Introduction to Machine Learning

Neural Networks

Mingchen Gao

March 4, 2019
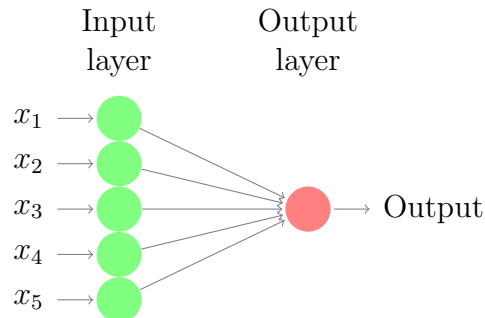
Outline

Contents

1 Extending Perceptrons

- Questions?

- Why not work with thresholded perceptron?
  - Not differentiable
- How to learn non-linear surfaces?
- How to generalize to multiple outputs, numeric output?

The reason we do not use the thresholded perceptron is because the objective function is not differentiable. To understand this, recall that to compute the gradient for perceptron learning we compute the partial derivative of the objective function with respect to every component of the weight vector.

$$\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i} \frac{1}{2} \sum_j (y_j - \mathbf{w}^\top \mathbf{x}_j)^2$$

Now if we use the thresholded perceptron, we need to replace $\mathbf{w}^\top \mathbf{x}_j$ with o in the above equation, where o is -1 if $\mathbf{w}^\top \mathbf{x}_j < 0$ and 1, otherwise. Obviously, given that o is not smooth, the function is not differentiable. Hence we work with the unthresholded perceptron unit.

2 Multi Layered Perceptrons

2.1 Generalizing to Multiple Labels

- Distinguishing between multiple categories
- *Solution:* Add another layer - **Multi Layer Neural Networks**

Multi-class classification is more applicable than binary classification. Applications include, handwritten digit recognition, robotics, etc.