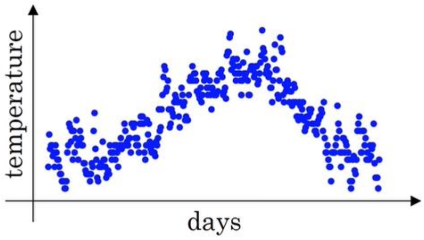
Exponentially Weighted Averages (with/without bias correction):

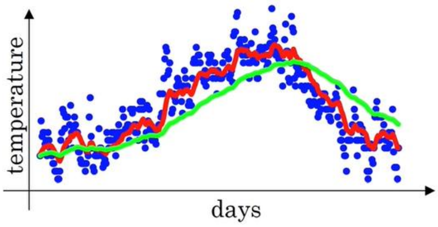
为了理解这个概念，我们可以用一个气温的例子引入，记录了全年每一天的天气气温信息：



如果我们想观察到气温在这年中的一个动态平均的总体趋势，我们可以使用 exponential weighted average：

$$V_t = \beta V_{t-1} + (1-\beta)\theta_t$$

以上是 exponential weighted average 的基础表达式，其中 V_t 可以被看作是 $\frac{1}{1-\beta}$ 天气温的一个近似动态平均值，这意味着参数 β 决定了我们观测的窗口大小（约等于弱化了当前的值 θ_t 对于 V_t 的影响）。举个例子，如果 $\beta = 0.9$ 那么 V_t 就可以被看作是时间点t以及之前 10 天的温度的一个近似均值 (approximately average over 10 days). 下图中的红线对应着 $\beta = 0.9$，绿线对应 $\beta = 0.98$。



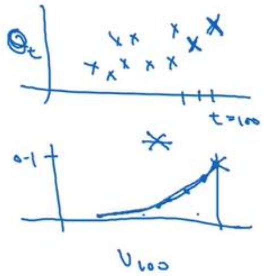
In a word, small β means more susceptible to outlier but adapts more quickly to the latest value changes

Example: Let consider v_{100}, which is the exponentially weighted average of the temperature on t=100

$$v_{100} = 0.9v_{99} + 0.1\theta_{100}$$
$$v_{99} = 0.9v_{98} + 0.1\theta_{99}$$
$$v_{98} = 0.9v_{97} + 0.1\theta_{98}$$
$$\dots$$
$$v_o = 0$$

$$v_{100} = 0.1\theta_{100} + 0.1 \times (0.9)\theta_{99} + 0.1(0.9)^2\theta_{98} + 0.1(0.9)^3\theta_{97} + 0.1(09)^4\theta_{96} + \cdots$$

It is taking the daily temperature multiplying with an exponentially decaying function and then summing it up.



$$And, \qquad 0.9^{10} \approx 0.35 \approx \frac{1}{e}$$

That is why, with β equal to 0.9, a weighted average has a window size of roughly 10.

Bias Correction in exponentially weighted average

Bias correction 主要用于解决 exponentially weighted average 初始阶段，由于 $t < \frac{1}{1-\beta}$ 导致 approximate value 不准确(偏小)的问题，Bias correction 的具体实施细节如下：

$$v_t = \beta v_{t-1} + (1-\beta)\theta_t \qquad \text{(Assume } \beta = 0.98\text{)}$$

Below is the expanded expression of v_{0-2}:

$$\begin{aligned} v_0 &= 0 \\ v_1 &= 0.98v_0 + 0.02\theta_1 \\ v_2 &= 0.98v_1 + 0.02\theta_2 \\ &= 0.98 \times 0.02 \times \theta_1 + 0.02\theta_2 \\ &= 0.0196\theta_1 + 0.02\theta_2 \end{aligned}$$

v_2 will be significantly less than both θ_1 and θ_2, so v_2 is not a very good estimation. To make the estimation more accurate, we can add a **bias correction term** to it:

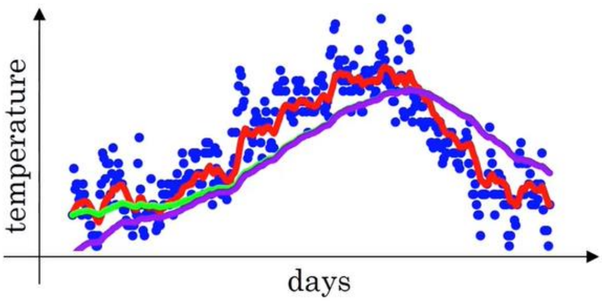
$$\frac{v_t}{1-\beta^t}$$

So that,

$$t = 2: \qquad 1 - \beta^t = 1 - (0.98)^2 = 0.0396$$

$$\frac{V_2}{0.0396} = \frac{0.0196\theta_1 + 0.02\theta_2}{0.0396}$$

As $t \to \infty$, $1 - \beta^t$ will approach 1. As a result we can get **the green curve** shown in the figure below (Compared with no bias correction added, **the purple curve**):



It can be seen that bias correction mainly correct the estimated value at the beginning. With the increase of t, the green curve and the purple curve gradually overlapped.