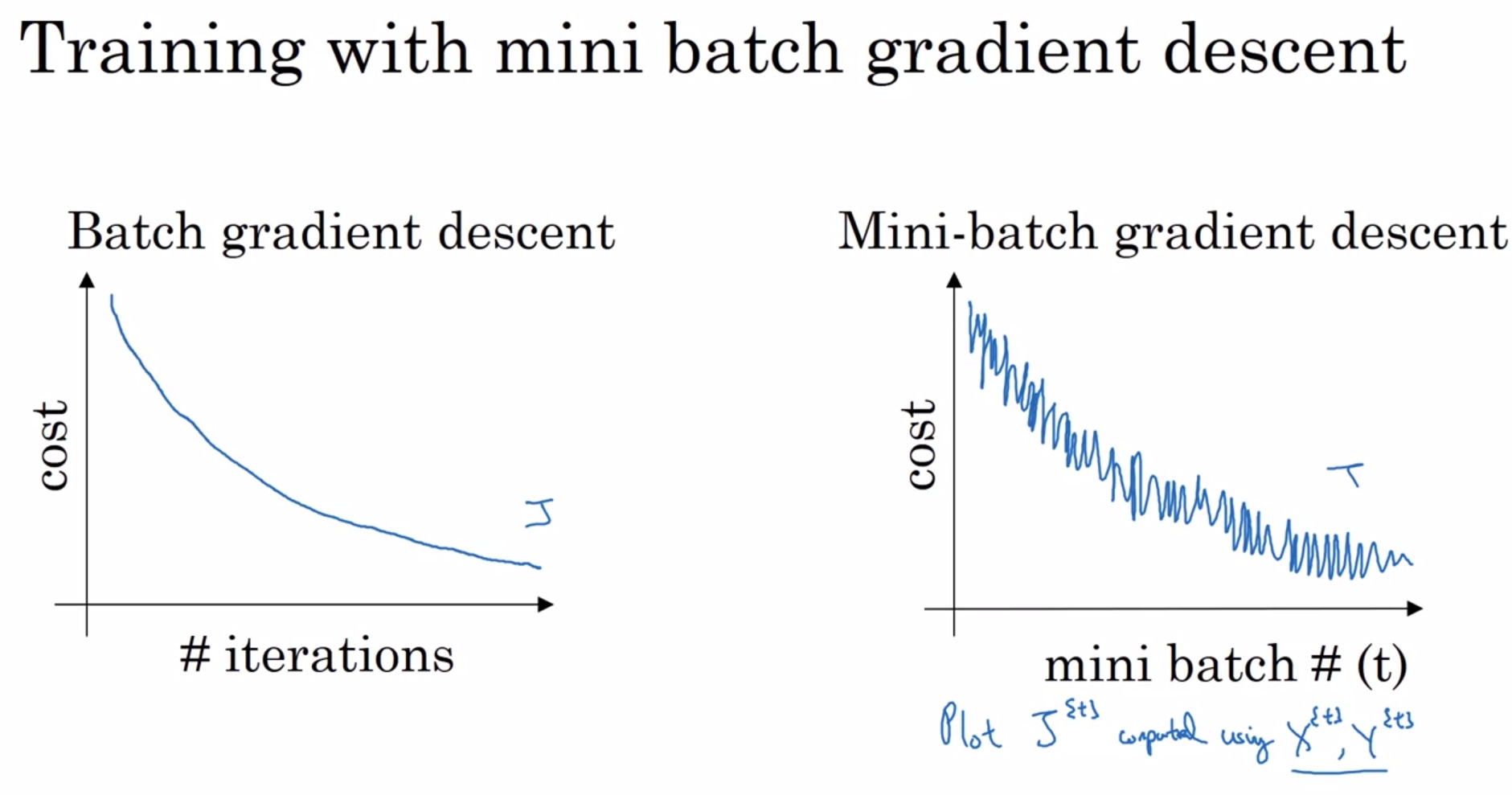
**6/1/2021**

**Optimization Algorithms**

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

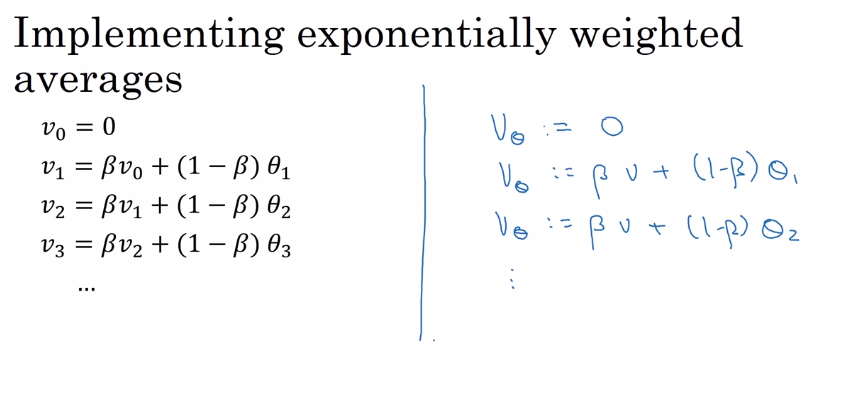
* Epoch = a pass through a training (full or mini batch) set.
* If not doing mini batch, then the agent only takes 1 big step. By doing mini batches, the agents can take many mini steps in 1 big dataset.
* But again, you want to pass thru the training dataset multiple time.
* 1 iteration = 1 time pass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



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

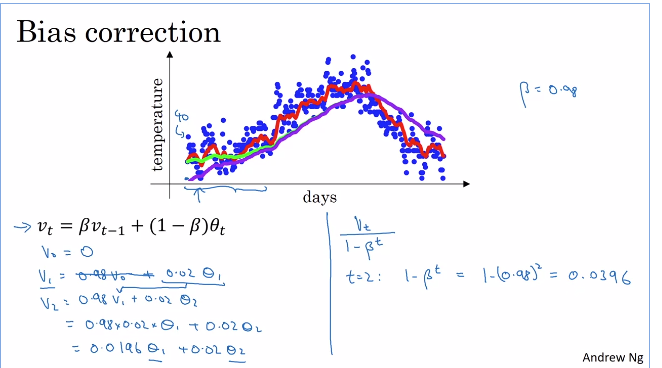
* Say we have a dataset if m values.
* If the mini batch has size = m, then we have the batch gradient descent => Take too long for 1 iteration
* If the mini-batch has the size = 1, then we have the stochastic gradient descent, and each example is its own mini-batch.
* In practice, the mini batch should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

