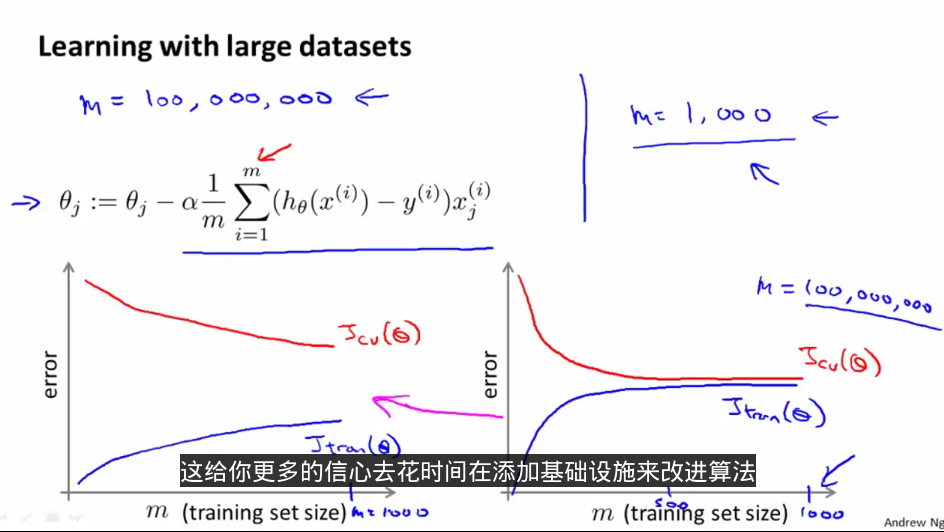
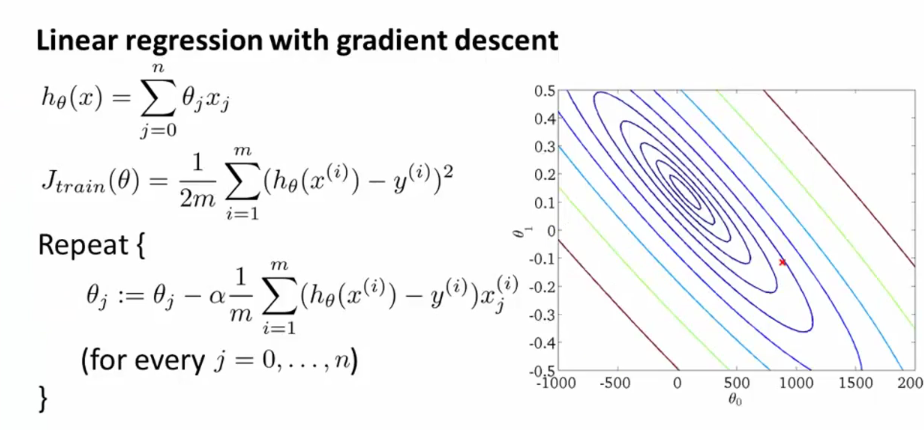
1. 大规模数据训练。模型在数据较少的时候呈现过拟合，这就需要增加数据。



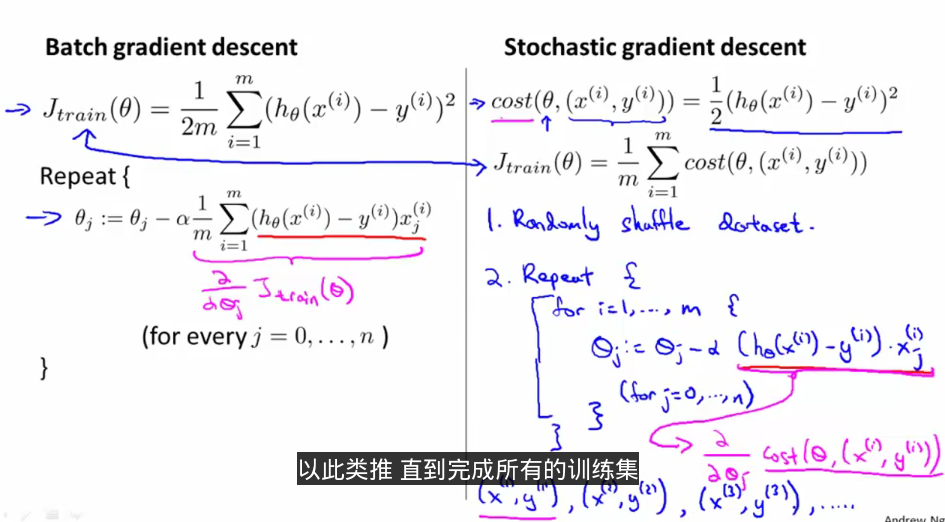
1. 随机梯度下降（Stochastic Gradient Descent）:用在大量的数据集上。

