task_8.1

Machine Learning (WiSe 2025/2026)

Author: Suvansh Shukla
Matriculation No. 256245

Assignment 8 Task 1

The gradient descent training rule for a single perceptron is the partial derivative of output function w.r.t each of the different weights.

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

The training rule, used to update weights uses this gradient like this:

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

$$\triangle w_i = -\eta \frac{\partial E}{\partial w_i}$$

The given error function is the SSE Error function:

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

The output function is:

$$o = w_0 + w_1 x_1 + w_1 x_1^2 + \ldots + w_n x_n + w_n x_n^2$$

This can be re-written as:

$$o = w_0 + \sum_{i=1}^{n} w(x_i + x_i^2)$$

The partial derivative of the SSE error function results in this:

$$\frac{\partial E}{\partial w_i} = \sum \frac{\partial E}{\partial o_d} \frac{\partial o_d}{\partial w_i}$$

First we try to find the value of $\frac{\partial E}{\partial o_d}$, where we know the value of E here is the given SSE error function. So,

$$\frac{\partial E}{\partial o_d} = \frac{\partial}{\partial o_d}\left[\frac{1}{2}(t_d - o_d)^2\right]$$

$$\frac{\partial E}{\partial o_d} = \frac{1}{2}.\, 2(t_d - o_d).\, \frac{\partial}{\partial o_d}(t_d - o_d)$$

$$\frac{\partial E}{\partial o_d} = (t_d - o_d).\, (-1)$$

$$\frac{\partial E}{\partial o_d} = (o_d - t_d)$$

Let this be equation 1.

Now we need to calculate the partial derivative to find the value of $\frac{\partial o_d}{\partial w_i}$, this is where we replace o_d with the output function we just derived.

$$\frac{\partial o_d}{\partial w_i} = \frac{\partial}{\partial w_i}\left[w_o + \sum_{j=1}^{n} w_j(x_{j,d} + x_{j,d}^2)\right]$$

$$\frac{\partial o_d}{\partial w_i} = x_{j,d} + x_{j,d}^2$$

Let this be equation 2.

Now a special case we need to consider is the fact that for weight w_0, the derivative would simply be 1, because the partial derivative of a variable is 1 and everything else in the equation doesn't have a w_0 variable so it can be treated as a constant, i.e. 0. Therefore,

$$\frac{\partial o_d}{\partial w_0} = 1$$

Final Gradient Descent

After all of this the final gradient descent becomes:

$$\frac{\partial E}{\partial w_i} = \sum \frac{\partial E}{\partial o_d}\frac{\partial o_d}{\partial w_i}$$

replacing partial derivatives with their solved forms, i.e. equation 1 & 2:

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

Finally in the format of weight updation it becomes:

$$\triangle w_i = -\eta \sum (o_d - t_d)(x_{i,d} + x_{i,d}^2)$$

The equation for the bias weight would be a special form of the equation above:

$$\triangle w_0 = -\eta \sum (o_d - t_d)$$