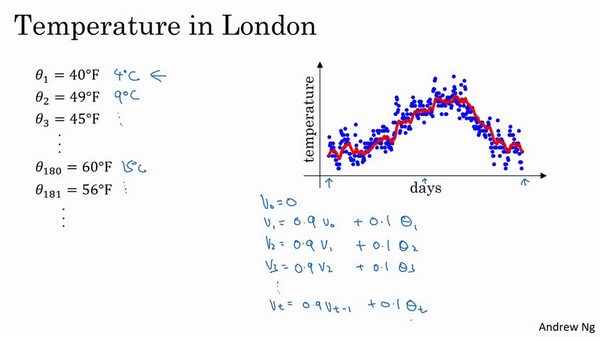
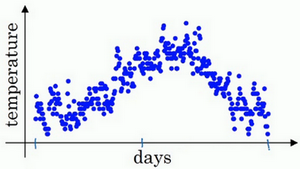
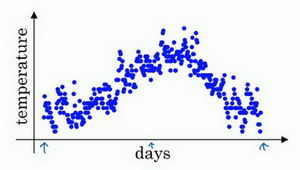
### 2.3 指数加权平均数（Exponentially weighted averages）

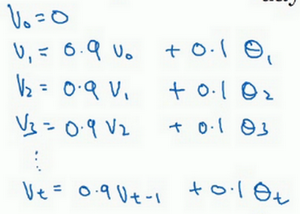
我想向你展示几个优化算法，它们比梯度下降法快，要理解这些算法，你需要用到指数加权平均，在统计中也叫做指数加权移动平均，我们首先讲这个，然后再来讲更复杂的优化算法。  虽然现在我生活在美国，实际上我生于英国伦敦。比如我这儿有去年伦敦的每日温度，所以1月1号，温度是40华氏度，相当于4摄氏度。我知道世界上大部分地区使用摄氏度，但是美国使用华氏度。在1月2号是9摄氏度等等。在年中的时候，一年365天，年中就是说，大概180天的样子，也就是5月末，温度是60华氏度，也就是15摄氏度等等。夏季温度转暖，然后冬季降温。



你用数据作图，可以得到以下结果，起始日在1月份，这里是夏季初，这里是年末，相当于12月末。



这里是1月1号，年中接近夏季的时候，随后就是年末的数据，看起来有些杂乱，如果要计算趋势的话，也就是温度的局部平均值，或者说移动平均值。



你要做的是，首先使，每天，需要使用0.9的加权数之前的数值加上当日温度的0.1倍，即，所以这里是第一天的温度值。

第二天，又可以获得一个加权平均数，0.9乘以之前的值加上当日的温度0.1倍，即，以此类推。

第二天值加上第三日数据的0.1，如此往下。大体公式就是某天的等于前一天值的0.9加上当日温度的0.1。

如此计算，然后用红线作图的话，便得到这样的结果。

图片包含 地图, 文字

描述已自动生成

你得到了移动平均值，每日温度的指数加权平均值。

看一下上一张幻灯片里的公式，，我们把0.9这个常数变成，将之前的0.1变成，即

图片包含 文字, 白板

描述已自动生成

由于以后我们要考虑的原因，在计算时可视大概是的每日温度，如果是0.9，你会想，这是十天的平均值，也就是红线部分。

我们来试试别的，将设置为接近1的一个值，比如0.98，计算，这就是粗略平均了一下，过去50天的温度，这时作图可以得到绿线。

图片包含 文字, 地图

描述已自动生成

这个高值要注意几点，你得到的曲线要平坦一些，原因在于你多平均了几天的温度，所以这个曲线，波动更小，更加平坦，缺点是曲线进一步右移，因为现在平均的温度值更多，要平均更多的值，指数加权平均公式在温度变化时，适应地更缓慢一些，所以会出现一定延迟，因为当，相当于给前一天的值加了太多权重，只有0.02的权重给了当日的值，所以温度变化时，温度上下起伏，当 较大时，指数加权平均值适应地更缓慢一些。