传统的梯度下降方法：

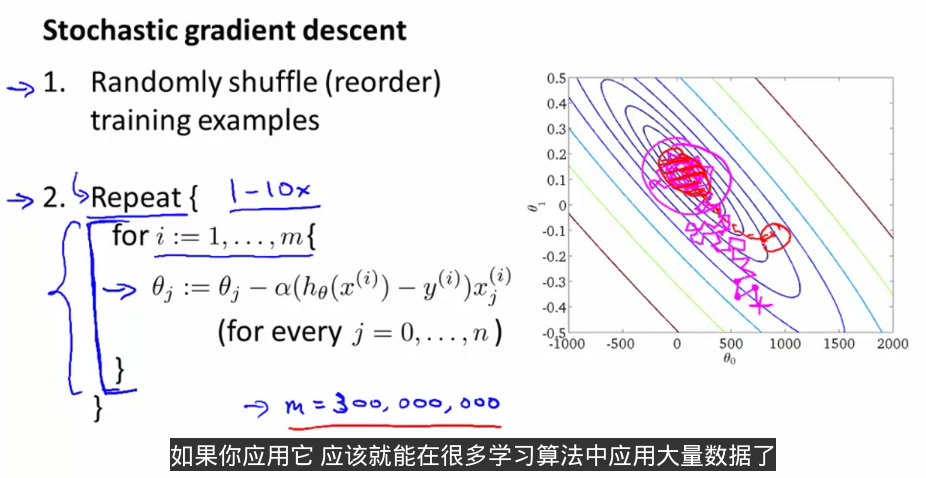


假如m特别大，梯度下降方法很慢。随机梯度下降步骤：

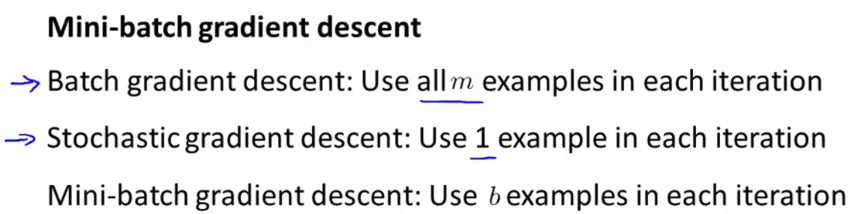
1. 随机打乱数据集
2. 进行梯度下降(不需要对所有训练样本求和，只需要一个样本)

