Sheet 2 Exercise 2

November 6, 2024

Ex 2.1 Derive the updaate funciton for w

to be added...

Ex 2.2 Gradient

To find the function that has the minimal sum when it comes to the distance between funciton and datapoints, we take the gradiant of our sum function. The Gradient finds the minimum because it alwys "moves" in the direction of the negativle "slope of out function.

Ex 2.3: One Step of Gradient Descent

Given:

$$\mathbf{w}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \alpha = 0.25$$

and the data points:

x_1	x_2	y
-5	0	0
-3	-2	0
2	5	1
4	1	1

1. Include the bias term in each feature vector:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -5 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ -3 \\ -2 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}$$

2. Initial predictions $p_n = \sigma(\mathbf{x}_n^T \mathbf{w}_0) = 0.5$ for all n because w_0 is 0 and $\sigma(0) = \frac{1}{1+e^0} = \frac{1}{1+1}$

3. Computing the gradient:

$$\nabla J = \frac{1}{4}\sum_{n=1}^{4}(p_n - y_n)\mathbf{x}_n$$

with $p_n = 0.5$ and y_n values:

$$\nabla J = \frac{1}{4}\left((0.5-0)\mathbf{x}_1 + (0.5-0)\mathbf{x}_2 + (0.5-1)\mathbf{x}_3 + (0.5-1)\mathbf{x}_4\right)$$

$$= \begin{bmatrix} 0,5 \\ -2,5 \\ 0 \end{bmatrix} + \begin{bmatrix} 0,5 \\ 1,5 \\ -1,0 \end{bmatrix} + \begin{bmatrix} -0,5 \\ -1,0 \\ -2,5 \end{bmatrix} + \begin{bmatrix} -0,5 \\ -2,0 \\ -0,5 \end{bmatrix} = \begin{bmatrix} 0 \\ -7 \\ -4 \end{bmatrix}$$

Calculate each term and sum them up to find ∇J, then update the weights:

$$\mathbf{w}_1 = \mathbf{w}_0 - \alpha \nabla J$$

$$\mathbf{w}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -7 \\ -4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1,75 \\ 1 \end{bmatrix}$$

Thus, after one step of gradient descent, the updated weight vector $\mathbf{w}_1$ is:

$$\mathbf{w}_1 = \begin{bmatrix} 0 \\ 1,75 \\ 1 \end{bmatrix}$$

Ex 2.4: predicting the Probability

Using the updated weight vector:

$$\mathbf{w}_1 = \begin{bmatrix} 0 \\ 1.75 \\ 1 \end{bmatrix}$$

and $P(y = 1|X = [-1, 1]^T)$.

input vector $\mathbf{x}$ with added bias term:

$$\mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

computing $\mathbf{w}_1^T\mathbf{x}$:

$$\mathbf{w}_1^T\mathbf{x} = \begin{bmatrix} 0 \\ 1.75 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = 0 \cdot 1 + 1.75 \cdot (-1) + 1 \cdot 1 = -1.75 + 1 = -0.75$$

calculation P usong sigmoid...

$$P(y = 1|X = [-1, 1]^T) = \sigma(-0.75) = \frac{1}{1 + e^{0.75}} \approx 0.32$$

So, the predicted probability is approximately 0.32.