Artificial Neural Networks

Dr. Muhammad Aqib

University Institute of Information Technology

PMAS-Arid Agriculture University Rawalpindi

Optimization Algorithms

Derivatives

- The derivative of a constant equals 0 (m is a constant in this case, as it's not a parameter that we are deriving with respect to, which is x in this example).
  - $\frac{d}{dx}\, 1 = 0$
  - $\frac{d}{dx} m = 0$
- The derivative of x equals 1.
  - $\frac{d}{dx} x = 1$
- The derivative of a linear function equals its slope.
  - $\frac{d}{dx} mx + b = m$

Derivatives (cont.)

- The derivative of a constant multiple of the function equals the constant multiple of the function's derivative.
  - $\frac{d}{dx}[k.f(x)] = k.\frac{d}{dx}\, f(x)$

- The derivative of a sum of functions equals the sum of their derivatives.
- $\frac{d}{dx}[f(x) + g(x)] = \frac{d}{dx}\, f(x) + \frac{d}{dx} g(x) = f'(x) + g'(x)$

- The derivative of a difference of functions equals the differnce of their derivatives.
- $\frac{d}{dx}[f(x) - g(x)] = \frac{d}{dx}\, f(x) - \frac{d}{dx} g(x) = f'(x) - g'(x)$

Derivatives (cont.)

- The derivative of an exponentiation
  - $\frac{d}{dx}x^n = n \cdot x^{n-1}$

- We used the value x instead of the whole function f(x) here since the derivative of an entire function is calculated a bit differently.

Partial Derivatives

- The partial derivative measures how much impact a single input has on a function's output.
- The method for calculating a partial derivative is the same as for derivatives discussed before.
- We simply have to repeat this process for each of the independent inputs.

- Each of the function's inputs has some impact on this function's output, even if the impact is 0.
- We need to know these impacts.
  - This means that we have to calculate the derivative with respect to each input separately to learn about each of them. That's why we call these partial derivatives.

Partial Derivative (cont.)

- We are interested in the impact of singular inputs since our goal, in the model, is to update parameters. The ∂ operator means explicitly that; the partial derivative

- $f(x, y, z) = \frac{\partial}{\partial x} f(x, y, z), \ \frac{\partial}{\partial y} f(x, y, z), \ \frac{\partial}{\partial z} f(x, y, z)$

Partial Derivative of a Sum

Calculating the partial derivative with respect to a given input means to calculate it like the regular derivative of one input, just while treating other inputs as constants. For example:

$$f(x, y) = x + y \quad \rightarrow \quad \frac{\partial}{\partial x} f(x, y) = \frac{\partial}{\partial x}[x + y] = \frac{\partial}{\partial x}x + \frac{\partial}{\partial x}y = 1 + 0 = 1$$

$$\frac{\partial}{\partial y} f(x, y) = \frac{\partial}{\partial y}[x + y] = \frac{\partial}{\partial y}x + \frac{\partial}{\partial y}y = 0 + 1 = 1$$

First, we applied the sum rule — the derivative of a sum is the sum of derivatives. Then, we already know that the derivative of *x* with respect to *x* equals *1*. The new thing is the derivative of *y* with respect to *x*. As we mentioned, *y* is treated as a constant, as it does not change when we are deriving with respect to *x*, and the derivative of a constant equals *0*. In the second case, we derived with respect to *y*, thus treating *x* as constant. Put another way, regardless of the value of y in this example, the slope of *x* does not depend on *y*. This will not always be the case, though, as we will soon see.

Partial Derivative of a Sum (cont.)

$$f(x,y) = 2x + 3y^2 \quad \rightarrow \quad \frac{\partial}{\partial x} f(x,y) = \frac{\partial}{\partial x}[2x + 3y^2] = \frac{\partial}{\partial x}2x + \frac{\partial}{\partial x}3y^2 =$$

$$= 2 \cdot \frac{\partial}{\partial x}x + 3 \cdot \frac{\partial}{\partial x}y^2 = 2 \cdot 1 + 3 \cdot 0 = 2$$

$$\frac{\partial}{\partial y} f(x,y) = \frac{\partial}{\partial y}[2x + 3y^2] = \frac{\partial}{\partial y}2x + \frac{\partial}{\partial y}3y^2 =$$

$$= 2 \cdot \frac{\partial}{\partial y}x + 3 \cdot \frac{\partial}{\partial y}y^2 = 2 \cdot 0 + 3 \cdot 2y^1 = 6y$$

Partial Derivative of Multiplication

$$f(x, y) = x \cdot y \quad \rightarrow \quad \frac{\partial}{\partial x} f(x, y) = \frac{\partial}{\partial x}[x \cdot y] = y \frac{\partial}{\partial x} x = y \cdot 1 = y$$

$$\frac{\partial}{\partial y} f(x, y) = \frac{\partial}{\partial y}[x \cdot y] = x \frac{\partial}{\partial y} y = x \cdot 1 = x$$