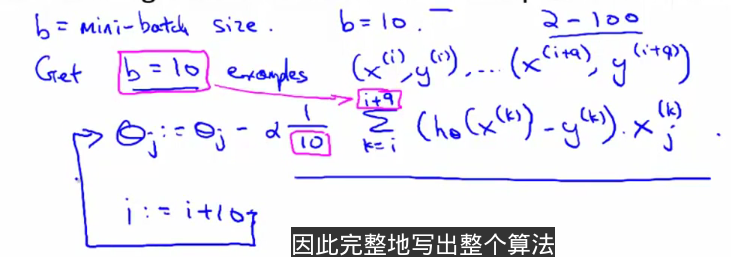


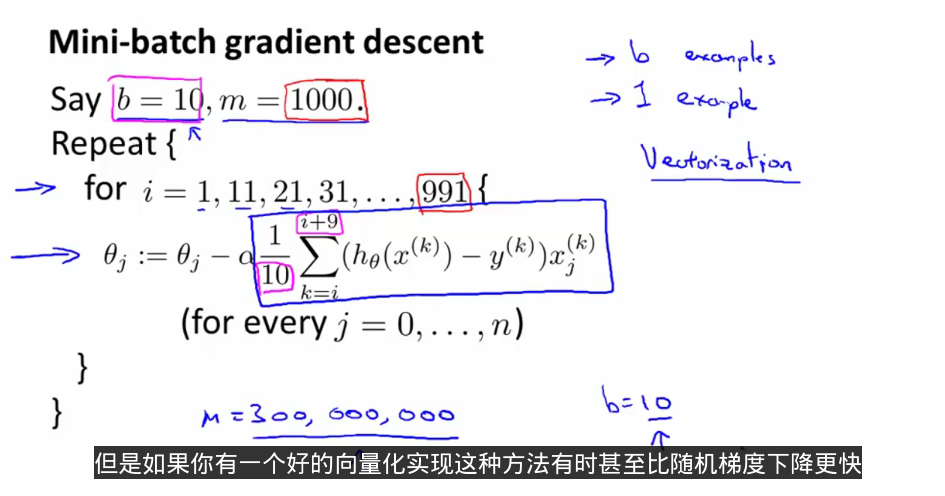
随机梯度下降方法不是每次选择梯度最大的地方不断逼近全局最小值，而是方向随机，但大体方向是向着全局最小，最后会在全局最小附近徘徊。



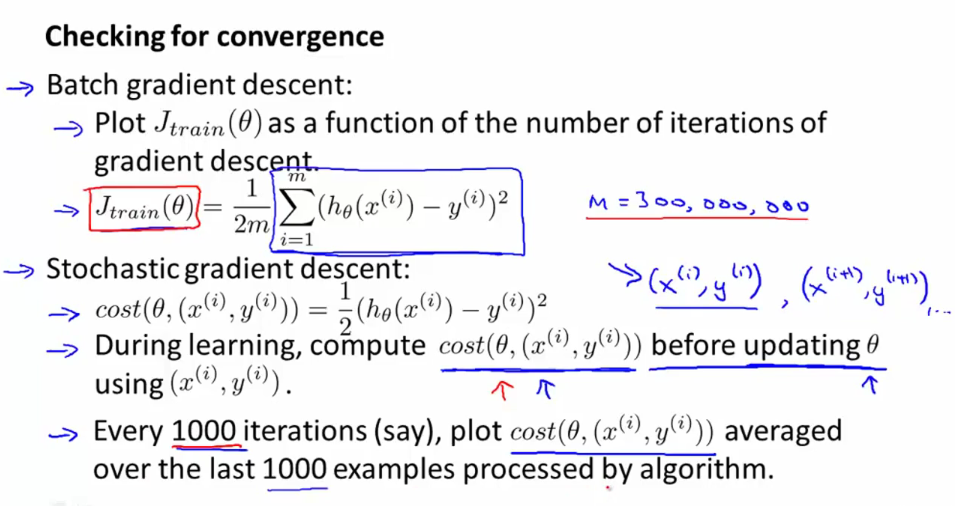
使用一个样本来更新梯度似乎显得非常随机，而选择传统的梯度下降方法的全部求和（批量梯度下降，Batch Gradient Descent）又太慢，选择进行折中，选择b（b一般在2-100范围内）个样本进行梯度更新，这就是小批量梯度下降（Mini-Batch）：

