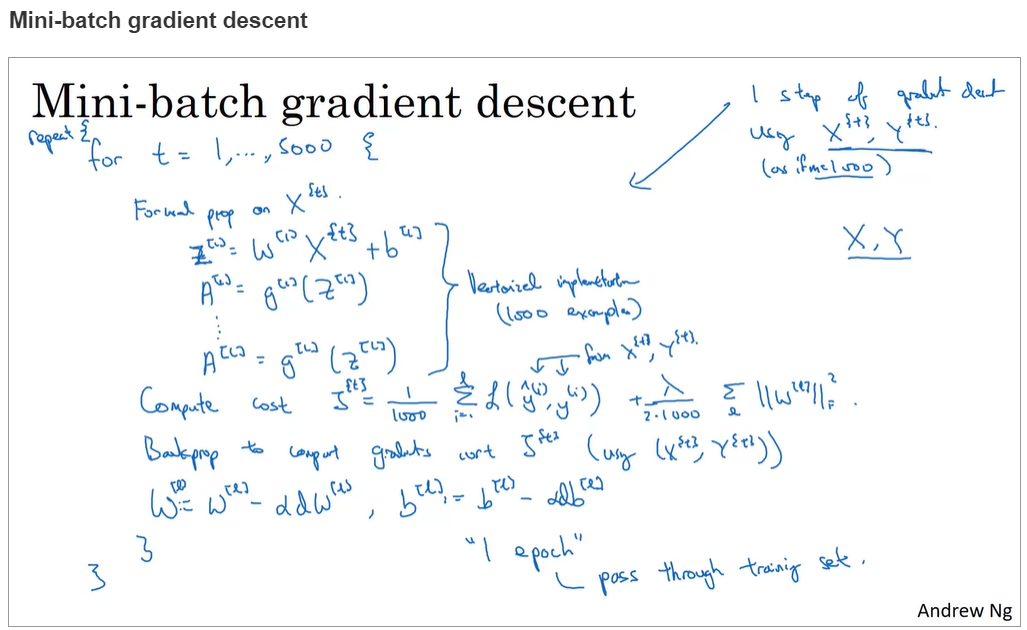
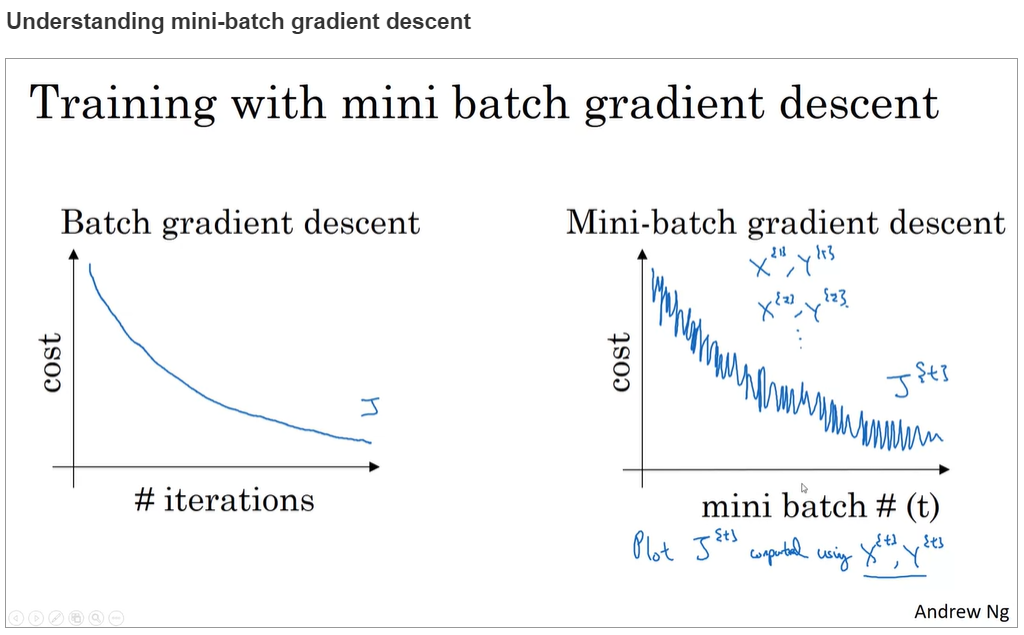
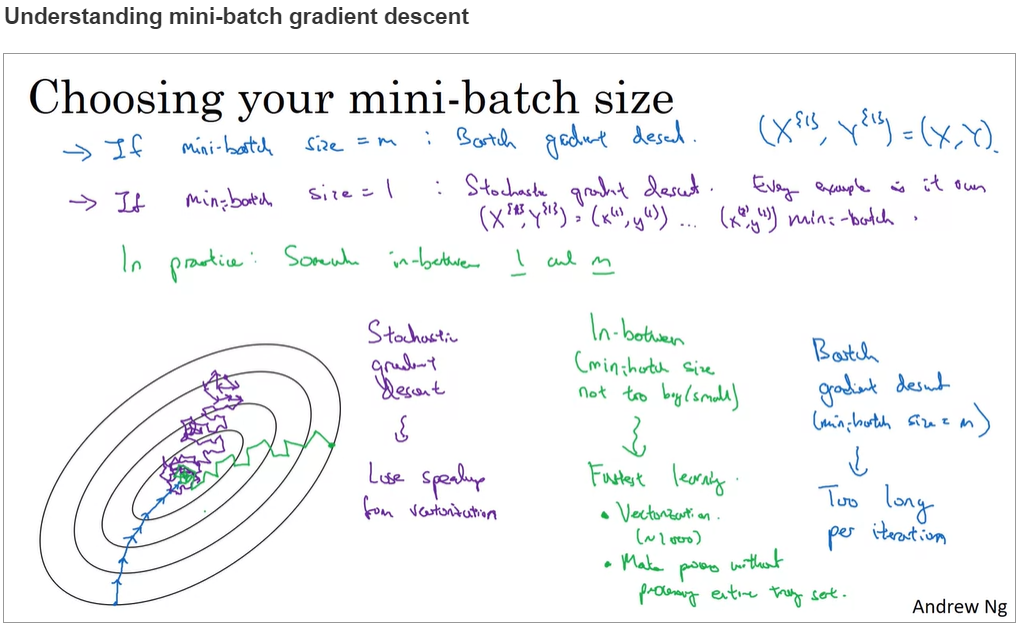


* **x(i)** 🡪 **i th traning example**
* **z[L] 🡪 z for L th layer**
* **X{t} , Y{t}** 🡪 t th mini batch X{t}.shape = (nx ,size of minibatch) Y{t} .shape = (1, size of minibatch)



* **1000 🡪 minibatch size**
* **5000 🡪 number of minibatches**





* **When we use batch training that is size of minibatch = m then the gradient descent takes relatively big steps and relatively low noise and we just head towards minimum.**
* **When we use Stochostic gradient descent that is batch size = 1 it takes relatively small steps and may be extremely noisy and may lead to wrong direction also it doesn’t necessarily converges to minimum it will oscillate around the region around the minimum.**
* **In between size would be fast little noisy also this might not converge to minimum but it will oscillate in very minimum region around minimum.**