Partial Derivative of Multiplication (cont.)

$$f(x,y,z) = 3x^3z - y^2 + 5z + 2yz \quad \rightarrow$$

$$\begin{aligned}
\frac{\partial}{\partial x} f(x,y,z) &= \frac{\partial}{\partial x}[3x^3z - y^2 + 5z + 2yz] = \\
&= \frac{\partial}{\partial x}3x^3z - \frac{\partial}{\partial x}y^2 + \frac{\partial}{\partial x}5z + \frac{\partial}{\partial x}2yz = \\
&= 3z \cdot \frac{\partial}{\partial x}x^3 - \frac{\partial}{\partial x}y^2 + 5 \cdot \frac{\partial}{\partial x}z + 2 \cdot \frac{\partial}{\partial x}yz = \\
&= 3z \cdot 3x^2 - 0 + 5 \cdot 0 + 2 \cdot 0 = 9x^2z
\end{aligned}$$

Partial Derivative of Multiplication (cont.)

$$f(x, y, z) = 3x^3z - y^2 + 5z + 2yz \quad \rightarrow$$

$$\begin{aligned}
\frac{\partial}{\partial y} f(x, y, z) &= \frac{\partial}{\partial y}[3x^3z - y^2 + 5z + 2yz] = \\
&= \frac{\partial}{\partial y} 3x^3z - \frac{\partial}{\partial y} y^2 + \frac{\partial}{\partial y} 5z + \frac{\partial}{\partial y} 2yz = \\
&= 3 \cdot \frac{\partial}{\partial y} x^3z - \frac{\partial}{\partial y} y^2 + 5 \cdot \frac{\partial}{\partial y} z + 2z \cdot \frac{\partial}{\partial y} y = \\
&= 3 \cdot 0 - 2y + 5 \cdot 0 + 2z \cdot 1 = -2y + 2z
\end{aligned}$$

Partial Derivative of Multiplication (cont.)

$$f(x,y,z) = 3x^3z - y^2 + 5z + 2yz \quad \rightarrow$$

$$\begin{aligned}
\frac{\partial}{\partial z} f(x,y,z) &= \frac{\partial}{\partial z}[3x^3z - y^2 + 5z + 2yz] = \\
&= \frac{\partial}{\partial z}3x^3z - \frac{\partial}{\partial z}y^2 + \frac{\partial}{\partial z}5z + \frac{\partial}{\partial z}2yz = \\
&= 3x^3 \cdot \frac{\partial}{\partial z}z - \frac{\partial}{\partial z}y^2 + 5 \cdot \frac{\partial}{\partial z}z + 2y \cdot \frac{\partial}{\partial z}z = \\
&= 3x^3 \cdot 1 - 0 + 5 \cdot 1 + 2y \cdot 1 = 3x^3 + 5 + 2y
\end{aligned}$$

Gradient

- The partial derivative is a single equation, and the full multivariate function's derivative consists of a set of equations called the gradient.
- In other words, the gradient is a vector of the size of inputs containing partial derivative solutions with respect to each of the inputs.
- Let's see the partial derivative we calculated earlier.

$$f(x, y, z) = 3x^3z - y^2 + 5z + 2yz \quad \rightarrow$$

$$\frac{\partial}{\partial x} f(x, y, z) = 9x^2z$$

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

$$\frac{\partial}{\partial z} f(x, y, z) = 3x^3 + 5 + 2y$$

Gradient (cont.)

- If we calculate all of the partial derivatives, we can form a gradient of the function. Using different notations, it looks as follows.

$$\nabla f(x,y,z) = \begin{bmatrix} \frac{\partial}{\partial x} f(x,y,z) \\ \frac{\partial}{\partial y} f(x,y,z) \\ \frac{\partial}{\partial z} f(x,y,z) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix} f(x,y,z) = \begin{bmatrix} 9x^2 z \\ -2y + 2z \\ 3x^3 + 5 + 2y \end{bmatrix}$$

- It's a vector of all of the possible partial derivatives of the function, and we denote it using the ∇ — nabla symbol that looks like an inverted delta symbol.

Optimization Algorithms

- Loss tells us how poorly the model is performing at that current instant.
- We use this loss to train our network such that it performs better.
- We take the loss and try to minimize it because a lower loss means our model is going to perform better.
- The process of minimizing (or maximizing) any mathematical expression is called optimization.

- Optimizers are algorithms or methods used to change the attributes of the neural network such as weights and learning rate to reduce the losses.
- Optimizers are used to solve optimization problems by minimizing the function.

Working of Optimizers

- Think of a hiker trying to get down a mountain with a blindfold on. It's impossible to know which direction to go in, but there's one thing he can know.
  - If he's going down (making progress) or going up (losing progress).
- Eventually, if she keeps taking steps that lead her downwards, she'll reach the base.
- Similarly, it's impossible to know what your model's weights should be right from the start.
- But with some trial and error based on the loss function (whether the hiker is descending), you can end up getting there eventually.
- How you should change your weights or learning rates of your neural network to reduce the losses is defined by the optimizers you use.
- Optimization algorithms are responsible for reducing the losses and to provide the most accurate results possible.

