# 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.