**3/ Exponentially weighted averages:**



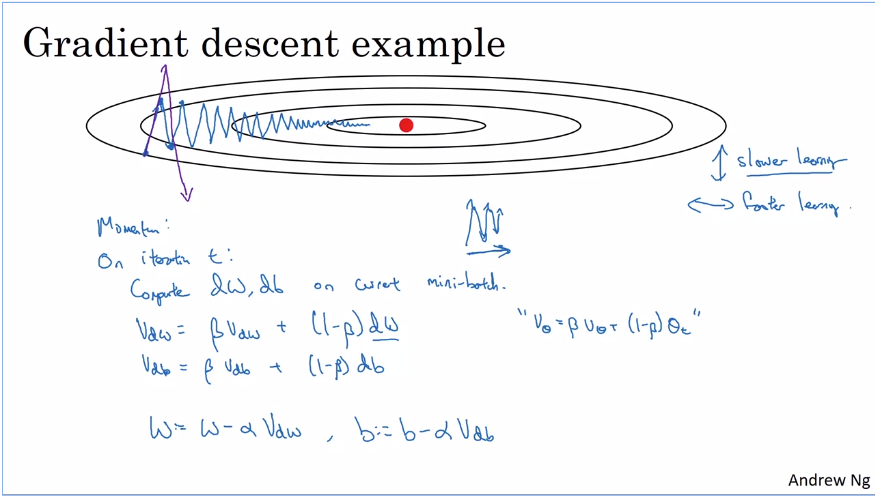
**4/ Bias Correction:**

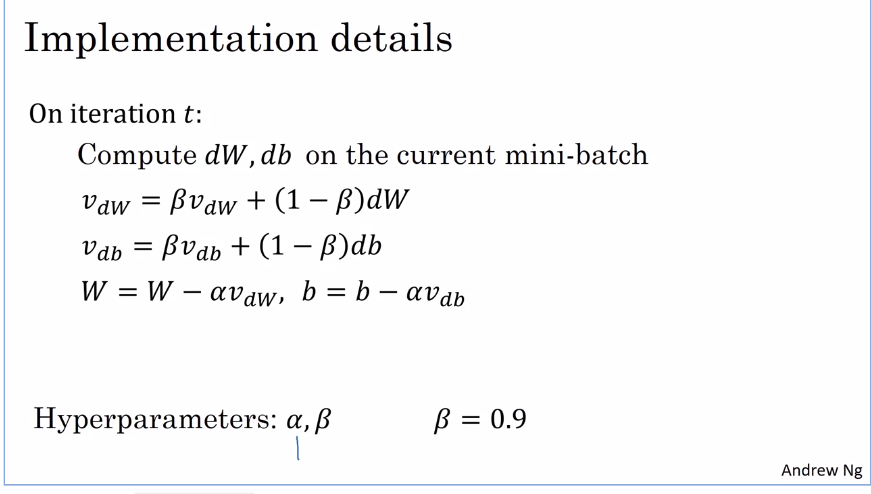
* Helps with Exponentially Weighted Average better.



