Machine Learning

Assignment 7.1

Submitted By: Ranji Raj

December 9, 2020

Derivation

Given output of the single perceptron:

$$o = w_0 + w_1 x_1 + w_1 x_1^2 + ... + w_n x_n + w_n x_n^2$$

Above equation can be generalised in vector form as,

$$o = \vec{w}.\vec{x}_d + \vec{w}.\vec{x}_d^2 \tag{1}$$

The training error of a weight vector is defined as,

$$E(\vec{w}) \equiv \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2 \tag{2}$$

And gradient is given as:

$$\nabla E[\vec{w}] \equiv [\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, ..., \frac{\partial E}{\partial w_n}]$$

Differentiating, (2) wrt w_i,

$$\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i} \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

$$\frac{\partial E}{\partial w_i} = \frac{1}{2} \sum_{d \in D} \frac{\partial}{\partial w_i} (t_d - o_d)^2$$

$$\frac{\partial E}{\partial w_i} = \frac{1}{2} \sum_{d \in D} 2(t_d - o_d) \frac{\partial}{\partial w_i} (t_d - o_d)$$

$$\therefore \quad \frac{\partial E}{\partial w_i} = \sum_{d \in D} (t_d - o_d) \frac{\partial}{\partial w_i} (t_d - \vec{w}.\vec{x}_d - \vec{w}.\vec{x}_d^2)$$

from (1)

$$\frac{\partial E}{\partial w_i} = \sum_{d \in D} (t_d - o_d)(-x_{i,d} - x_{i,d}^2)$$

$$\frac{\partial E}{\partial w_i} = -\sum_{d \in D} (t_d - o_d)(x_{i,d} + x_{i,d}^2) \tag{3}$$

Above equation is nothing but, $\nabla E[\vec{w}]$ Therefore, the gradient descent training rule can be denoted as,

$$\Delta \vec{w} = -\eta \nabla E[\vec{w}]$$

here η is a positive constant called as the *learning rate*, which determines the step size in the gradient descent search.
The negative sign indicates the movement of the weight vector in the direction that *decreases E*.