1 Notes

This document will cover notes for artificial neural networks.

Characteristics of problems for which neural nets are appropriate:

- Instances/examples are made of many attribute-value pairs. Input attributes can be highly correlated or independent of each other. Inputs can be any real values.
- The target function can be any discrete (categorical) value, real value or vector of real or discrete values.
- Neural networks are resilient to errors in the data, so the inputs *may* contain errors.
- Training times are allowed to be long. Training neural networks can take large amounts of time which are dependent
- Quick evaluation/prediction times are required. Though they take a long time to learn/train, neural networks are able to compute predictions given inputs fairly quickly.
- It is not required that the learned target function is easily understandable by humans. Node number and weights is large in complicated neural networks and it is not likely that humans will easily understand them.

1.1 Perceptrons

1.1.1 Definition

Perceptrons are the most basic unit in a neural network. Given a set of real value inputs a perceptron calculates a weighted linear combination of those inputs and outputs 1 if the calculated value is greater than some threshold and -1 otherwise.

$$\sigma(x_1, ..., x_n) = \begin{cases} 1 \text{ if } \omega_0 + \omega_1 x_1 + ... + \omega_n x_n > 0 \\ -1 \text{ otherwise} \end{cases}$$

Here the value of $\sum_1^n \omega_n x_n$ must be greater than some threshold $-\omega_0$ in order to output a value of 1. This is the activation threshold.

Given the definition of a perceptron, the hypothesis space is the possible values of the weight vectosr $\vec{\omega}$ and can be defined as:

$$H = \{\vec{\omega} \mid \vec{\omega} \in \mathbb{R}^{n+1}\}$$

1.1.2 Representation

Conceptually a perceptron represents a hyper plane. The dimension of the space in which the hyper plane is located is determined by the number of inputs to the perceptron. For example, a perceptron which has two inputs represents a line while a perceptron with three inputs represents a plane. Given this a single preceptron has the ability to linearly separate

2 Questions

Hopefully these questions will be answered in the course of writing these notes.

1. How are categorical inputs/outputs handled? Is each category assigned an integer value?
2. What kinds of activation functions are suited for categorical inputs? Is the Heavyside function still used or is some other activation function more suited?
3. How do certain factors affect the training time of a neural network. Which parameters have the largest effect? How does training time scale with respect to data set size and complexity?