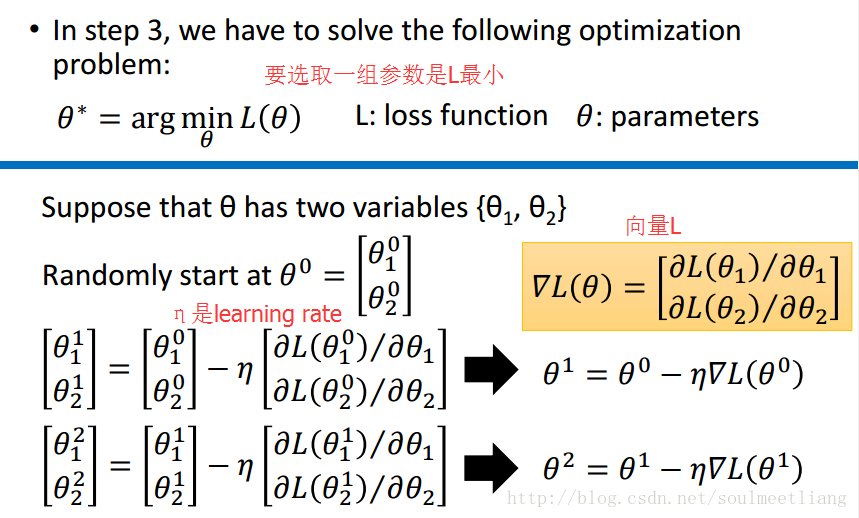
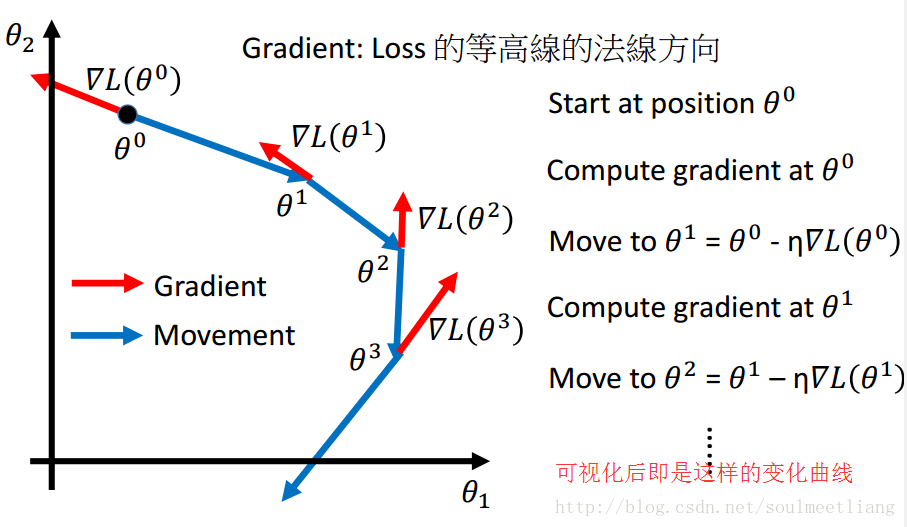
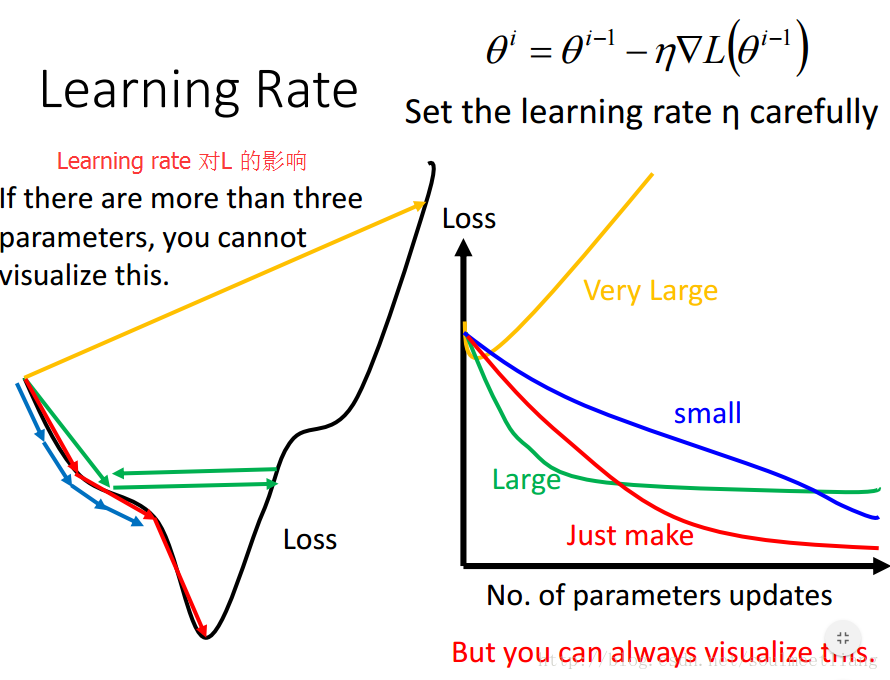
**三．**[**[机器学习入门] 李宏毅机器学习笔记-3 （Gradient Descent ；梯度下降）**](http://blog.csdn.net/soulmeetliang/article/details/72830179)