* Help you start from the purple line to green line

**5/ Gradient Descent with Momentum**

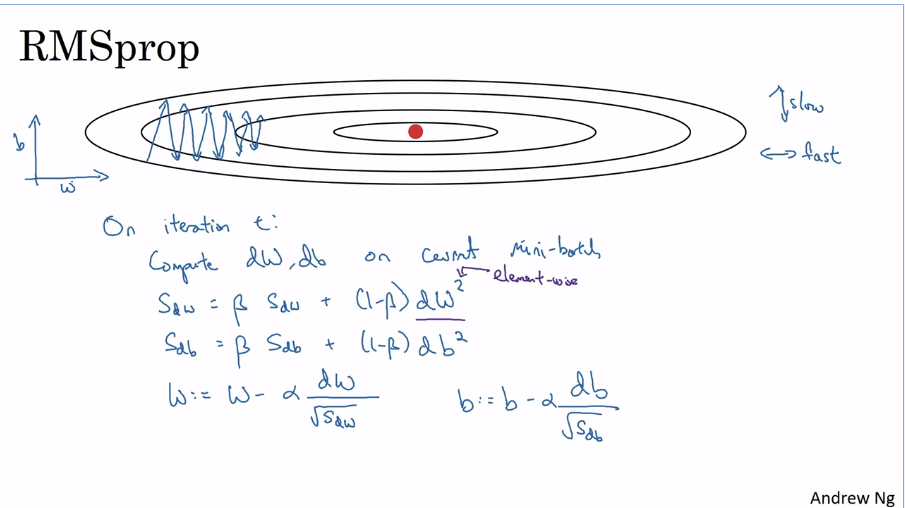




* Momentum can speed up Gradient Descent

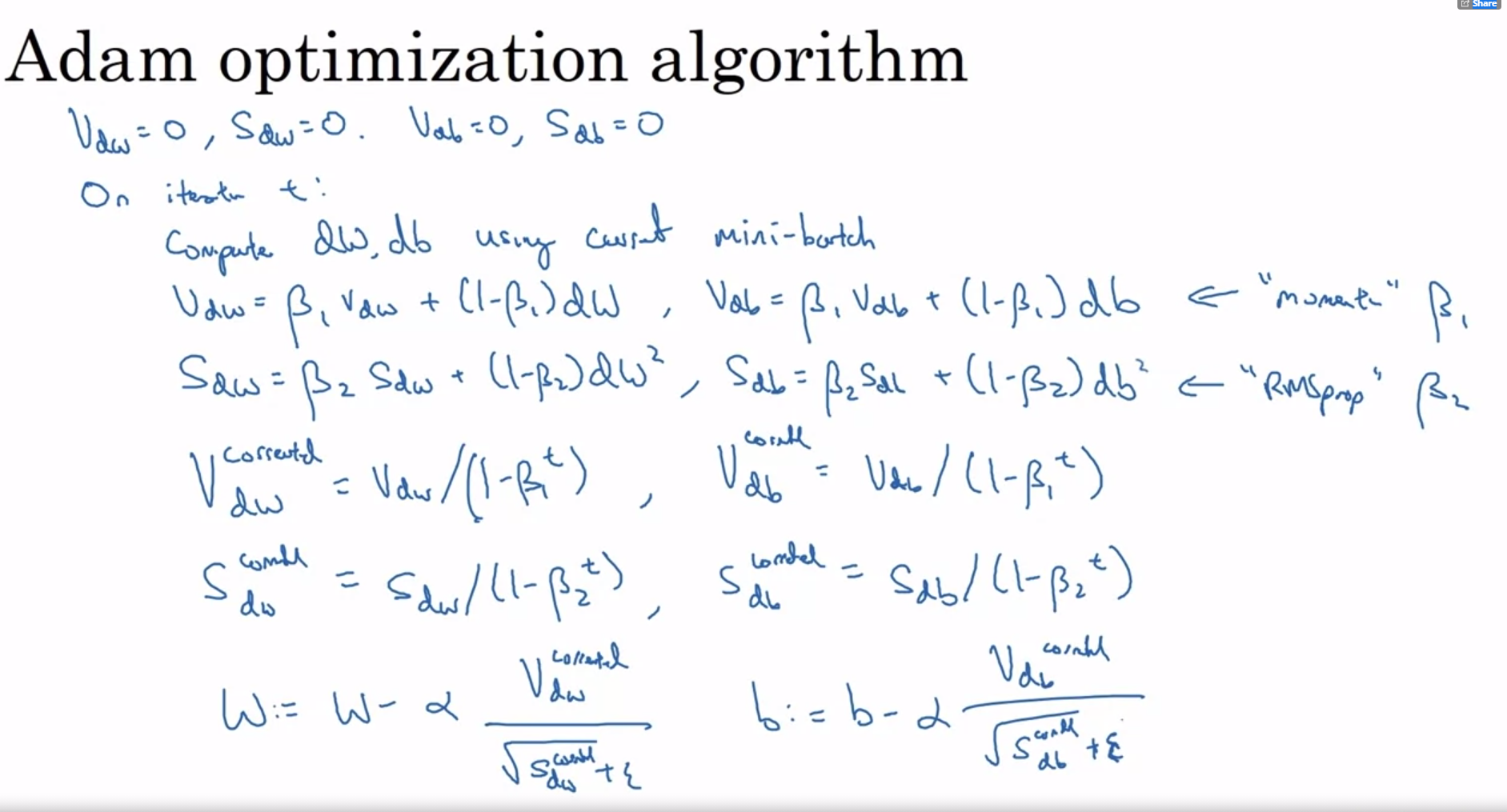
**6/ RMSprop – Root mean square prop – can also speed up gradient descent**

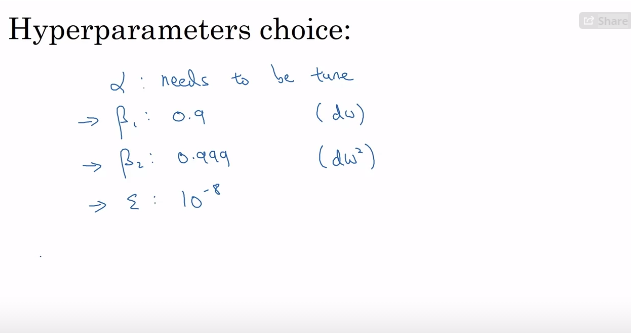
* You want to speed up W but slow down b => RMSprop



