Berkeley Engineering | BerkeleyHaas

PROFESSIONAL CERTIFICATE IN MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Office Hour #15 with Matilde D’Amelio

June 30, 2022 at 9 pm UTC

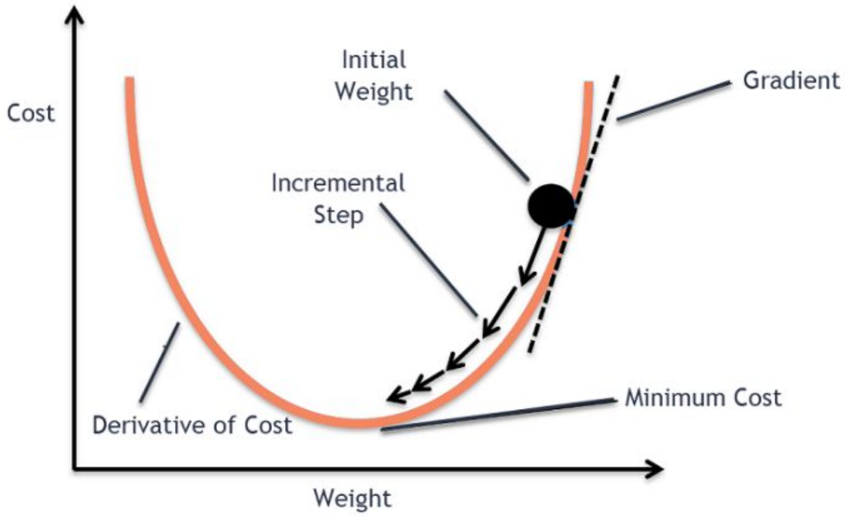
Gradient Descendent

Gradient descent is by far the most popular optimization strategy used in machine learning and deep learning at the moment

Training data helps these models learn over time, and the cost function within gradient descent specifically acts as a barometer, gauging its accuracy with each iteration of parameter updates. Until the function is close to or equal to zero, the model will continue to adjust its parameters to yield the smallest possible error. Once machine learning models are optimized for accuracy, they can be powerful tools for artificial intelligence

The starting point is just an arbitrary point for us to evaluate the performance. From that starting point, we will find the derivative (or slope), and from there, we can use a tangent line to observe the steepness of the slope. The slope will inform the updates to the parameters—i.e. the weights and bias. The slope at the starting point will be steeper, but as new parameters are generated, the steepness should gradually reduce until it reaches the lowest point on the curve, known as the **point of convergence**

The cost (or loss) function measures the difference, or error, between actual y and predicted y at its current position.



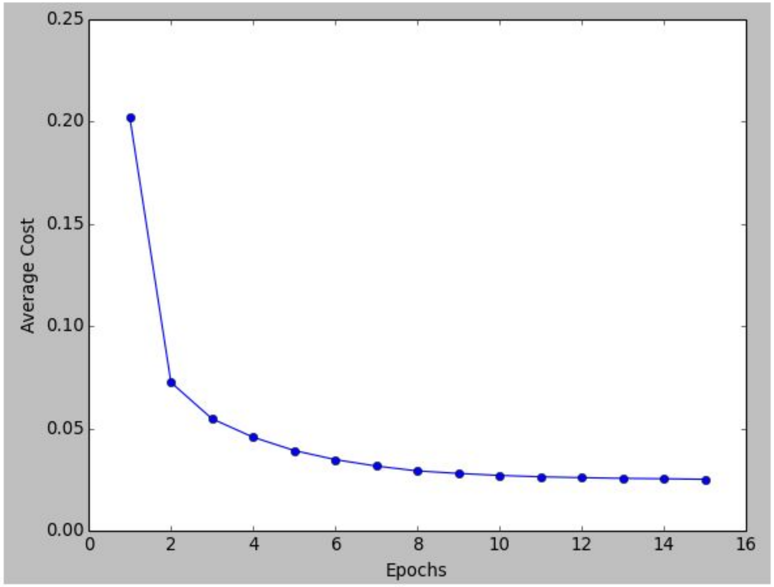
Types of Gradient Descent

Batch gradient descent

We take the average of the gradients of all the training examples and then use that mean gradient to update our parameters. So that's just one step of gradient descent in one epoch.

Batch Gradient Descent is great for convex or relatively smooth error manifolds. In this case, we move somewhat directly towards an optimum solution.

While this batching provides computation **efficiency**, it can still have a **long processing time for large training datasets** as it still needs to store all of the data into memory. Batch gradient descent also usually produces a stable error gradient and convergence, but sometimes that convergence point isn't the most ideal, finding the **local minimum** versus the global one.

