

Ch 6.3.1: Multidimensional Gradient Descent

author: Riley Primeau date: '2022-12-07'

autosize: true

Gradient Descent

Gradient descent is used for optimization problems where you need to find a maximum/minimum. Like the Newton-Raphson method for finding roots, this method relies on knowing the first derivative and is an iterative process.

- Finds derivative at x
- Takes a step down/up of size h in the direction of the slope found at x
- As the function approaches a peak/trough through each iteration, it will result in smaller changes in x
- Once the change in x decreases below the tolerance level, a maximum or minimum has been reached

Gradient Descent in 1 Dimension

Gradient Descent can be Expanded to Multiple Dimensions

- Altering the 1-dimensional gradient descent function to work with multiple dimensions is fairly simple
- The variables for x and the derivative must be made into vectors to accommodate for more dimensions
- Partial derivatives for each variable input are now needed

Implementation

```
gd = function(fp, x, h = 1e2, tol = 1e-4, m = 1e3)
{
  iter = 0

  oldx = x
  x = x - h * fp(x)

  while(vecnorm(x - oldx) > tol)
  {
    iter = iter + 1
    if(iter > m)
```

```

    return(x)
    oldx = x
    x = x - h * fp(x)
}
return(x)
}

```

Things to note...

- `fp` is a vector of functions
- `x` is a vector of values
- The function `vecnorm(x, p=2)` finds the magnitude of the vector. The power `p` can be changed but the default value of `p` is 2. This is known as the Euclidean Norm

More important aspects

- Each variable will converge to their values at different rates
- Increasing the step limit improves results but increases flops
- Larger step size typically means faster convergence BUT too large of a step size can lead to unpredictable results

Analogy

A person is stuck in the mountains and is trying to get down (i.e., trying to find the global minimum). There is heavy fog such that visibility is extremely low. Therefore, the path down the mountain is not visible, so they must use local information to find the minimum. They can use the method of gradient descent, which involves looking at the steepness of the hill at their current position, then proceeding in the direction with the steepest descent (i.e., downhill).

2D Visualizations

Applications of Multidimensional Gradient Descent

- Training machine learning neural networks
- <https://youtu.be/aircAruvnKk?t=334>
- Gradient descent is used during back propagation in order to update weights to make more accurate predictions

References

- Howard, II, J.P. (2017). Computational Methods for Numerical Analysis with R (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315120195>
- Wikipedia
- “SGD.” Hasty.ai, 16 Nov. 2022, <https://hasty.ai/docs/mp-wiki/solvers-optimizers/sgd>.
- Jiang, Lili. “A Visual Explanation of Gradient Descent Methods (Momentum, AdaGrad, RMSProp, Adam).” Medium, Towards Data Science, 21 Sept. 2020, <https://towardsdatascience.com/a-visual-explanation-of-gradient-descent-methods-momentum-adagrad-rmsprop-adam-f898b102325c>.