neurons (and networks) was done in the 1940s, before “Artificial Intelligence” was established as a field of study. (The year 2006, by the way, marks the fiftieth anniversary of the Dartmouth Conference, at which AI was formally established as a field of research and was given its name.) Since that time, they have found support among computer scientists as a mechanism for solving optimization problems, while also finding popularity in the world of science fiction. Neural networks are of central importance, for example, in 2001: A Space Odyssey, in which they are highlighted as the key reason for the existence of HAL.

This week we will study artificial neural networks and the backpropagation learning algorithm.

1 Backpropagation

1.1 Reading

Please read

- Mitchell, Chapter 4 (at least through Section 4.6.2), and
- Witten and Frank, pages 227-233.

While Alpaydin’s text includes a chapter on this topic, I thought I’d point you to the appropriate readings in Mitchell, as there seems to be clear agreement among you that Mitchell is better written. I trust that you will be able to follow the derivation of the update rules, as they are simply extensions of the update rules for simple perceptrons, which we considered during the first week of the semester.

1.2 Exercises

This week your primary exercise will involve the implementation of the backpropagation algorithm. While this algorithm is conceptually fairly simple (the big idea, after all, is walking along the error surface to find the weights that minimize squared error), the details can be tricky. The algorithm is certainly non-trivial to debug. Please don’t wait until the last minute to begin your implementation.

As all of the texts above point out, neural networks are not necessarily the representation of choice for classification tasks. They are difficult to understand (compare them decision trees, for example), and they take a long time to train. They are, however, useful in many related contexts, which is why they are so worthwhile to study.

While classification is not necessarily the best context in which to view neural networks from the standpoint of performance, we will focus on applying them to this task. We will do this in order to better understand how neural networks and backpropagation compare to the techniques we have already studied.

Please implement the stochastic version of backpropagation this week. (Mitchell explains why this is often the best backpropagation option.) In order to constrain the task somewhat, you may make the following assumptions:

- You never need to model a neural network with more than one hidden layer.
• The neural network will be used to represent classification tasks.

• The inputs and classes are nominal- (i.e., discrete-)valued.

• The input to your system will be a data file in the ”ARFF” format that weka assumes. This should allow you to reuse your file-reading code from the earlier Naive Bayes assignment.

In order to handle nominal attributes and classes in a neural network, you will need to treat the values as if they are numeric. Say that you have a binary-valued attribute or class. This is easy – you can simply treat such an attribute or class as if it has one of the two possible values 0 and 1. If an attribute or class has multiple possible values, then you will need to create a node for each possible value. Say, for instance, that you are implementing a neural network for the “weather” data. The “outlook” attribute can take one of three values – sunny, rainy, cloudy. You would handle this by creating three input nodes. For any given example, only one of these nodes would have the value 1; the other two would be 0.

You can test your implementation on the weather and contact lens data sets. You might find it useful to compare your results with the results of weka’s implementation of backpropagation. To find the classifier in weka, look for “Multilayer perceptron” under “functions”.

In addition to writing the program, you should go over it so that you will be prepared to present it in detail during your tutorial session. A detailed presentation includes not only a trace through the program, but an explanation of the choices you made in design.

2 Multitask Learning

Witten and Frank tell us that:

The same technique can be applied to predict several targets, or attribute values, simultaneously by creating a separate output unit for each one. Intuitively, this may give better predictive accuracy than building a separate classifier for each class attribute if the underlying learning tasks are in some way related.

This idea – multitask learning – has been extensively studied.

2.1 Reading

Please read


You can find the link to this paper by going to the Assignments page for this course. This paper describes multitask learning first in the context of backpropagation but then extends it to other learning techniques that we have studied.

2.2 Exercise

By this point in the semester your experience with machine learning algorithms and familiarity with at least some research should place you in a good position to critically analyze this paper. For the tutorial session you should be prepared to not only discuss the general results presented in the paper, but should be able to contribute to a lively discussion that critiques the work.
In order to better prepare for such a discussion (to be guided by you, the students, rather than me), please write a review of the paper. I would like you to follow the general format that we researchers use when reviewing papers for conferences or journals. The review should begin with a very brief assessment of your own expertise/comfort level in reviewing the paper. It should then give a summary of the paper, including all of the major results. Next it should include a critique. You should include both positive and negative comments. You should comment on such areas as importance, technical soundness, and clarity.

Please type your report, which should be a minimum of two pages of text (12-point font, 1.5 spacing).