# LARGE SCALE MACHINE LEARNING

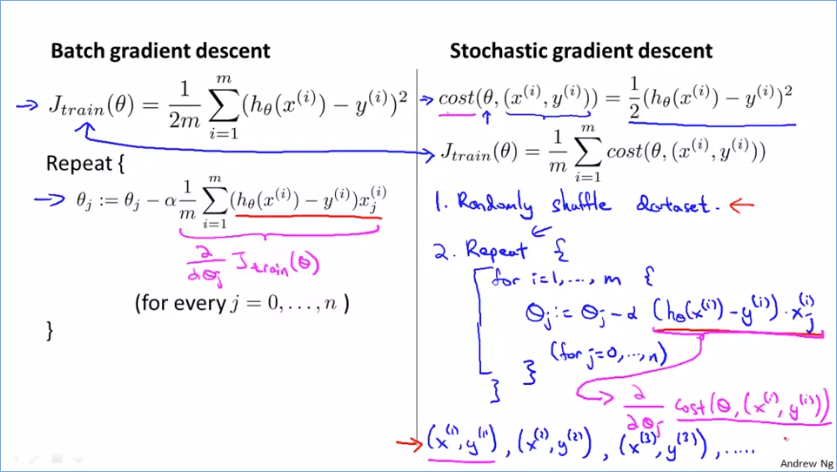
## Gradient Descent With Large Datasets

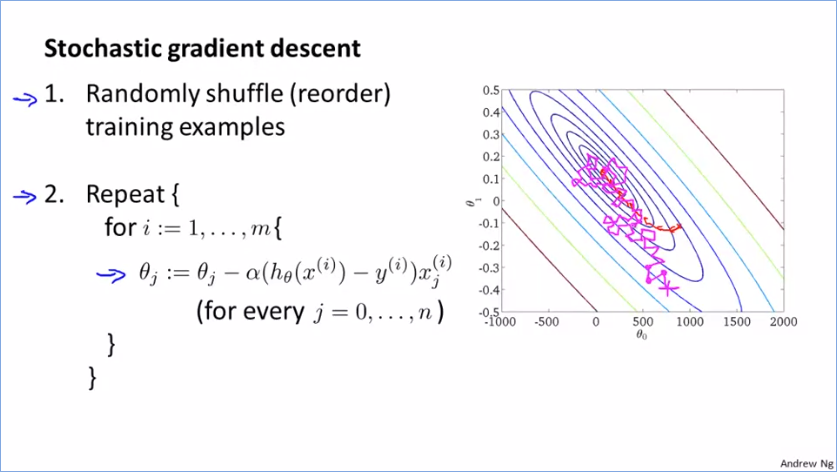
Learning With Large Datasets

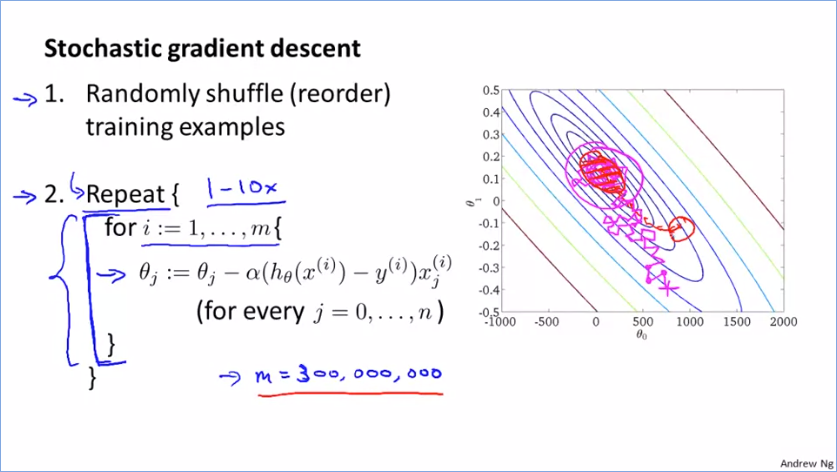
* Plot learning curve for range of values of m.