**7/ Adam optimization algorithm:**

* Momentum + RMSProp

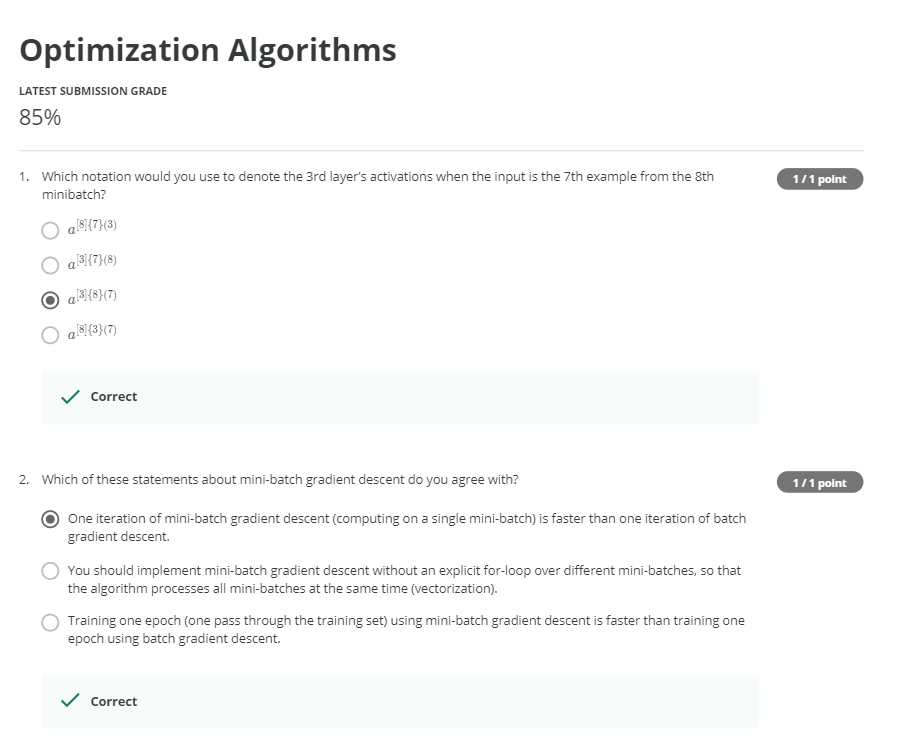


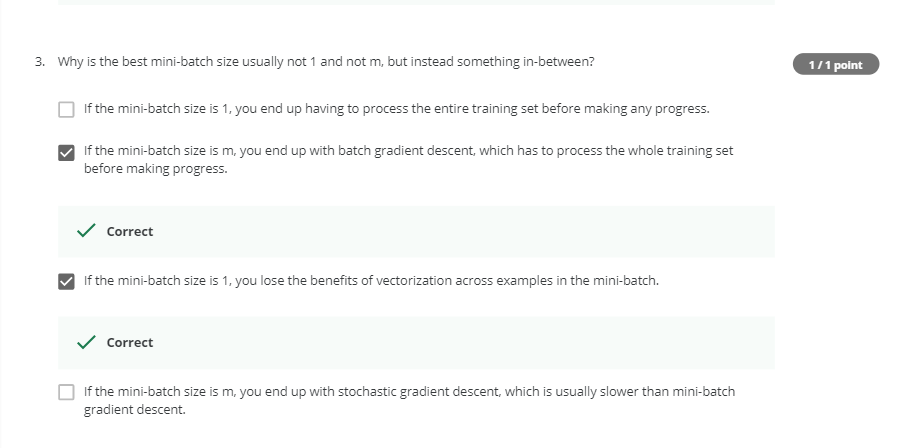


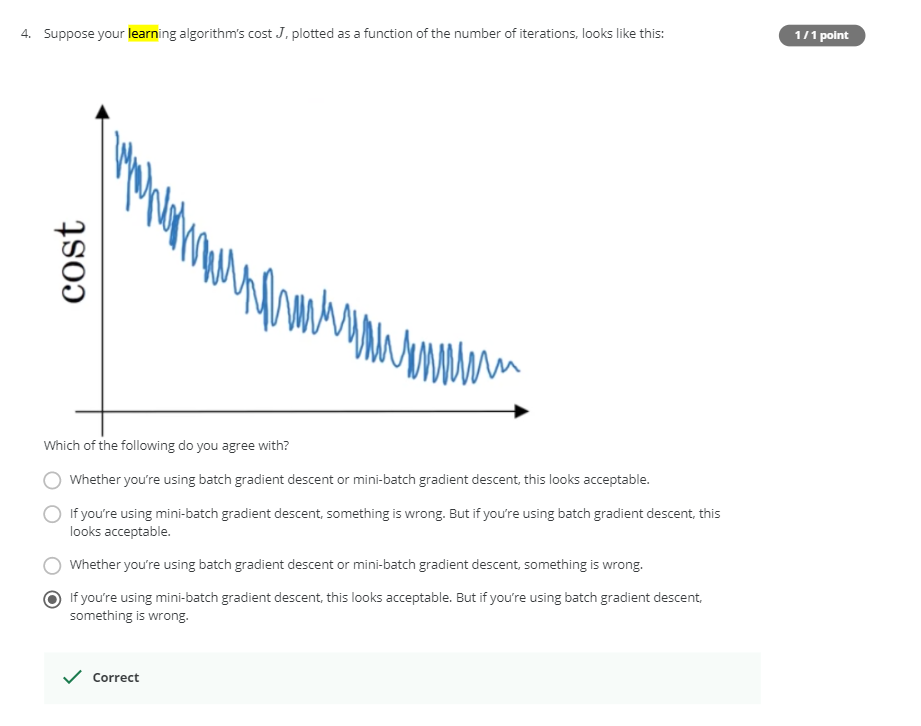
