Adaptive Linear Neuron

Nick R. Porter

March 31, 2017

1 Adaline Overview

- Main difference to perceptron is the activation function.

- Activation function is define as $\phi(x, w) = x \cdot w$

- Activation function is used for learning the weights

- A new function called the quantizer is used to predict class labels

- Cost function SSE: $J(w) = \frac{1}{2}\Sigma_i((y^{(i)} - \phi(z^{(i)}))^2$

- Use gradient descent to minimize the cost function.

- Update of w is same, $w := w + \Delta w$

- $\Delta w = -\alpha \Delta J(w)$

- $\Delta J(w) = -\Sigma_i(y^{(i)} - \phi(z^i))x_j^{(i)}$

- Also referred to as 'batch gradient descent'