我们可以再换一个值试一试，如果是另一个极端值，比如说0.5，根据右边的公式（），这是平均了两天的温度。



作图运行后得到黄线。

图片包含 地图, 文字

描述已自动生成

由于仅平均了两天的温度，平均的数据太少，所以得到的曲线有更多的噪声，有可能出现异常值，但是这个曲线能够更快适应温度变化。

所以指数加权平均数经常被使用，再说一次，它在统计学中被称为**指数加权移动平均值**，我们就简称为指数加权平均数。通过调整这个参数（），或者说后面的算法学习，你会发现这是一个很重要的参数，可以取得稍微不同的效果，往往中间有某个值效果最好，为中间值时得到的红色曲线，比起绿线和黄线更好地平均了温度。

现在你知道计算指数加权平均数的基本原理，下一个视频中，我们再聊聊它的本质作用。

### 2.4 理解指数加权平均数（Understanding exponentially weighted averages）

上个视频中，我们讲到了指数加权平均数，这是几个优化算法中的关键一环，而这几个优化算法能帮助你训练神经网络。本视频中，我希望进一步探讨算法的本质作用。

回忆一下这个计算指数加权平均数的关键方程。

的时候，得到的结果是红线，如果它更接近于1，比如0.98，结果就是绿线，如果小一点，如果是0.5，结果就是黄线。

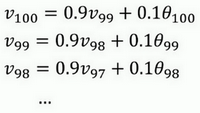
图片包含 地图, 文字

描述已自动生成

我们进一步地分析，来理解如何计算出每日温度的平均值。

同样的公式，

使，写下相应的几个公式，所以在执行的时候，从0到1到2到3，的值在不断增加，为了更好地分析，我写的时候使得的值不断减小，然后继续往下写。



首先看第一个公式，理解是什么？我们调换一下这两项（），。

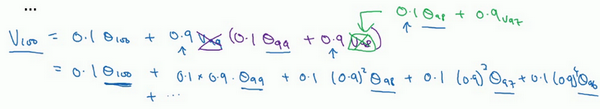
那么是什么？我们就代入这个公式（），所以：

。

那么是什么？你可以用这个公式计算（），把公式代进去，所以：

。

以此类推，如果你把这些括号都展开，



所以这是一个加和并平均，100号数据，也就是当日温度。我们分析的组成，也就是在一年第100天计算的数据，但是这个是总和，包括100号数据，99号数据，97号数据等等。画图的一个办法是，假设我们有一些日期的温度，所以这是数据，这是，所以100号数据有个数值，99号数据有个数值，98号数据等等，为100，99，98等等，这就是数日的温度数值。

图片包含 白板

描述已自动生成

然后我们构建一个指数衰减函数，从0.1开始，到，到，以此类推，所以就有了这个指数衰减函数。

图片包含 天空

描述已自动生成

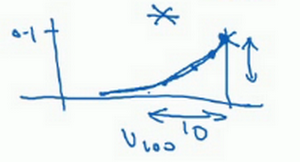
计算是通过，把两个函数对应的元素，然后求和，用这个数值100号数据值乘以0.1，99号数据值乘以0.1乘以，这是第二项，以此类推，所以选取的是每日温度，将其与指数衰减函数相乘，然后求和，就得到了。

图片包含 文字, 天空, 白板

描述已自动生成

结果是，稍后我们详细讲解，不过所有的这些系数（），相加起来为1或者逼近1，我们称之为偏差修正，下个视频会涉及。

最后也许你会问，到底需要平均多少天的温度。实际上大约为0.35，这大约是，**e是自然算法的基础之一。**大体上说，如果有，在这个例子中，，所以，约等于，大约是0.34，0.35，换句话说，10天后，曲线的高度下降到，相当于在峰值的。