# Review

# 梯度下降的三个小贴士

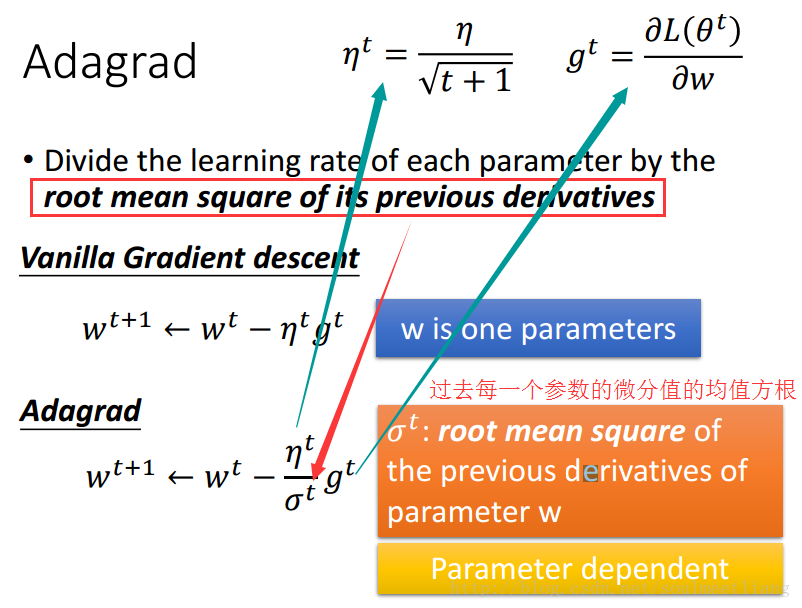
## Tip 1 Tuning your learning rates



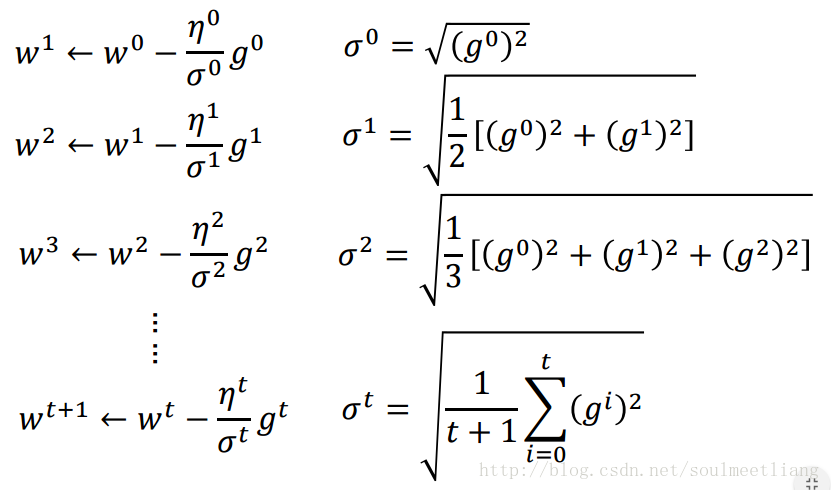
* 最流行也是最简单的做法就是：在每一轮都通过一些因子来减小learning rate。   
  + 最开始时，我们距离最低点很远，所以我们用较大的步长。
  + 经过几轮后，我们接近了最低点，所以我们减少learning rate。
  + 比如： 1/t 衰减： ηn=ηt+1√
* learning rate 不能从一而终   
  + 要给不同的参数设置不同的learning rate。