**8/ Speed up learning algorithms with Learning Rate Decay:**

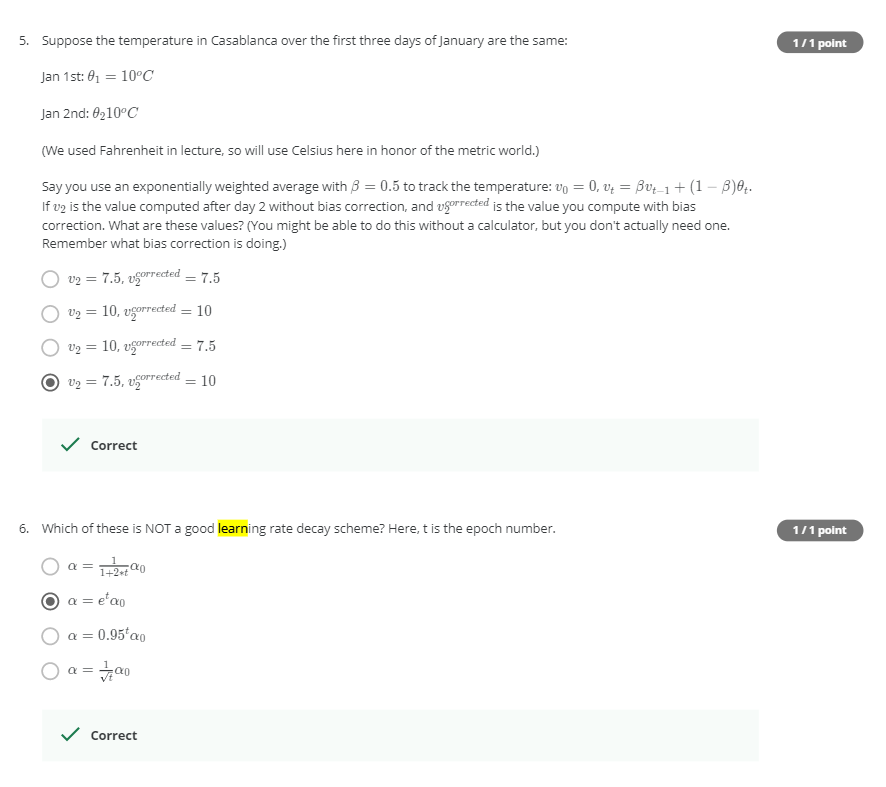
* 1 Epoch = 1 pass thru the data training set