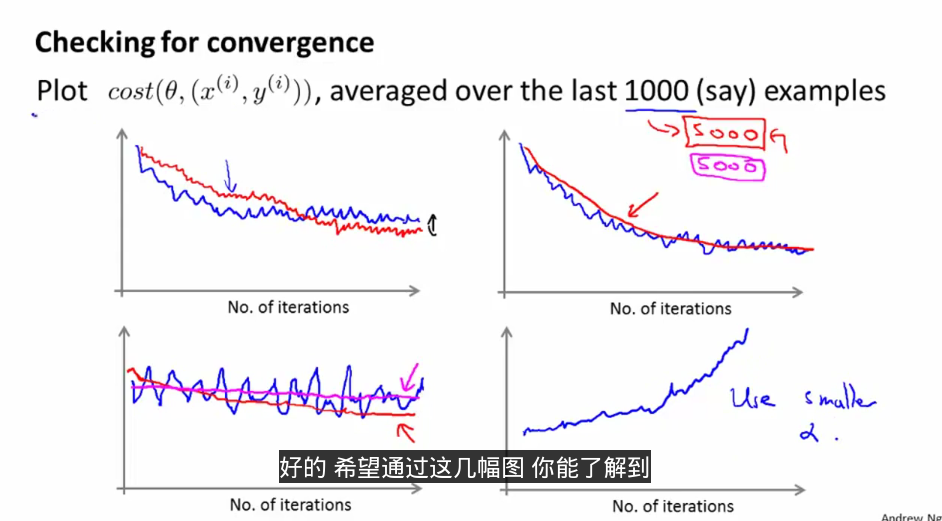




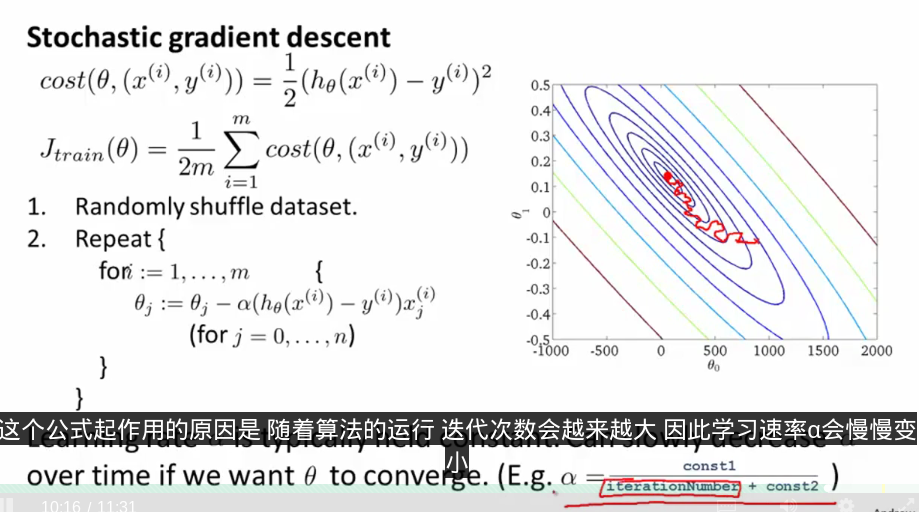


随机梯度下降的收敛问题：

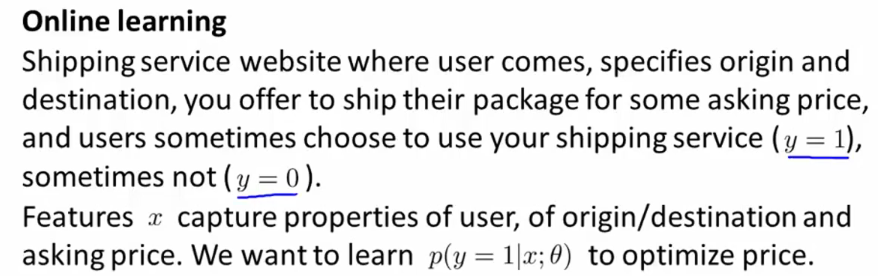


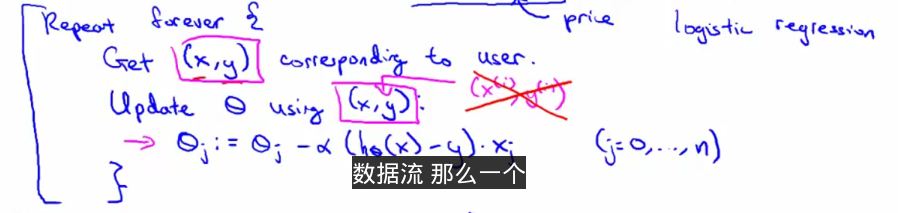


随机梯度下降会在全局最小值附近不断徘徊，如果想让他达到全局最小，可以通