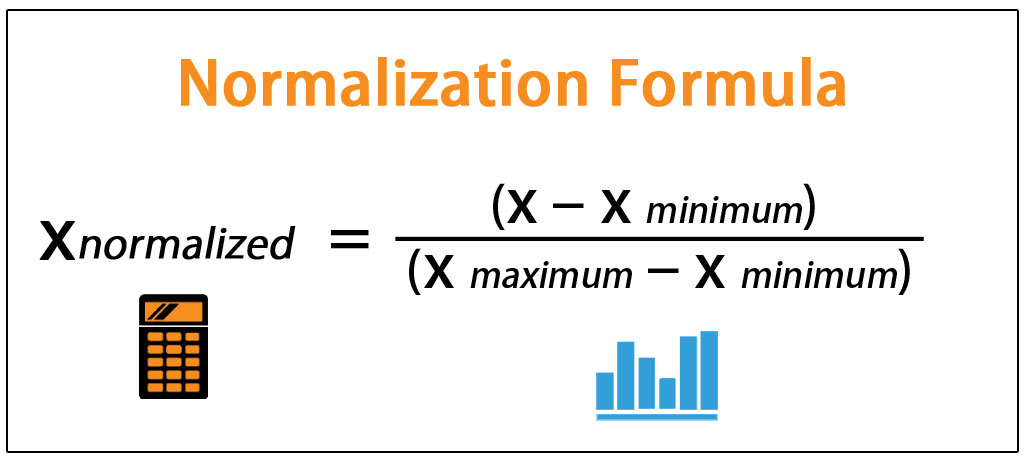
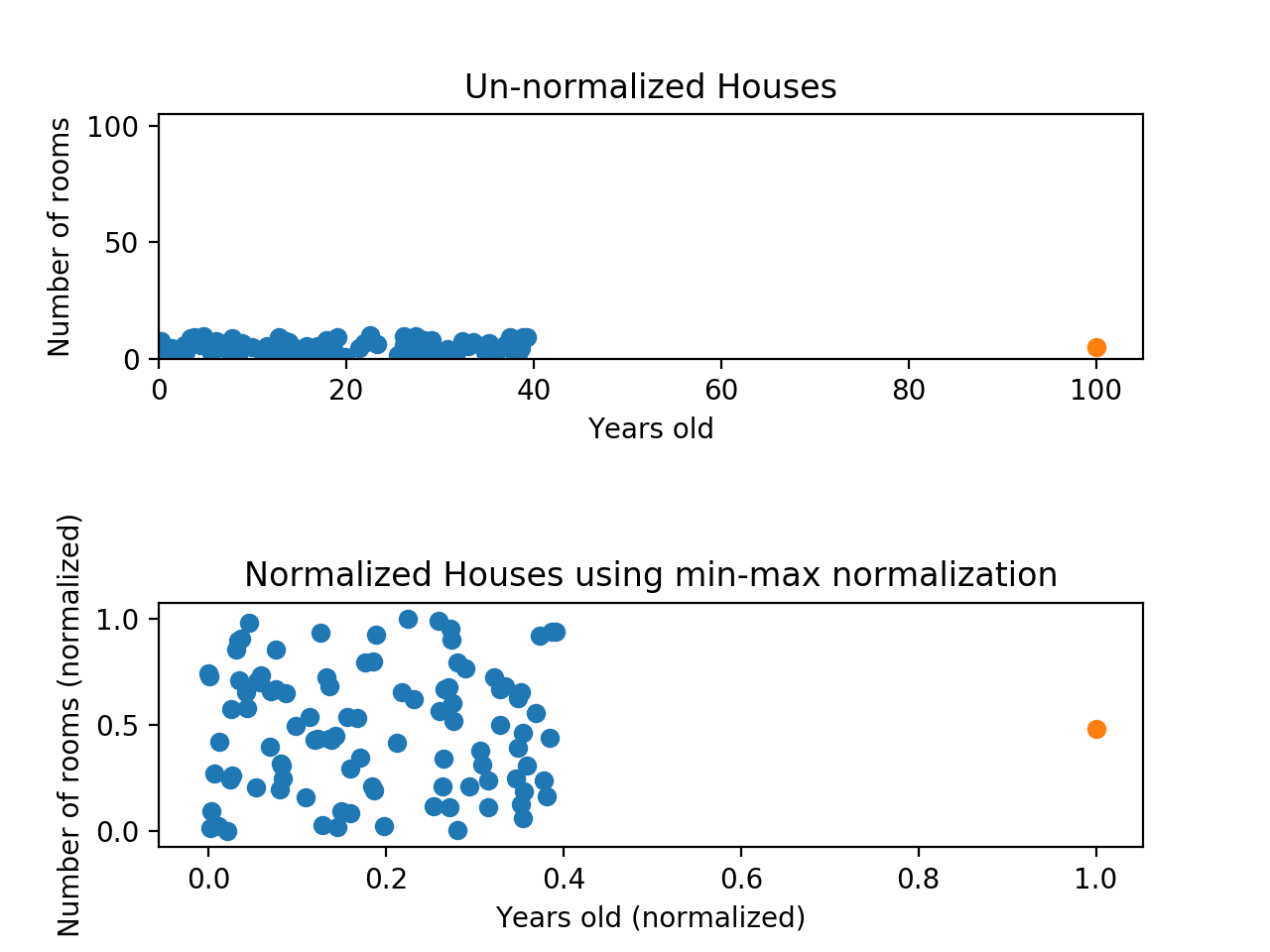
**归一化**

