MSE491: Application of Machine Learning in Mechatronic Systems

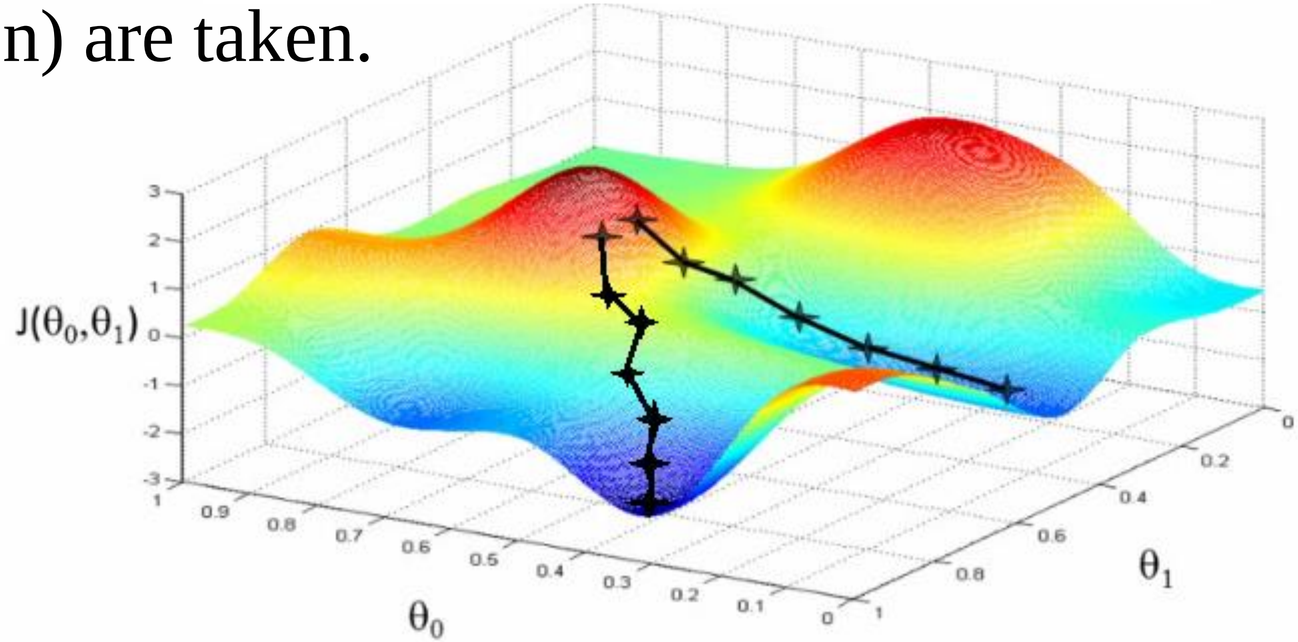
Types of Gradient Descent

Mohammad Narimani, *Ph.D., P.Eng.*
Lecturer
School of Mechatronic Systems Engineering
Simon Fraser University

Gradient descent

In gradient descent the step taken goes in the direction of the steepest descent/ascent, a direction given by the gradient of the function.

In the figure below, the journey starts at two locations on top of the hill and each two descent paths (minimization) are taken.



Gradient descent

-

– it will very slow

SFU SIMON FRASER UNIVERSITY SURREY

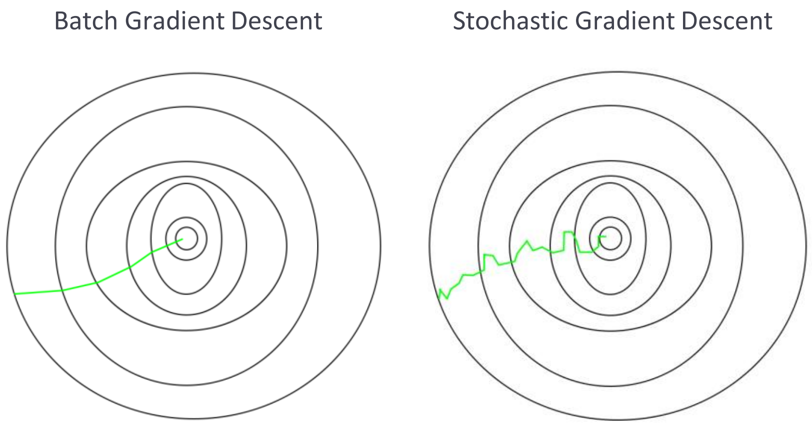
Types of Gradient Descent

- Typically, there are three types of Gradient Descent:
  - Batch Gradient Descent (discussed before)
  - Stochastic Gradient Descent (SGD)
  - Mini-batch Gradient Descent

Stochastic Gradient Descent (SGD)

- SGD uses only a single sample to perform each iteration, i.e., a batch size of one.

- The sample is randomly shuffled and selected for performing the iteration.



Taken from https://www.geeksforgeeks.org/

Sudo code for SGD:

Shuffle (m samples)

for i in range (m):

$$\theta_j = \theta_j - \alpha\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right).\, x_j^{(i)}$$

Stochastic Gradient Descent (SGD)

- Pros and cons:

Mini-batch Gradient Descent

- **Approach:**

Step 1: the whole set of training examples is divided to *t* groups (*t* mini-batches).

Step 2: Using a for loop, the GD formula is executed on all mini-batches (one mini-batch in each iteration), that is in each iteration the cost function is executed for only one mini-batch. Also, θ_j's are updated.

Step 3: When the process is executed for all mini-batches (the whole training set), one epoch is done.

Sudo code for mini-batch GD:

```
Shuffle (m samples)

Divide the training set to t mini-batches

for i in  range (t):      # t is the number of mini-batches
      - Compute the cost function J
      -Update θj′s
end
```