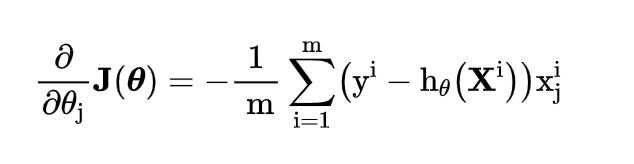
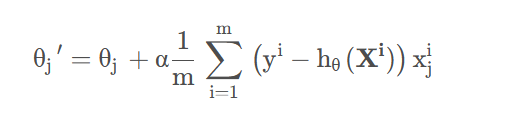
现在我们考虑m个数据 (只需要将m个值求平均即可):



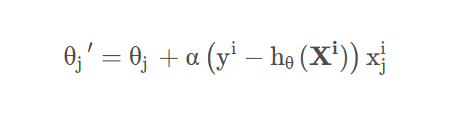
**(1).批梯度下降法(MGD）**

在最小化损失函数的过程中，需要不断反复的更新权重θ使得误差函数减小，更新过程如下:



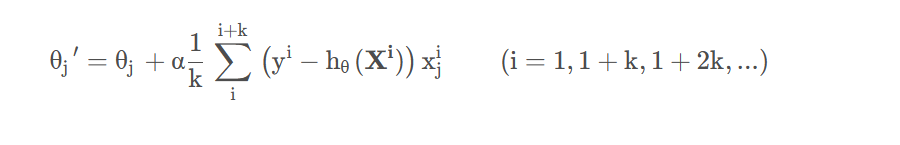
**(2).随机梯度下降法(SGD)**

由于MGD每更新一个参数时，都要用到所有的样本数，所以训练速度会随着样本数量的增加而变得很慢。而SGD每更新一个参数，仅会随机取一个样本：



#### (3).小批量梯度下降(MBGD)

MBGD小批量梯度下降法，假设每次更新参数时用到的样本数为 k kk ：



1.BGD(批梯度下降)

每次更新会使用所有的训练数据来最小化损失函数。 如果只有一个极小值，那么BGD是考虑了所有训练集数据，朝着最小值迭代运动。缺点是如果样本值很大的话，更新速度会很慢。

2.SGD(随机梯度下降)

在每次更新的时候，由于只考虑了一个样本点，大大加快了训练数据的速度，但是由于训练数据的噪声点较多，那么每一次利用噪声点进行更新，就不一定是朝着极小值方向更新，但是由于更新多次，所以整体方向上还是朝着极小值方向更新。

3.MBGD(小批量梯度下降)

MBGD为了综合BGD的训练速度慢以及SGD的准确性综合而来。但是这里要注意，不同问题的batch是不一样的。