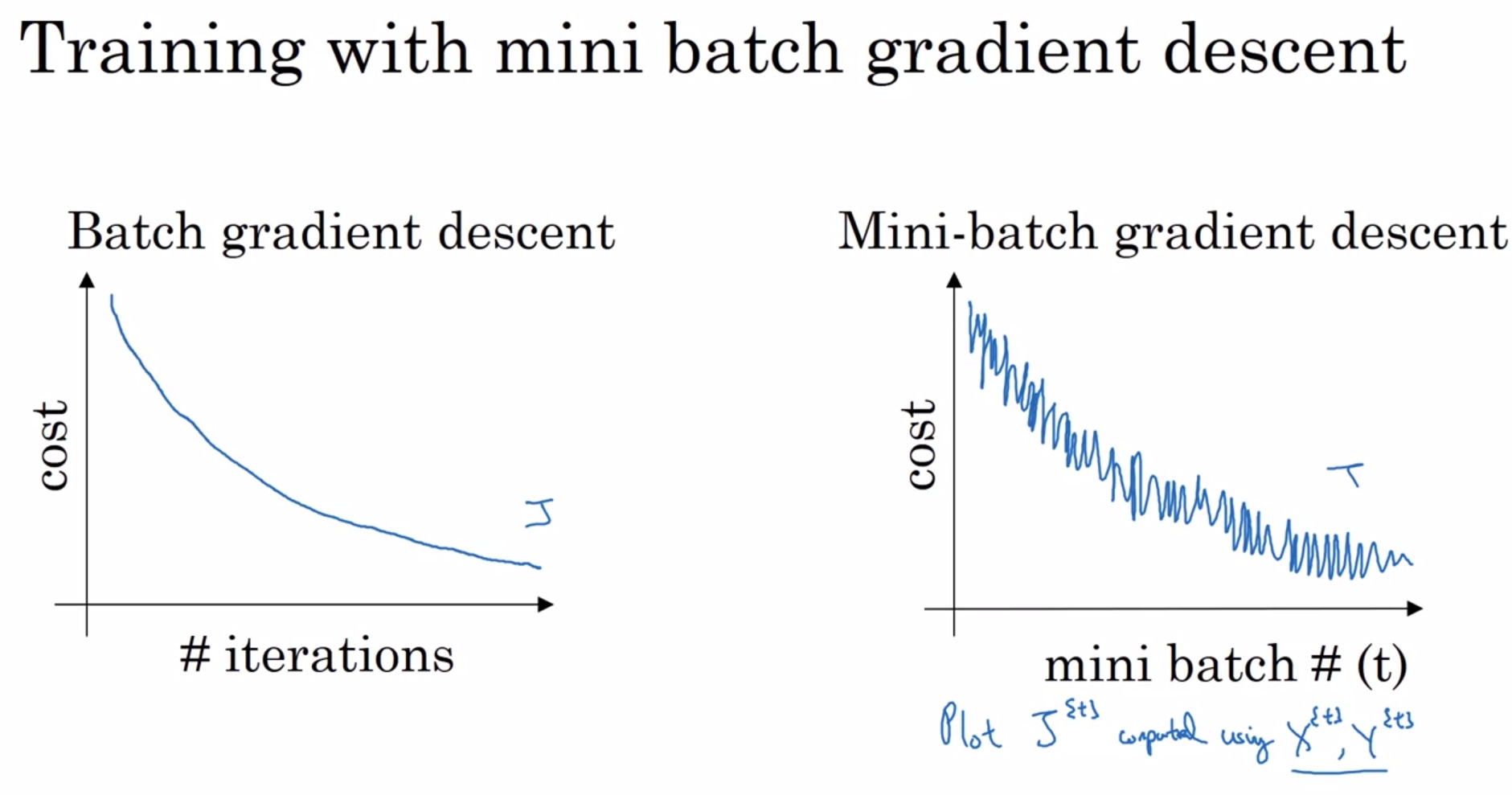
**6/1/2021**

**Optimization Algorithms**

**1/ Mini-batch Gradient Descent:**

* Epoch = a pass through a training (full or mini batch) set.
* If not doing mini batch, then the agent only takes 1 big step. By doing mini batches, the agents can take many mini steps in 1 big dataset.
* But again, you want to pass thru the training dataset multiple time.
* 1 iteration = 1 time pass thru the training dataset



**2/ Choosing your mini-batch size:**

* Say we have a dataset if m values.
* If the mini-batch has size = m, then we have the batch gradient descent => Take too long for 1 iteration
* If the mini-batch has the size = 1, then we have the stochastic gradient descent and each example is its own mini-batch.
* In practice, the mini batch should be in between 1 & m.