Homework 5

Sunday, October 6, 2024 9:41 PM

Part One

1. In your own words, describe what the purpose of gradient descent is in machine learning.

The purpose of gradient descent is to help a machine learning model make accurate predictions by constantly refining the best possible set of parameters.

2. Why is it called "descent"?

Gradient descent is called "descent" because it involves descending down the loss function until it is as accurate as possible.

3. Explain what a gradient is in the context of a function.

In the context of a function, a gradient is a vector of partial derivatives. Mathmatically:

$$\nabla f(x_1, x_2, \ldots, x_n) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$$

each component of the resultant vector tells us how much this function changes when an input changes.

4. How does the gradient guide the optimization process?

The gradient guides the optimization process by pointing in the direction of the steepest increase in the function's value. This helps us to minimize the function.

5. Define the term "learning rate" in gradient descent. What happens if the learning rate is too small? Too large?

The learning rate in gradient descent is a "hyper-parameter" that determines the size of the steps taken during the optimization process when updating parameters. If it is too small, the iterations down the curve will be too small. This makes optimization an impossibly slow process. If the learning rate is too large, the iterations will be too large causing the model to overshoot the minimum.

6. Explain the difference between global and local minima in optimization problems.

The optimization process requires us to find the global minima of the loss function, this it is incredibly important. Local minima occor on the loss function when it is lower than the adjacent but there may exist another point on the function that is lower. Global minima are also local minima. Local minima may hurt optimization, when it stops there instead of the global minima.

7. How can gradient descent be affected by local minima?

Because the gradient descent becomes very small, the optimization process may stop updating the parameters, incorrectly assuming it reached the global minima.

Part Two

1. Given the following function: $f(x) = x^2 + 3x + 5$. If the initial value $x_0 = 5$ and the learning rate $\alpha = 0.1$, calculate the new value of x after three iterations of gradient descent. Does x get closer to the minimum?

$$f'(x) = \frac{d(x^2 + 3x + 5)}{dx}$$

$$f'(x) = 2x + 3$$

Rul for gradient descent: $x_{t+1} = x_t - \alpha f'(x)$

Iteration 1:

$f'(5) = 2(5) + 3 = 13$ gradient

$x_1 = 5 - 0.1(13) = 3.7$

Iteration 2:

$f'(3.7) = 2(3.7) + 3 = 10.4$ gradient

$x_2 = 3.7 - 0.1(10.4) = 2.66$

Iteration 3:

$f'(2.66) = 2(2.66) + 3 = 8.32$ gradient

$x_3 = 2.66 - 0.1(8.32) = 1.828$

Because the values are decreasing and moving closer to zero we can determine it is getting closer to the minimum.

2. For a multivariable function $f(x) = x^2 + y^2 + 2x + 4y + 6$. Calculate the gradient of the function and then perform one step of gradient descent starting from the point $(x_0, y_0) = (1, 1)$ with a learning rate $\alpha = 0.1$.

$$\frac{\partial f(x,y)}{\partial x} = 2x + 2$$

$$\frac{\partial f(x,y)}{\partial y} = 2y + 4$$

$\nabla f(x,y) = (2x + 2, 2y + 4)$

$2(1) + 2 = 4$ - Partial X

$2(1) + 4 = 6$ - Partial Y

$\nabla f(1,1) = \langle 4, 6 \rangle$

$x_1 = 1 - 0.1(4) = 0.6$

$y_1 = 1 - 0.1(6) = 0.4$

$(x_1, y_1) = (0.6, 0.4)$ ☐