过不断更新学习速率α的方式，这样在接近全局最小的时候震荡幅度会变小，不断逼近全局最小。

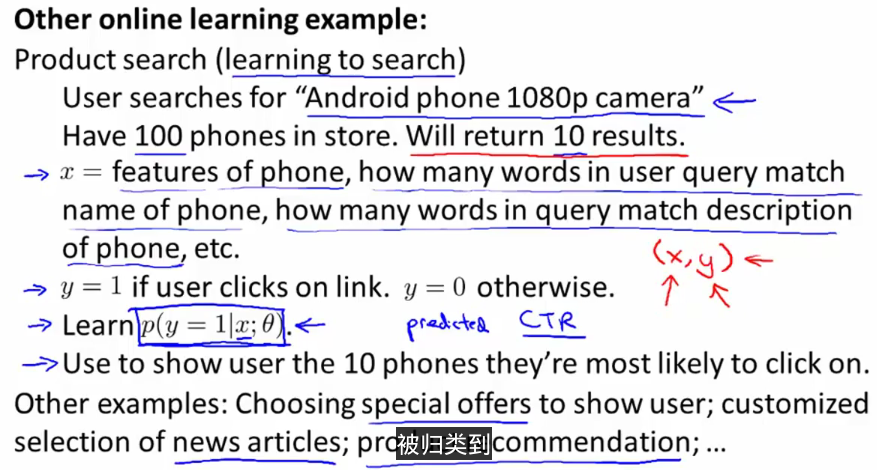


在线学习（Online Learning）:

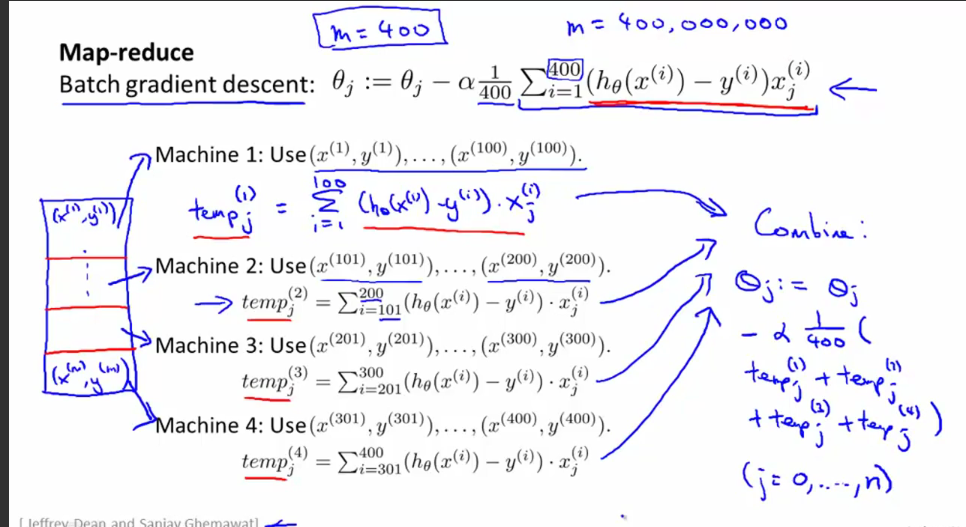


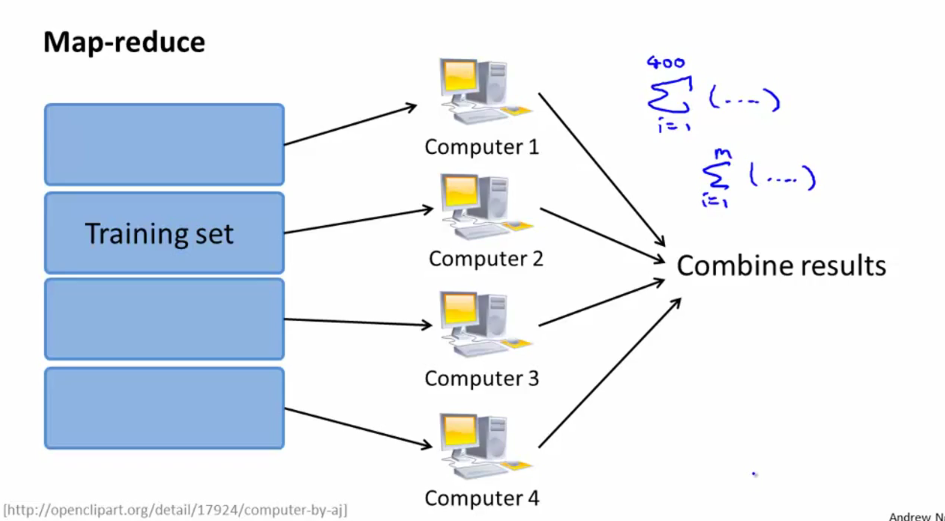


一个例子：



Map Reduce：



、