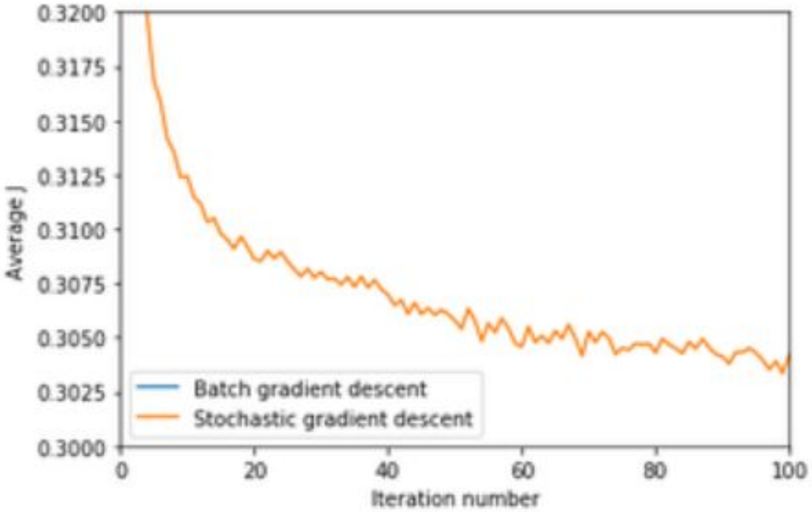


Types of Gradient Descent

Stochastic gradient descent

Stochastic gradient descent (SGD) runs a training epoch for **each example** within the dataset **and it updates each training example's parameters one at a time.** Since you only need to hold one training example, they are easier to store in memory. While these frequent updates can offer more detail and speed, it can result in **losses in computational efficiency** when compared to batch gradient descent. Its frequent updates can result in noisy gradients, but this can also be helpful in escaping the local minimum and finding the global one.

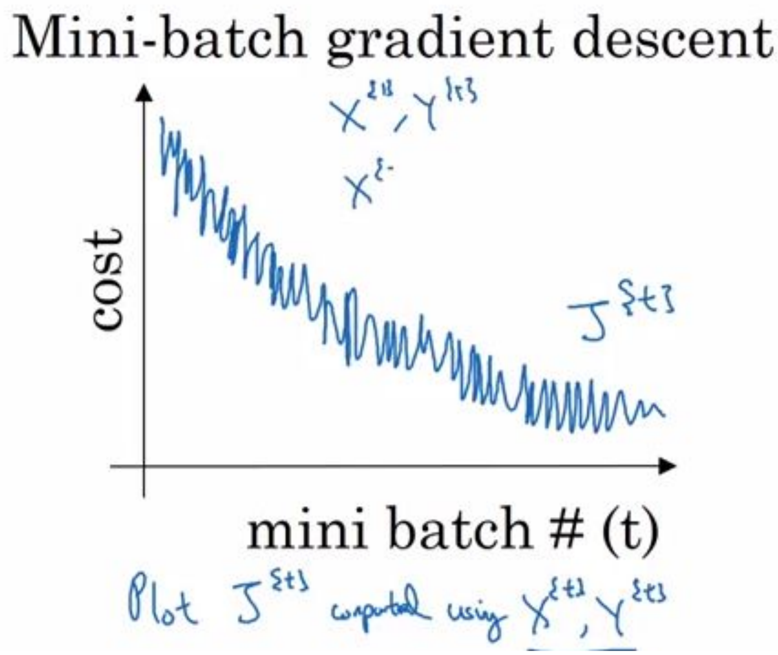
SGD can be used for larger datasets. It converges faster when the dataset is large as it causes updates to the parameters more frequently.



