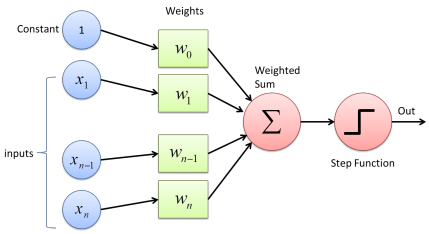
1 The Beginning

The ability to learn and memorize has always fascinated scientist and philosophers. With the surfacing of computers the comparisons between computers and brains has influenced research on both sides. In 1958 Frank Rosenblat, an American psychologist, published “The perceptron: a probabilistic model for information storage and organization in the Brain”. This work marks the beginning of deep learning.

1.1 The Perceptron

An electronic device/program that is built according to biological principles and shows an ability to learn. The perceptron is, in essence, a single artificial neuron. It has an input layer, which is equivalent to the dendrites of a real neuron. These inputs are then summed up at the ”cell”body. The summation output (Net) is passed through and a step activation-function which outputs the perceptron output. The step function emits 0 or 1 depending whether the inputs summation reached a threshold or not. This is similar to the action-potential of real neurons which is elicited if the inputs to the neurons have summed up and reached a threshold membrane potential. Noticed in the below figure that we have added to the input X an additional constant (1). This is called the bias. Why do we add this value? This will become clear later.



1.2 Perceptron Learning Algorithm

The perceptron learning algorithm belongs to the supervised learning family, i.e. learning from labelled examples.
The learning algorithm is called the delta-rule and is given bellow:
1. Initialize weights vector (W) with small random numbers
2. Repeat until convergence:
    a. Loop over feature vector (X_j) and labels (l_i) in training set D:
    b. Take X and pass it through the perceptron, calculating the output values:

$$y_j = f(wt) \cdot x_j) \qquad (1)$$

    c. update weights

$$w_i(t+1) = w_i(t) + \alpha(l_j - y_j)x_{j,i} \qquad \forall 0 <= i <= n \qquad (2)$$

3. Termination Criteria:
    a. When all samples are correctly labeled
    b. After a set number of epochs (iteration over entire data set is called an epoch)
    c. After a pre-set number of epochs where the percentage of misclassified labels remains stable.

The learning step:

$$w_i(t+1) = w_i(t) + l_j - y_j)x_{j,i} = w_i(t) + \alpha(l_j - w(t)_j)x_{j,i} \qquad \forall 0 <= i <= n \qquad (3)$$

alpha is the learning rate - a very important hyper-parameter!

NEXT IS PRACTICAL 1 - THE PERCEPTRON