Perceptron Learning Algorithm

Nick R. Porter

March 30, 2017

1 Perceptron Overview

- Binary classifier, can only be used to predict -1 or 1.
- If the data is not linearly separable then the algorithm will never terminate unless a max number of iterations is set.
- The algorithm takes a vector x for each feature of the data.
- We also learn a vector w of weights each weight corresponds to a feature in x. Sometimes have a w_0 which is a constant/offset.
- Returns a value (-1, 1) that corresponds to a class of the data.
- We can define a function $g(x, w) = w_0x_0 + w_1x_1 + ... + w_mx_m = w \cdot x$
- If $g(x, w)$ returns a value 0 or greater(or some threshold $theta$) return 1. Otherwise return -1
- Leads us to define an activation function $\phi(g(x, w)) = 1$ if $g(x, w) \geq 1$, else -1
- We learn these weights from training data

2 Training

- Algorithm outline:
  - Initialize all weights to 0 or small random numbers.
  - For each training sample $x^{(i)}$ perform the following setps:
    - Compute the predicted value, $\hat{y}$
    - Update weights
    - Update to each weight is defined as $w_j := w_j + \Delta w_j$
    - Δw is computed using the perceptron learning rule.
    - $\Delta w_j = \alpha(y^{(i)} - \hat{y}^{(i)})x_j^{(i)}$
  - Here alpha is the learning rate, usually a small number around 0.01.
  - We see that if the prediction is correct then $\Delta w_j = 0$