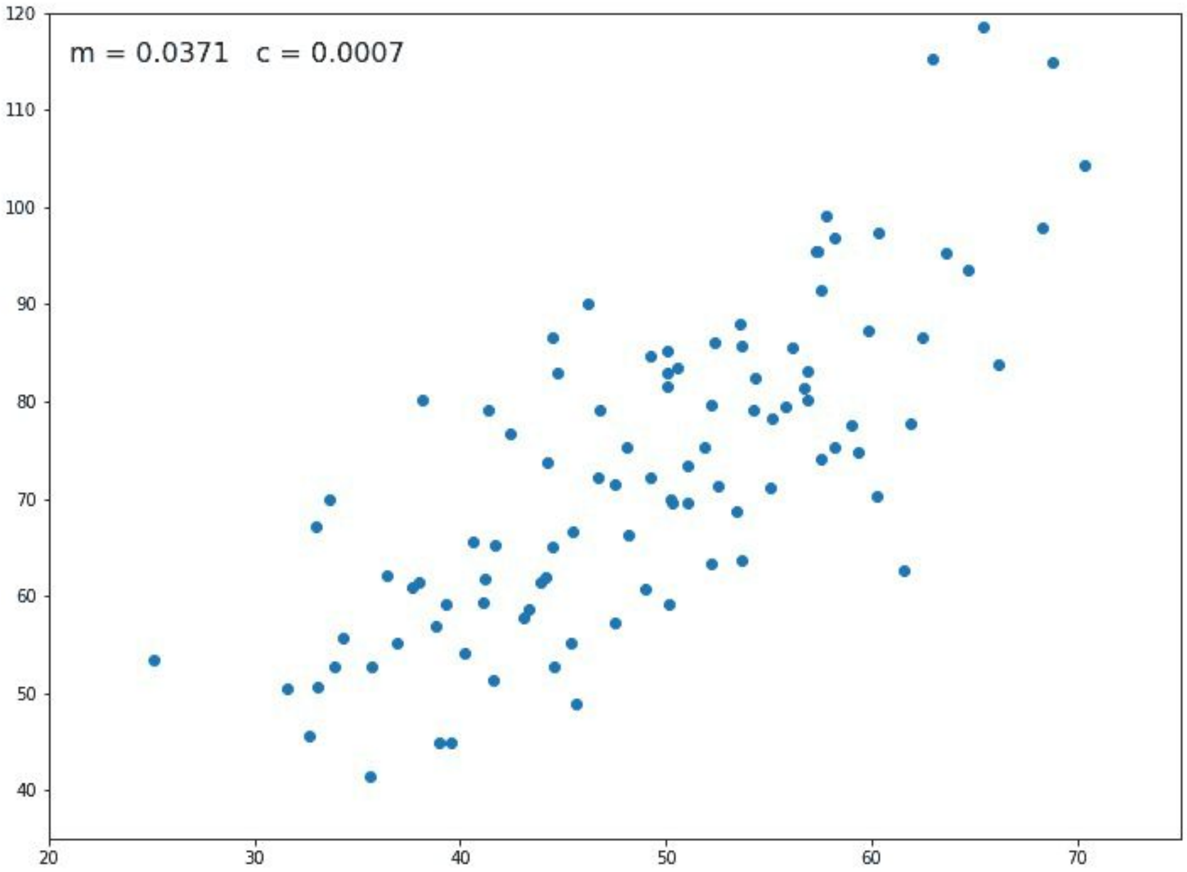
Types of Gradient Descent

Mini-batch gradient descent

Mini-batch gradient descent combines concepts from both batch gradient descent and stochastic gradient descent. **It splits the training dataset into small batch sizes and performs updates on each of those batches. This approach strikes a balance between the computational efficiency of batch gradient descent and the speed of stochastic gradient descent.**



Gradient Descendent and Linear Regression

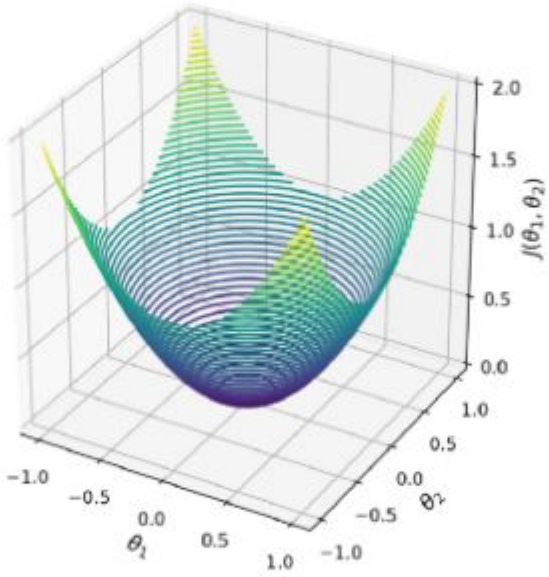


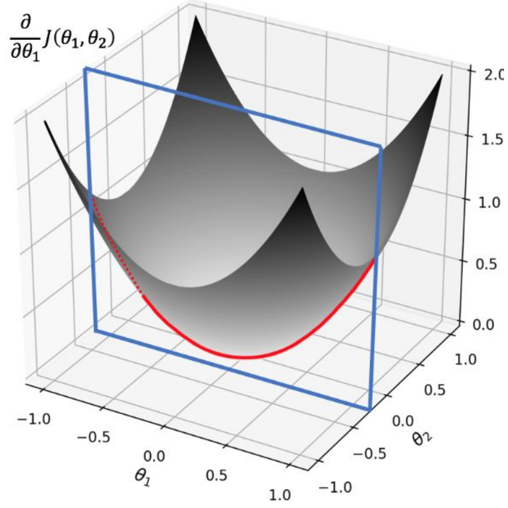
2D Gradient Descendent

Our aim here is to find the minimum of a function with more than one variable. When applying gradient descent to this function, our objective still remains the same, except that now we have two parameters, *θ1* and *θ2*, to optimise

Essentially, we cannot move both *θ1* and *θ2* at the same time when looking at a tangent. Therefore, we focus on only one variable at a time, whilst holding the other constant (partial derivative).

Fig.1a 3D plot for $J(\theta_1, \theta_2) = (\theta_1^2 + \theta_2^2)$





Group Discussion

What can be applications of the Gradient Descendent in your Capstone Project?

.

QUESTIONS?