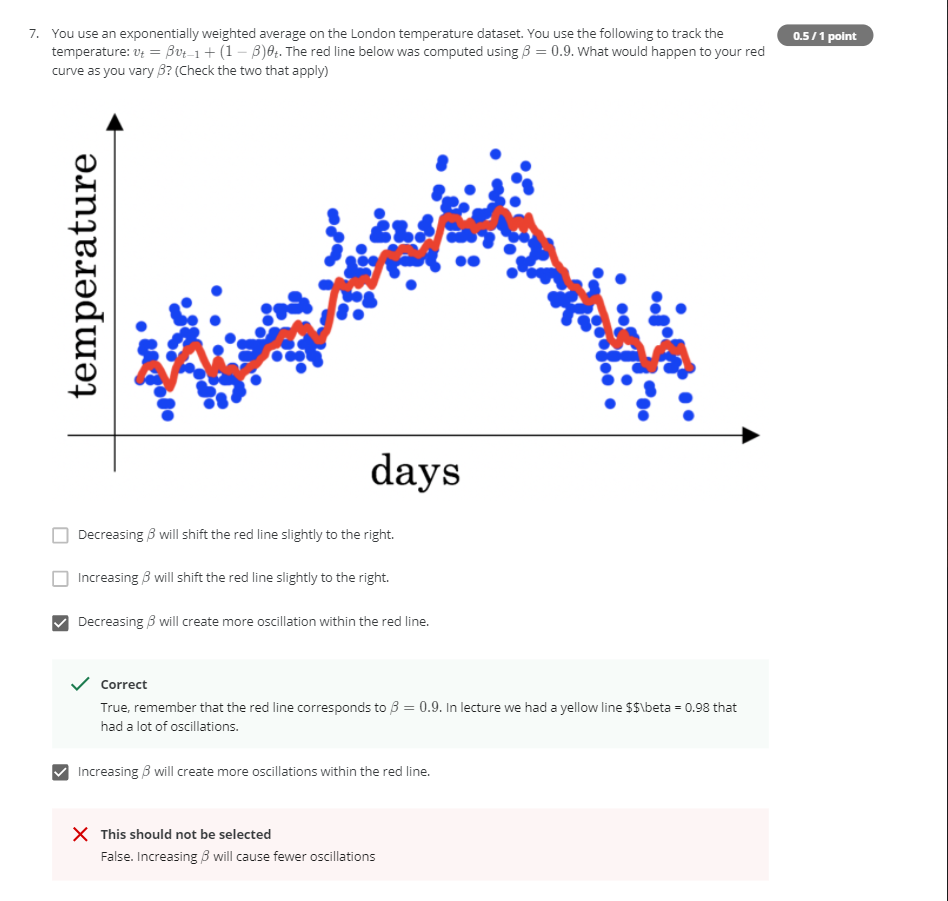
**9/ The Problem of Local Optima**

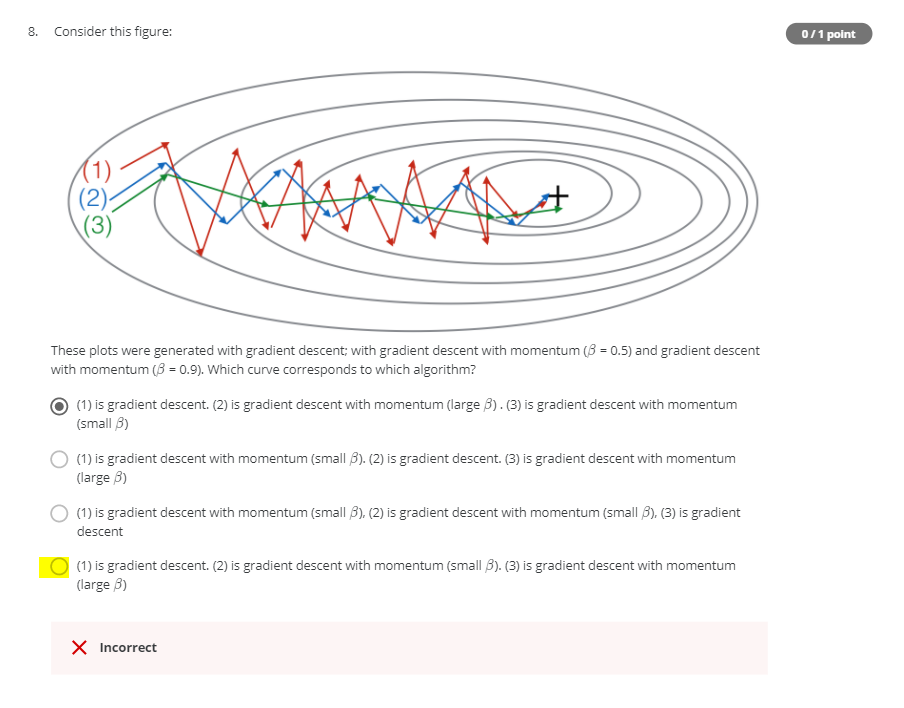


