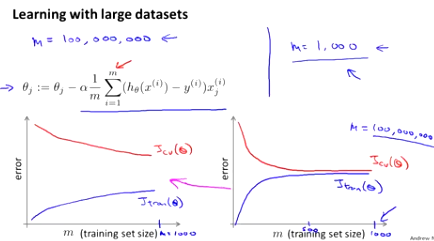
# 大规模机器学习

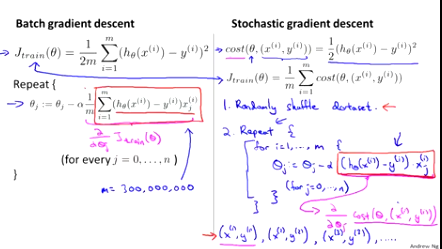
## 大数据的机器学习

大数据的模型构建需要花费大量的时间，在进行大数据的构建模型之前，需要知道模型是处于高偏差或高方差模型。

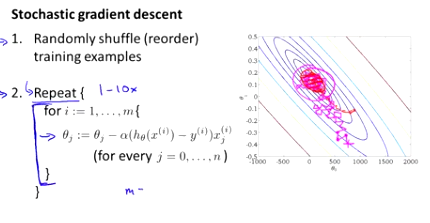


## 随机梯度下降算法

梯度下降算法计算大数据集的计算量非常大，耗时较久，因此采用随机梯度下降算法可以减少运算时间。



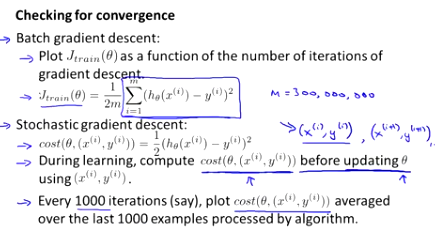
随机梯度下降算法步骤：



### 2.1 收敛性

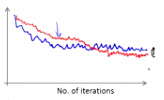
批处理梯度下降算法每次参数更新都需要计算整个训练集，随机梯度下降算法参数更新只需要计算一个训练样本，每次就立刻进步一点点，不需要在算法当中时不时地扫描一遍全部的训练集来计算整个训练集的代价函数。

批处理梯度下降算法和随机梯度下降算法的损失函数：

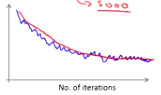


随机梯度下降算法的损失函数是每1000次迭代运算中，对最后1000个样本的cost值求平均然后画出来，通过观察这些画出来的图，我们就能检查出随机梯度下降是否在收敛。

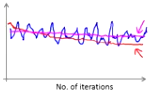




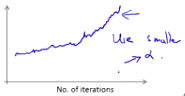
如上图所示，蓝色图为每次迭代计算最后1000组样本中损失的平均值，红色线是更小的学习速率α对应的损失曲线图。随机梯度下降不是直接收敛到全局最小值，而是在局部最小附近反复振荡，所以使用一个更小的学习速率，最终的振荡就会更小。



上图蓝色线是1000组样本的平均值，红色线是5000组样本的平均值。

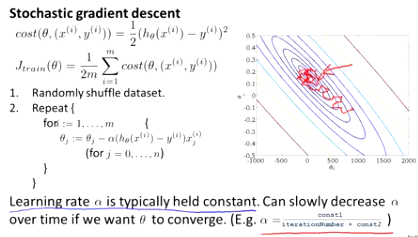


蓝色线表示1000组样本平均值的损失函数，红色线表示5000组样本平均值的损失函数，若样本损失函数变化的较缓慢，那么需要调整学习速率，或者改变特征变量或者改变其他的什么。



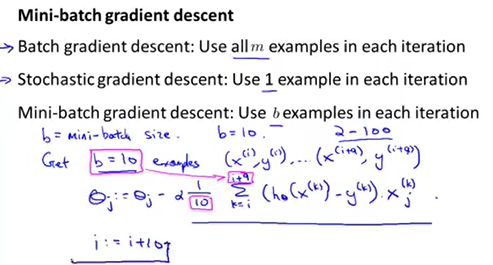
如上图，若损失函数曲线在发散，则使用更小的学习速率α。

随机梯度下降算法可以通过缓慢的较小α来使损失函数收敛。

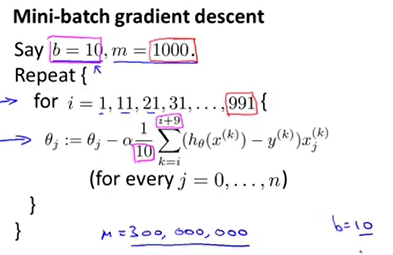


随机梯度下降算法会在最小值附近振荡，但当它越来越靠近最小值的时候，由于见笑了学习速率，因此这个振荡也会越来越小，知道落到几乎靠近全局最小的地方。上图学习速率α的公式是随着算法的运行，迭代次数会越来越大，因此学习速率α会慢慢变小，因此你的每一步就会越来越小，直到最终收敛到全局最小值。

## 小批量梯度下降算法

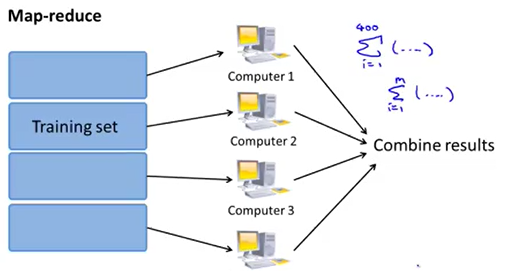


小批量梯度下降算法步骤：

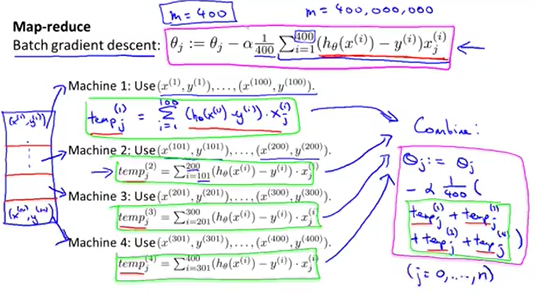


### 映射约减算法

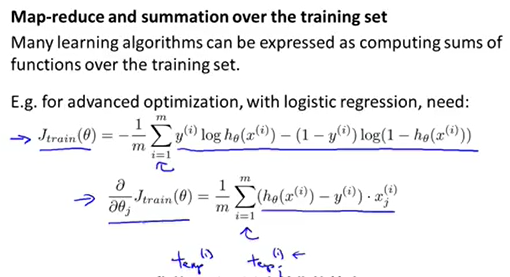
原理：通过多台计算机共同处理训练数据，把各计算机计算的结果送到中心计算服务器将这些变量合并起来。



例子:



映射约减算法需要尽可能的使机器学习算法可以表示为多个训练集的求和形式：



多核计算机的映射约减算法：