Stochastic Gradient Descent

* Batch vs. stochastic

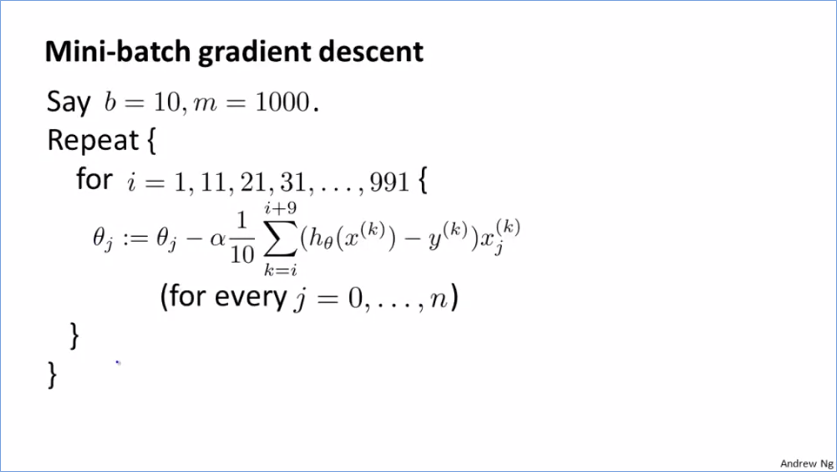




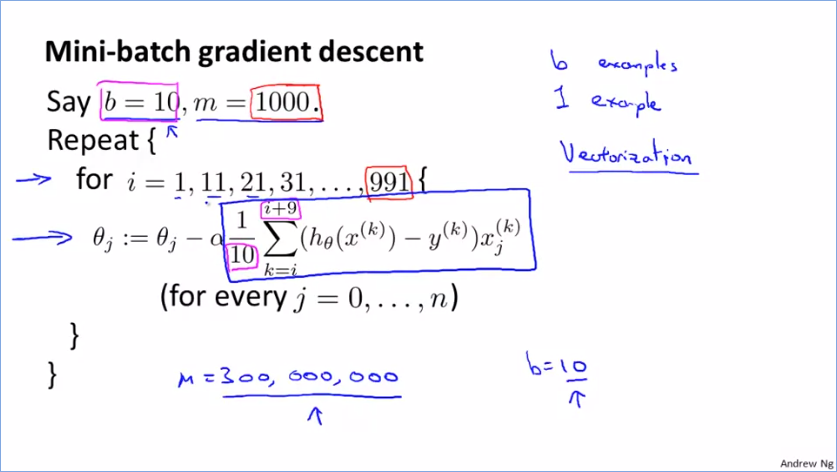


Mini-Batch Gradient Descent

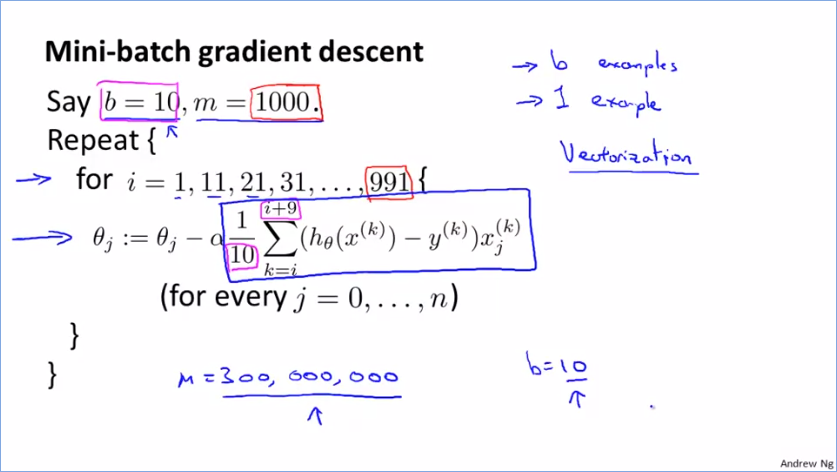
* Algorithm:



* Faster progress than batch gd:

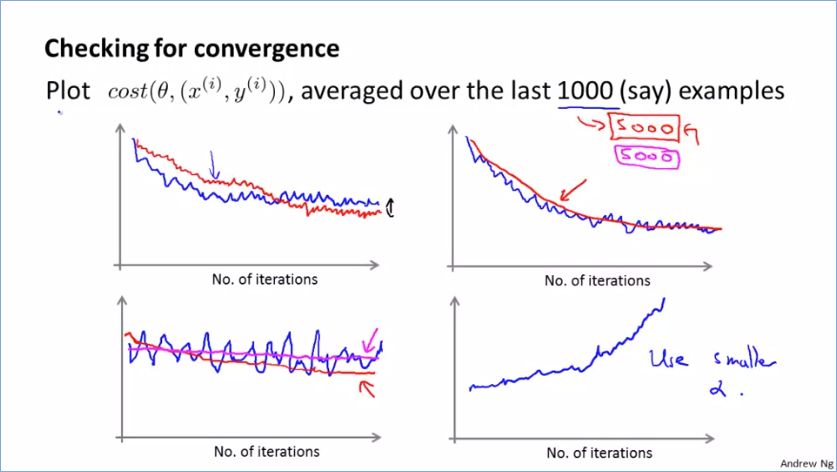


* Outperforms stochastic with good vectorized implementation
* Disadvantage: extra parameter b

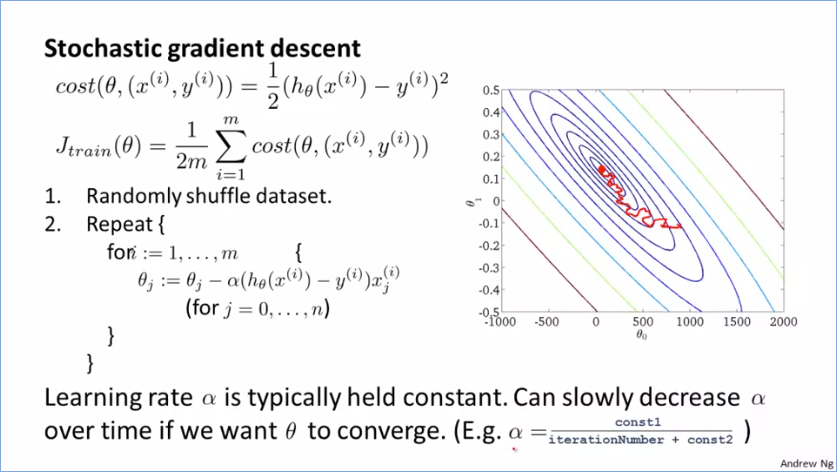


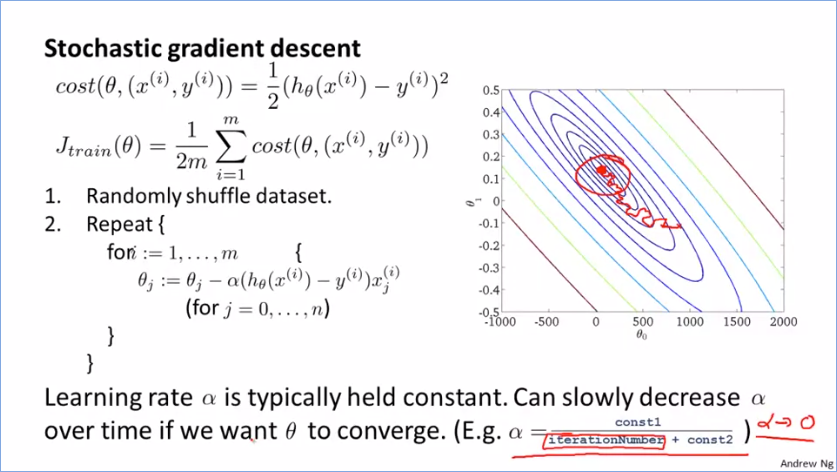
Stochastic Gradient Descent Convergence

* Compute cost associated with example right before updating θ, plot average over every last 1000 examples or so
* Learning rate α
* Number of examples to average over



* More common to keep α constant, but can decrease it to converge





## Advanced Topics

Online Learning

* Similar to stochastic gd algorithm
* Prices
* Product search - predicted click-through rate CTR
* Collaborative filtering 🡪 features LR classifier
* Enough data to need not store fixed training sets
* Adaptation

Map Reduce and Data Parallelism

* Can algorithm be expressed as computing sums of functions over training set?
* Hadoop open-source