# 一 mini batch梯度下降

为了解决什么问题?

为了提高训练速度.

具体实现如图:



将500万个样本,分成5000份, 每份1000个样本.这样是为一个mini batch, 每次对一个mini batch做梯度下降更新.



上图右侧规定了几个符号. X(i)是第i个样本. Z[l]是第l层的神经元的z值, X{t},Y{t}表示第t个mini batch的样本集.

Minibatch的训练方法:

下图展示了一个epoch的训练顺序.

首先, 最外层轮询5000份mini batch

然后, 里层,每次执行1000个单位的Loss计算及bp分解以及权值更新(以向量化运行).

其中Loss公式为(带有正则化):





# 二 mini batch深入理解

对比batch梯度下降和minibatch 梯度下降, Loss收敛如下图.

1. Batch的收敛平滑, 如果出现不收敛可能是学习率太大的问题.
2. Minibatch的收敛有波动,但是趋势是下降的.波动原因是不同的batch的Loss的差异.



miniBatch的batchsize取值的影响:

1. 取值为样本数m,minibatch梯度下降就是batch梯度下降.

收敛如下图蓝色线段

1. 取值为1, 则为随机梯度下降.

收敛如下图紫色线段,不能在极值点收敛,而是在极值点震荡.



Batch梯度下降和mini batch梯度下降的选择.

1. M小于2000时(认为数据集太小),使用Batch效果好.
2. 数据集大的情况,采用mini batch效果好,但也要注意GPU/CPU是否有足够的内存能够让你用.
3. Batchsize尽量选择2的次方.

# 三 指数加权平均

介绍指数加权平均的目的是, 下面一些算法比梯度下降法快, 但它们的基础是指数加权平均.

## 指数加权平均是什么

**指数加权平均**(exponentially weighted averges)，也叫指数加权移动平均，是一种常用的序列数据处理方式。

它的计算公式如下：



其中，

* θt：为第 t 天的实际观察值，
* Vt: 是要代替 θt 的估计值，也就是第 t 天的指数加权平均值，
* β： 为 V\_{t-1} 的权重，是可调节的超参。( 0 < β < 1 )

比如例子

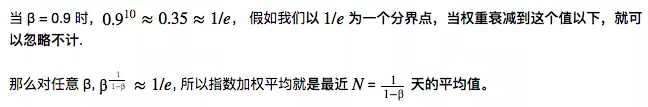
1. 蓝色点是过去一年的温度值.
2. 根据这个温度值,带入上面指数加权平均公式,就能求出加权平均之后的一种个平均值.用红色线表示.
3. 红色线(加权平均值)更加平滑,能够反映出温度趋势.



## 指数加权平均的简化及特点应用

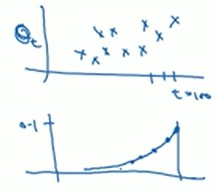
* 1. 展开如下式.
  2. 考虑一个指数特性. 必然有一个**ax**指数在**x**取到**X**时,使得**ax**约等于1/e.
  3. 实际使用过程中,可以忽略那些小于1/e的项.这样可以做到简化.
  4. 特别的, 得到一个简化的推论;

加权平均就是最近天的均值.



另一中理解加权均值

如下图, 将β和θ展开后,输出v可以认为是两者做的卷积.



## 指数加权平均推论应用

在推论下:

加权平均就是最近天的均值.

当β=0.9时, υ取的最近10天的均值.

当β=0.98时, υ取的最近50天的均值.

当β=0.5时, υ取的最近2天的均值.如下图黄色曲线,由于是最近两天的均值,波动很大.





## 指数加权平均的前期误差修正.

指数加权平均会有一个问题:

1. 按如下公式.
   1. 当v0=0时, v1=0.02θ1, v1<<θ1
   2. V2=0.0196θ1=0.02θ2,也远远小于θ1和θ2.会导致如下曲线图的异常.
      1. 红色曲线是β=0.9的
      2. 紫色曲线是β=0.98的
      3. 绿色曲线是期望β=0.98的, 紫色和绿色有差距,但是在中后期会重合.





为了解决这个偏差问题.引入一个修改方法:

**Vt修改=vt原始/(1-βt) , t为天数.**



1. 这个修改可以弥补t较小时的偏差
2. 在t较大的时候,分母**(1-βt)**接近1,不会导致影响后期行为.

动量梯度下降法

为了解决什么问题?

它比标准梯度下降法快很多.(标准梯度下降的学习率不能太大,避免发散).

是什么?

本质利用在梯度上应用指数加权平均,用该”指数加权平均”的梯度更新权重.

和标准梯度下降法的对比:

1. 蓝色描述的是标准梯度下降法.无论minibatch,batch都会有波动.也需要太多下降法的lr不能太大,太大就会导致发散(如紫色箭头的波动).
2. 期望在y轴上学习慢点,在x轴上加快学习.