为了达到此目的，有许多种技巧，而Adagrad就是一种不错的选择。

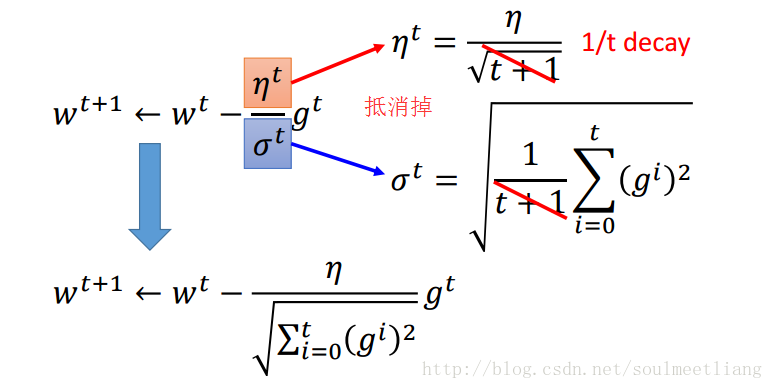
### Adagrad



这样操作后，每组参数的learning rate 都不同。

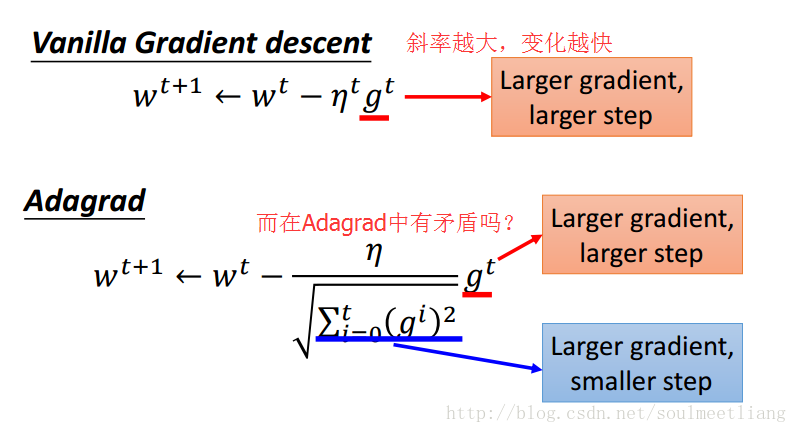
举个例子：   