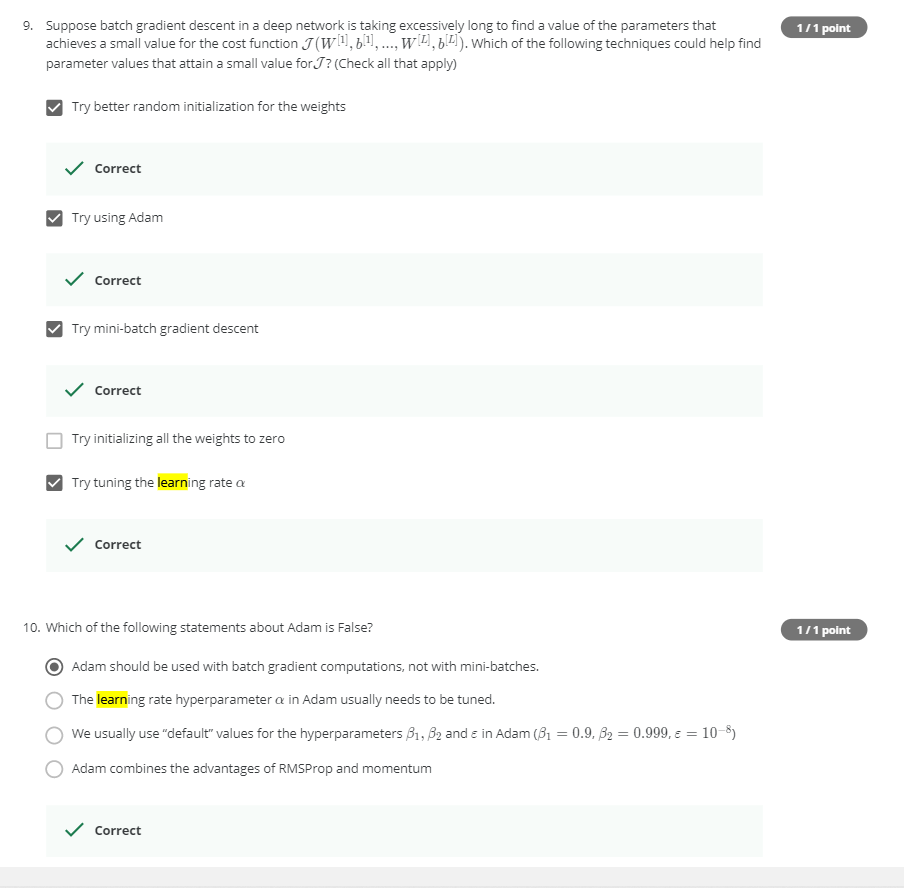












**1/ Batch GD vs SGD:**

