Perceptrons

Kristian Wichmann

January 24, 2017

1 Definition

A *perceptron* is the simplest possible linear feed-forward neural net. It is a binary classifier consisting of a list of input features feeding into one, linear threshold binary neuron.

So, if there's n features, a data point x is a n-dimensional vector. The classifier is characterized by another n-dimensional weight vector w and a constant b. The classification can be summed up as:

$$f(x) = \text{sign}(w^T x + b) \tag{1.1}$$

Here, the two possible outcomes are +1 (positive) and -1 negative.

1.1 Bias as a weight

It is possible to treat b as a weight term by making each sample vector into a $n+1$-dimensional one where the zero'th feature is simply a 1. Then b is then simply the zero'th element of the weight vector. The classification algorithm can then be written:

$$f(x) = \text{sign}(w^T x) \tag{1.2}$$

2 Picking features vs. learning

So if we wanted to set up a perceptron classifier, the first step would be to decide on the features. This is a step that is done by us - human beings - and thus no machine learning is happening in this step. But based on common sense and domain knowledge, a set of features is extracted from the raw input.

When that has been done, the training set is used for letting the perceptron learn appropriate values for w. This is the learning part of the perceptron.

3 Training algorithm (for a separable traning set)

It turns out that if the training set is indeed linearly separable, then the following algorithm will always produce an appropriate weight vector w:

- Initialize the weight vector w
- Pick a random x_i from the training set and do the following:
  - Is x_i correctly classified according to the current weights?
  - If yes, do nothing.
  - If x_i is misidentified as a negative, then add x_i to w.
  - If x_i is misidentified as a positive, then subtract x_i from w.
- Repeat with a new training sample.

If indeed the training set is linearly separate (and this may not be the case!), then the weight vector w will almost surely (i.e. with probability 1) eventually correspond to one such classifier.