COMP 462 Introduction to Machine Learning
Assignment 4 - Gradient Descent

Elif Erden 041503038

20.03.19

1 Introduction

In this assignment, derivation of the Gradient Descent rule was made for both classification and regression cases. Given the perceptron output and the error term $E[\vec{w}] = \frac{1}{2}\sum_{d \epsilon D}(t_d - o_d)^2$, the update equation for the perceptron becomes $w_i \leftarrow w_i + \Delta w_i$ where the Δw_i is calculated as:

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i} = \eta \sum_{d \epsilon D}(t_d - o_d) x_{i,d}$$

2 Derivation for Regression

In regression, the output of the perceptron is $o(\vec{x}) = \vec{w}.\vec{x}$. In equation (1), this output is substituted into the error term equation. Then derivative of E is taken with respect to w_i in equations (2),(3),(4). Finally, if the derivation is substituted in the update equation for the perceptron, equation (5) is obtained.

$$E[\vec{w}] = \frac{1}{2}\sum_{d \epsilon D}(t_d - \sum_{i=0}^{n}(w_i x_i))^2 \tag{1}$$

$$\frac{\partial E}{\partial w_i} = \frac{1}{2}\sum_{d \epsilon D} 2(t_d - \sum_{i=0}^{n}(w_i x_i))(t_d - \sum_{i=0}^{n}(w_i x_i))' \tag{2}$$

$$\frac{\partial E}{\partial w_i} = \sum_{d \epsilon D}(t_d - \sum_{i=0}^{n}(w_i x_i))(-x_i) \tag{3}$$

$$\frac{\partial E}{\partial w_i} = \sum_{d \epsilon D}(t_d - o_d)(-x_i) \tag{4}$$

$$\Delta w_i = \eta \sum_{d \epsilon D}(t_d - o_d)(x_i) \tag{5}$$

3 Derivation for Classification

In classification, the output of the perceptron is $o(\vec{x}) = sigmoid(\vec{w}.\vec{x})$ where $sigmoid(a) = \frac{1}{1+e^{-a}}$. In equation (6), this output is substituted into the error term equation. Then derivative of E is taken with respect to w_i in equations (7),(8),(9). Finally, if the derivation is substituted in the update equation for the perceptron, equation (10) is obtained.

$$E[\vec{w}] = \frac{1}{2}\sum_{d \epsilon D}(t_d - \sum_{i=0}^{n}\frac{1}{1+e^{-w_i x_i}})^2 \tag{6}$$

$$\frac{\partial E}{\partial w_i} = \frac{1}{2}\sum_{d \epsilon D}2(t_d - \sum_{i=0}^{n}\frac{1}{1+e^{-w_i x_i}})(t_d - \sum_{i=0}^{n}\frac{1}{1+e^{-w_i x_i}})' \tag{7}$$

$$\frac{\partial E}{\partial w_i} = \sum_{d \epsilon D}(t_d - \sum_{i=0}^{n}\frac{1}{1+e^{-w_i x_i}})\frac{-x_i e^{-w_i x_i}}{(1+e^{-w_i x_i})^2} \tag{8}$$

$$\frac{\partial E}{\partial w_i} = \sum_{d \epsilon D}(t_d - o_d)\frac{-x_i e^{-w_i x_i}}{(1+e^{-w_i x_i})^2} \tag{9}$$

$$\Delta w_i = \eta \sum_{d \epsilon D}(t_d - o_d)\frac{x_i e^{-w_i x_i}}{(1+e^{-w_i x_i})^2} \tag{10}$$

3.1 Sigmoid Function

Sigmoid function that is used for classification is shown in Figure 1.

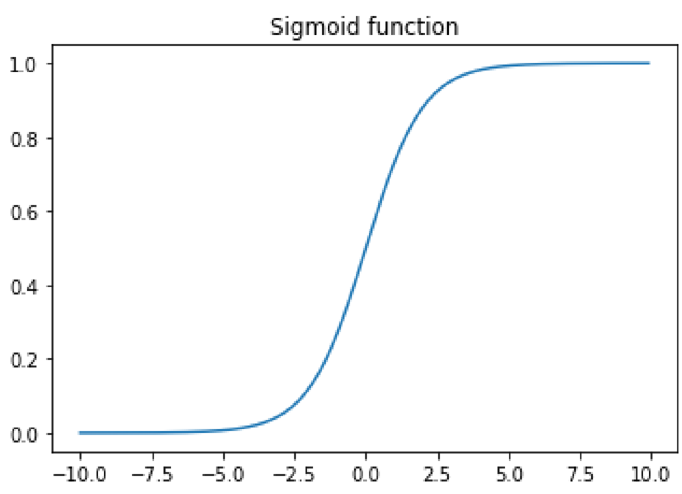

Figure 1: Sigmoid function