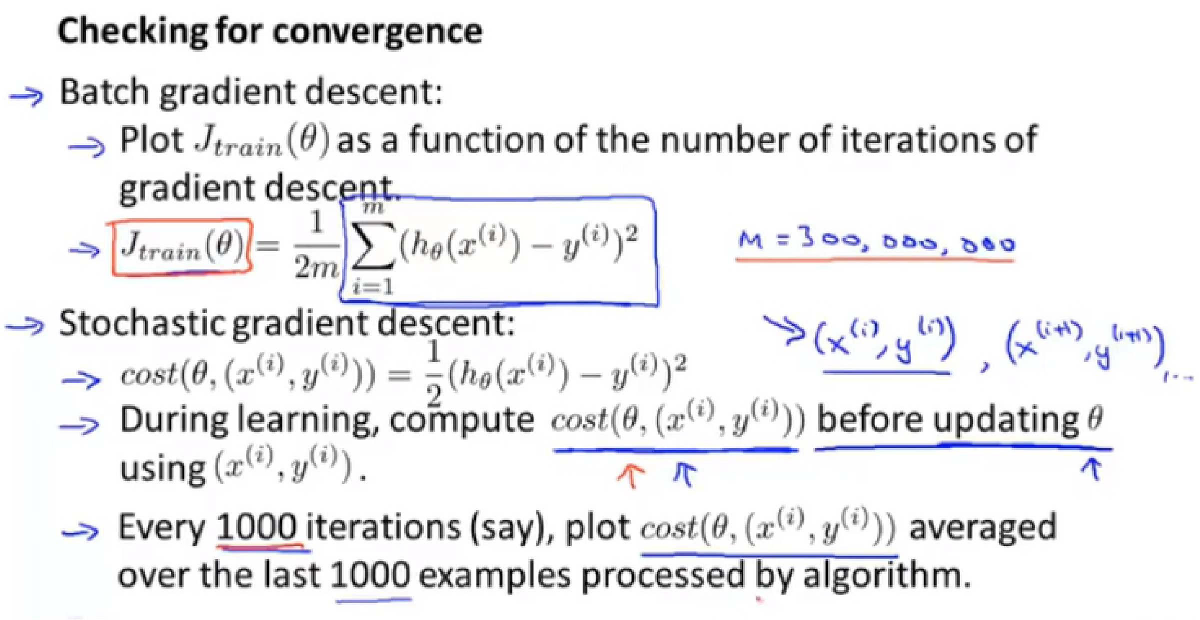
检查梯度下降是否收敛：

批量梯度下降：先画出迭代次数和代价函数的图像，再进行梯度下降，根据图像判断。

随机梯度下降：先计算代价函数，再更新参数sita。每1000此迭代，画出上1000个算法处理的图像，判断是否收敛。

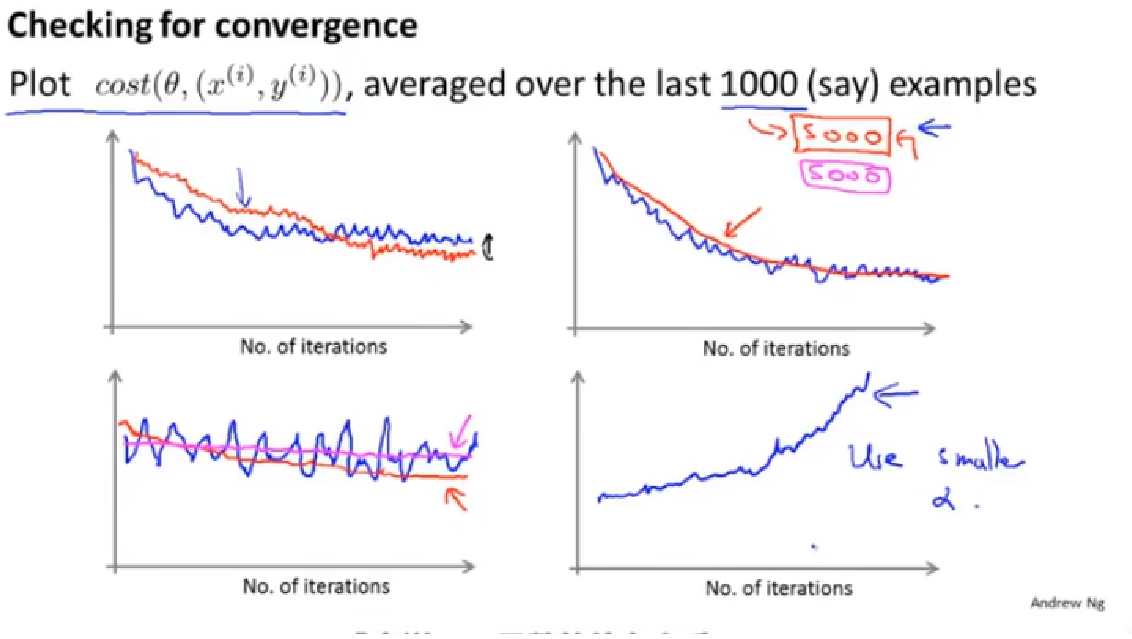


当减小学习率，函数噪声会更加大。如图一。

如果函数图像的噪声太大，可以加大每次迭代使用的数据数量，但这样可能造成反馈有延迟。如图二。

图像过于震荡，如图三，但他们的平均值总体是在下降的，那么可以增大训练样本。

当图像并不是在下降趋势，说明算法发散，如图四，可以使用更小的学习率。



我们也可以调整学习速率，让学习速率随着迭代次数逐渐减小，设置const1和const2，根据公式设置学习速率。