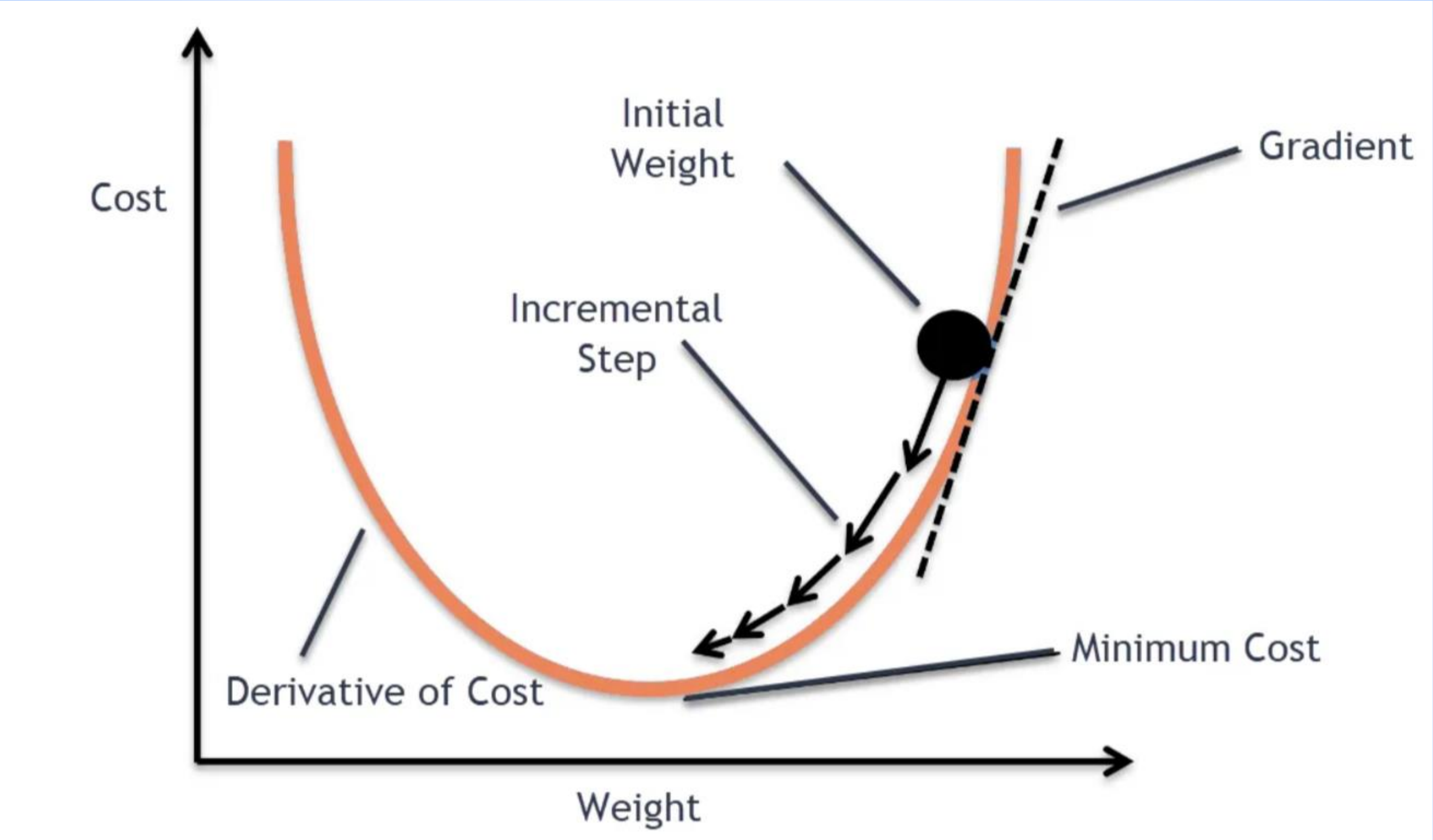
Types of Optimizers

- Gradient Descent
- Stochastic Gradient Descent (SGD)
- Mini Batch Stochastic Gradient Descent (MB-SGD)
- SGD with momentum
- Nesterov Accelerated Gradient (NAG)
- Adaptive Gradient (AdaGrad)
- AdaDelta
- RMSprop
- Adam
- …

Gradient Descent

- This is an iterative optimization algorithm for finding the minimum of a function.
- In deep learning neural networks are trained by defining a loss function and optimizing the parameters of the network to obtain the minimum of the function.
- The optimization is done using the gradient descent algorithm which operates in these two steps.
  - Compute the slope (gradient) that is first order derivative of the function at the current point.
  - Move in the opposite direction of the slope increase from the current point by the computed amount.

Gradient Descent (cont.)



Thank You