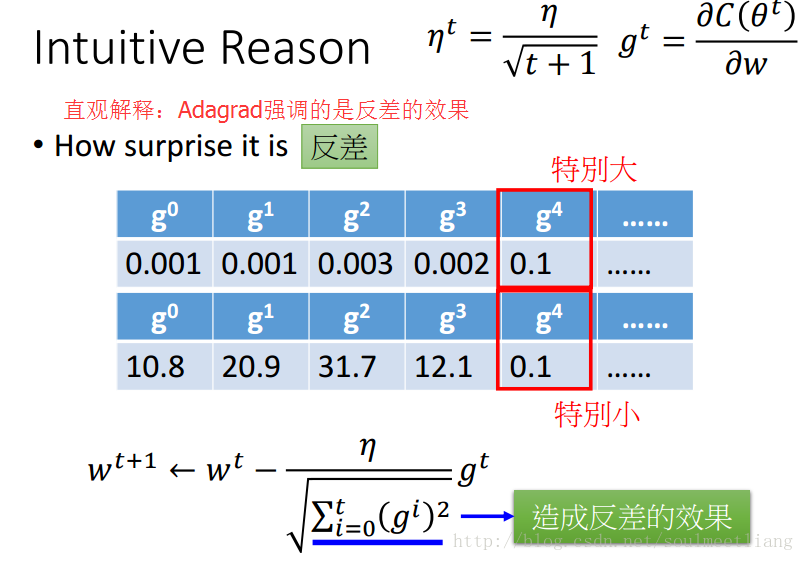
所以用Adagrad后，我们的参数变化要写成这样：

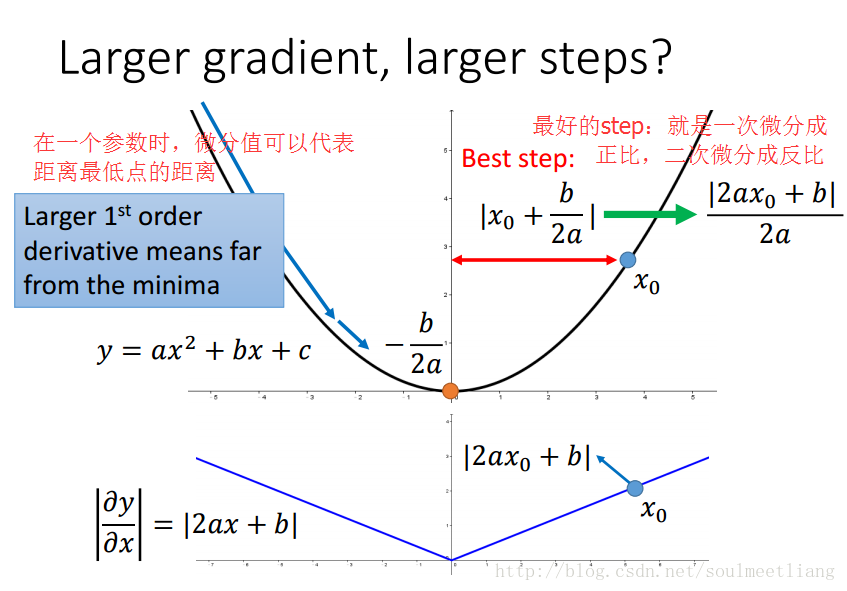
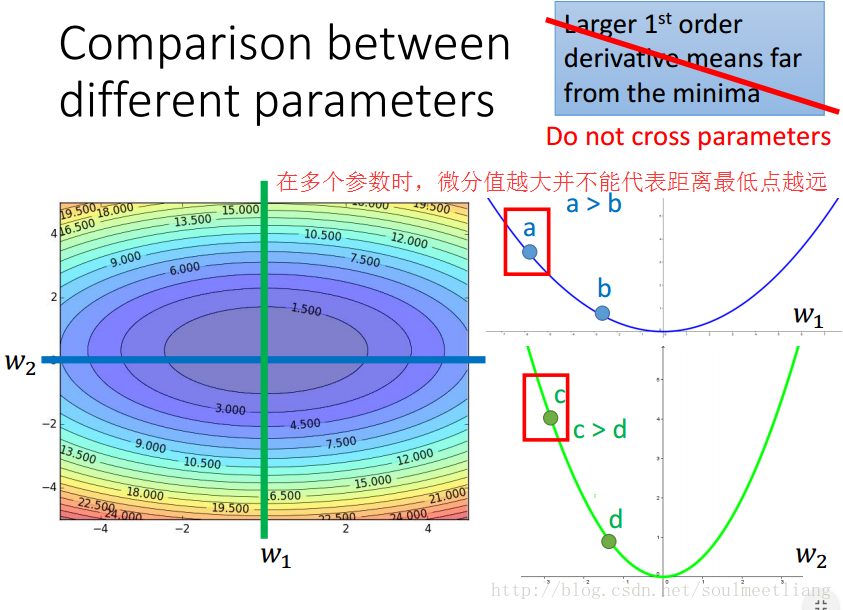
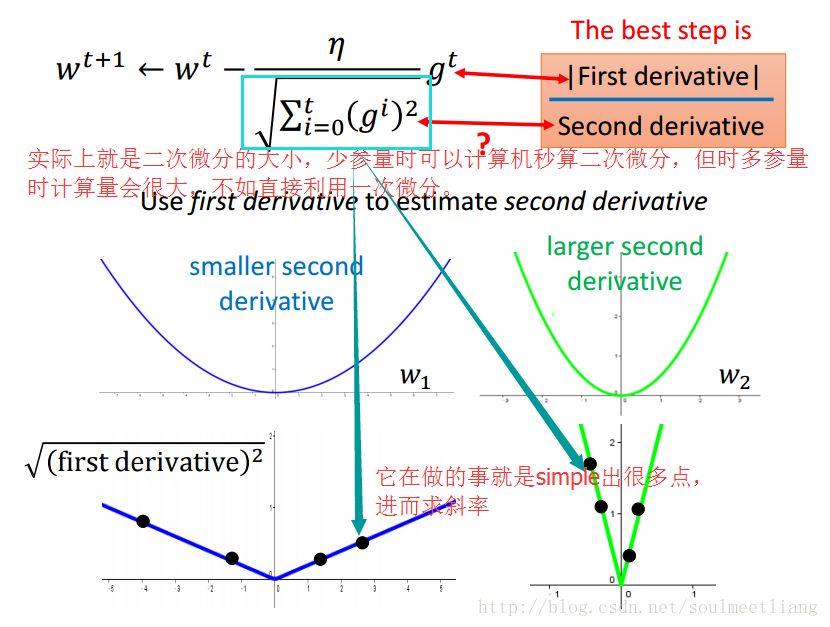