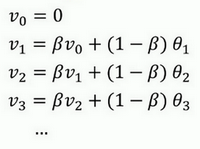


又因此当的时候，我们说仿佛你在计算一个指数加权平均数，只关注了过去10天的温度，因为10天后，权重下降到不到当日权重的三分之一。

图片包含 文字, 白板

描述已自动生成

相反，如果，那么0.98需要多少次方才能达到这么小的数值？大约等于，所以前50天这个数值比大，数值会快速衰减，所以本质上这是一个下降幅度很大的函数，你可以看作平均了50天的温度。因为在例子中，要代入等式的左边，，所以为50，我们由此得到公式，我们平均了大约天的温度，这里代替了，也就是说根据一些常数，你能大概知道能够平均多少日的温度，不过这只是思考的大致方向，并不是正式的数学证明。



最后讲讲如何在实际中执行，还记得吗？我们一开始将设置为0，然后计算第一天，然后，以此类推。

现在解释一下算法，可以将，，等等写成明确的变量，不过在实际中执行的话，你要做的是，**一开始将初始化为0**，然后在第一天使，然后第二天，更新值，，以此类推，有些人会把加下标，来表示是用来计算数据的指数加权平均数。

图片包含 文字

描述已自动生成

再说一次，但是换个说法，，然后每一天，拿到第天的数据，把更新为。

指数加权平均数公式的好处之一在于，它占用极少内存，电脑内存中只占用一行数字而已，然后把最新数据代入公式，不断覆盖就可以了，正因为这个原因，其效率，它基本上只占用一行代码，计算指数加权平均数也只占用单行数字的存储和内存，当然它并不是最好的，也不是最精准的计算平均数的方法。如果你要计算移动窗，你直接算出过去10天的总和，过去50天的总和，除以10和50就好，如此往往会得到更好的估测。但缺点是，如果保存所有最近的温度数据，和过去10天的总和，必须占用更多的内存，执行更加复杂，计算成本也更加高昂。

所以在接下来的视频中，我们会计算多个变量的平均值，从计算和内存效率来说，这是一个有效的方法，所以在机器学习中会经常使用，更不用说只要一行代码，这也是一个优势。

现在你学会了计算指数加权平均数，你还需要知道一个专业概念，叫做偏差修正，下一个视频我们会讲到它，接着你就可以用它构建更好的优化算法，而不是简单直接的梯度下降法。

### 2.5 指数加权平均的偏差修正（Bias correction in exponentially weighted averages）

你学过了如何计算指数加权平均数，有一个技术名词叫做偏差修正，可以让平均数运算更加准确，来看看它是怎么运行的。

图片包含 文字, 地图

描述已自动生成

在上一个视频中，这个（红色）曲线对应的值为0.9，这个（绿色）曲线对应的=0.98，如果你执行写在这里的公式，在等于0.98的时候，得到的并不是绿色曲线，而是紫色曲线，你可以注意到紫色曲线的起点较低，我们来看看怎么处理。

计算移动平均数的时候，初始化，，但是，所以这部分没有了（），所以，所以如果一天温度是40华氏度，那么，因此得到的值会小很多，所以第一天温度的估测不准。

，如果代入，然后相乘，所以，假设和都是正数，计算后要远小于和，所以不能很好估测出这一年前两天的温度。

图片包含 设备

描述已自动生成

有个办法可以修改这一估测，让估测变得更好，更准确，特别是在估测初期，也就是不用，而是用，t就是现在的天数。举个具体例子，当时，，因此对第二天温度的估测变成了，也就是和的加权平均数，并去除了偏差。你会发现随着增加，接近于0，所以当很大的时候，偏差修正几乎没有作用，因此当较大的时候，紫线基本和绿线重合了。不过在开始学习阶段，你才开始预测热身练习，偏差修正可以帮助你更好预测温度，偏差修正可以帮助你使结果从紫线变成绿线。

在机器学习中，在计算指数加权平均数的大部分时候，大家不在乎执行偏差修正，因为大部分人宁愿熬过初始时期，拿到具有偏差的估测，然后继续计算下去。如果你关心初始时期的偏差，在刚开始计算指数加权移动平均数的时候，偏差修正能帮助你在早期获取更好的估测。