归一化可以解决量纲不统一的问题，在需要计算距离的模型（LR、Kmeans、NN）中需要使用。

**规范化Max-min scalar (normalize)**

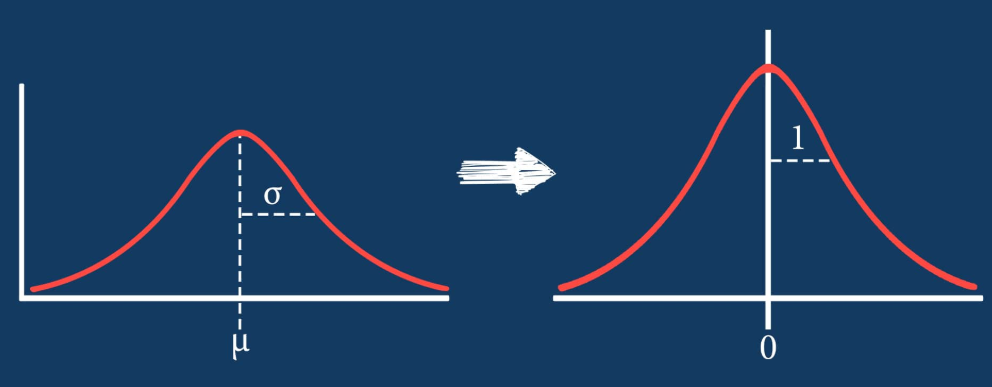


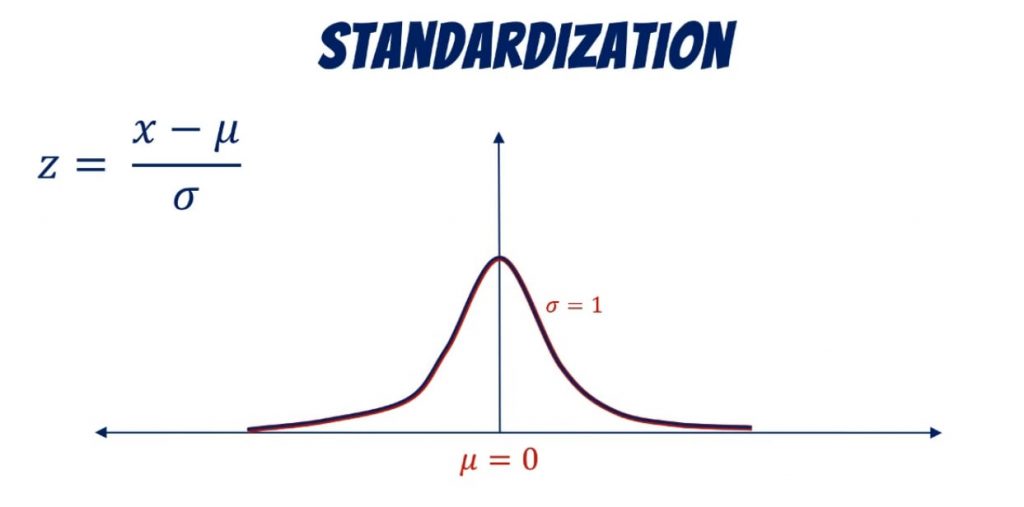


**标准化Data standardized**

数据分布比较陡的情况下，不利于机器进行运算。在很陡的区间Xi每次会变化很大，会在每次计算梯度的时候变化很大，引起梯度抖动严重；在很平的区间Xi每次变化很小接近没有变化，如y也没有什么变化，则遇到新的数据会预测不了；如果Xi每次变化很小，但y有变化（较大），也会使梯度抖动较大，使训练变得很困难。理想的情况是Xi每次变化不太大，也不太小。

遇到此种情况，不是不可以训练，而是训练起来比较困难。此时可以考虑使用标准化，让算法能稳定地更新参数。





如果数据是偏的情况下，可以先用log把数据变成正态分布，再做**standardized**