Adagrad越到后面改变会越慢，这是一个正常现象。

Q：有没有奇怪的地方？？是否有所冲突？？

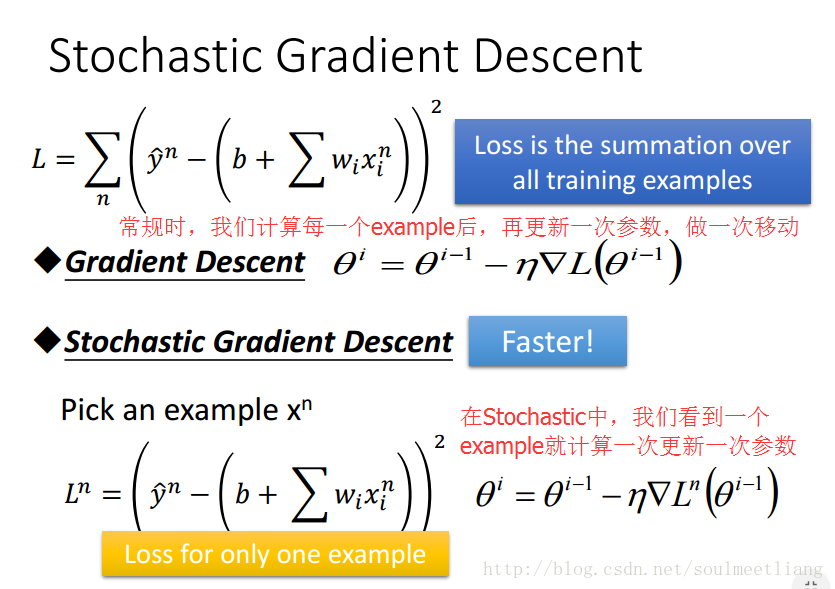


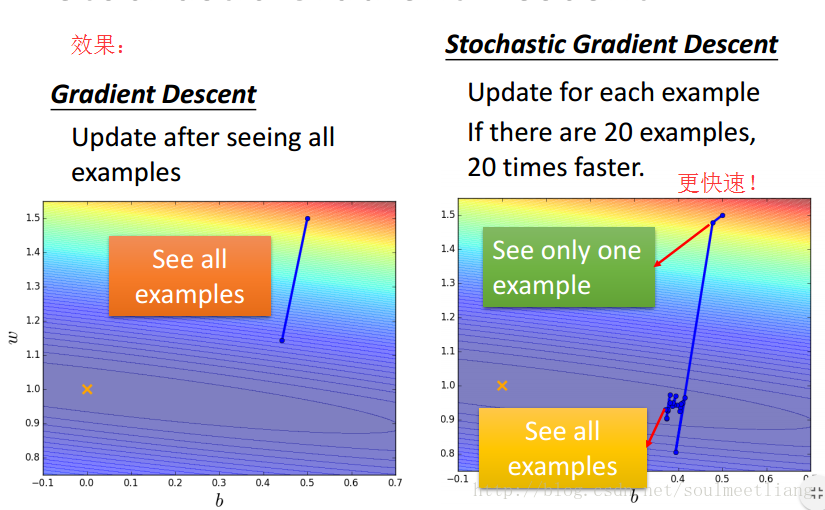


## Tip 2：Stochastic Gradient Descent

让你的Training更快一点！





## Tip 3 : Feature Scaling

让不同的特征值具有相同的缩放程度。

