

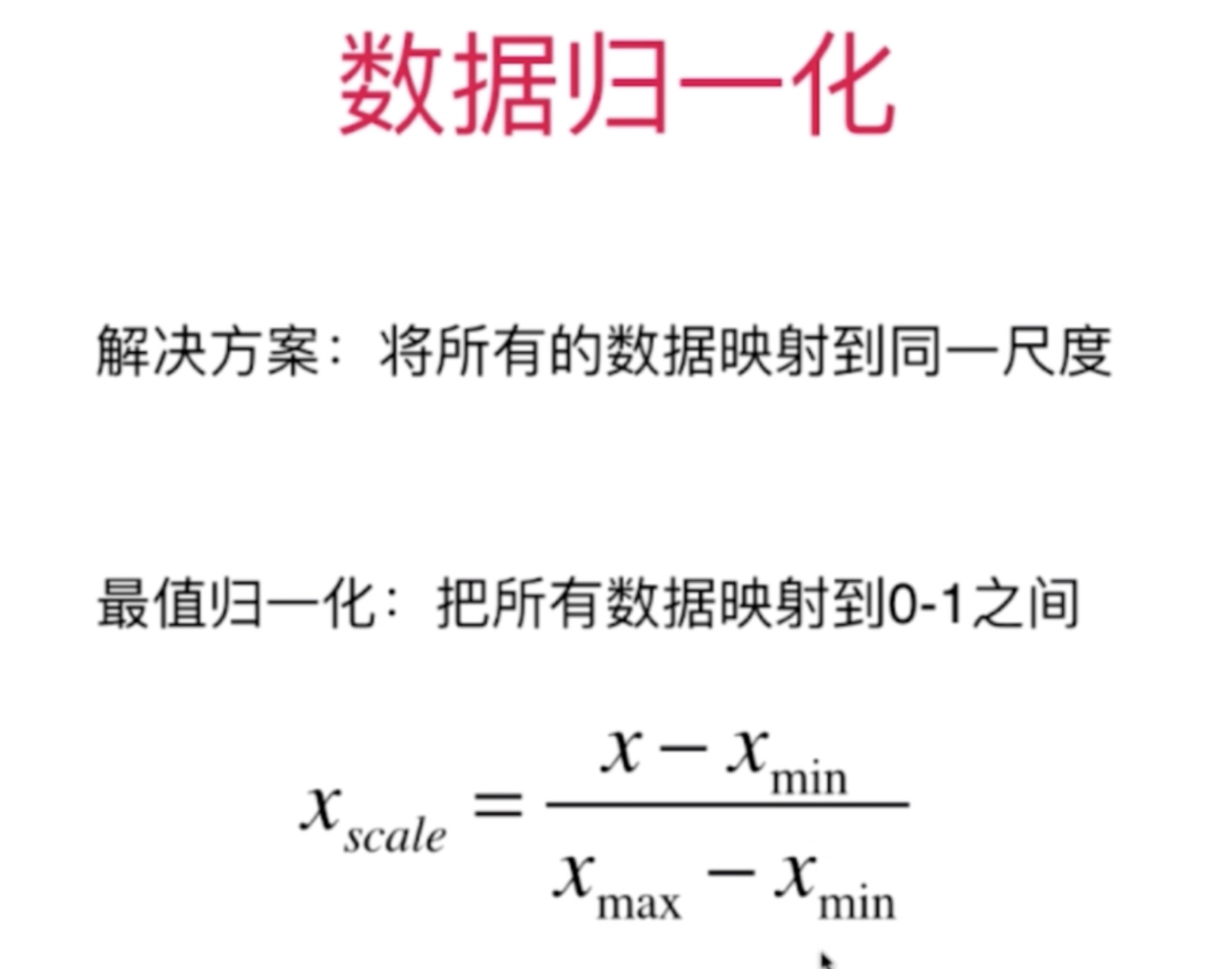


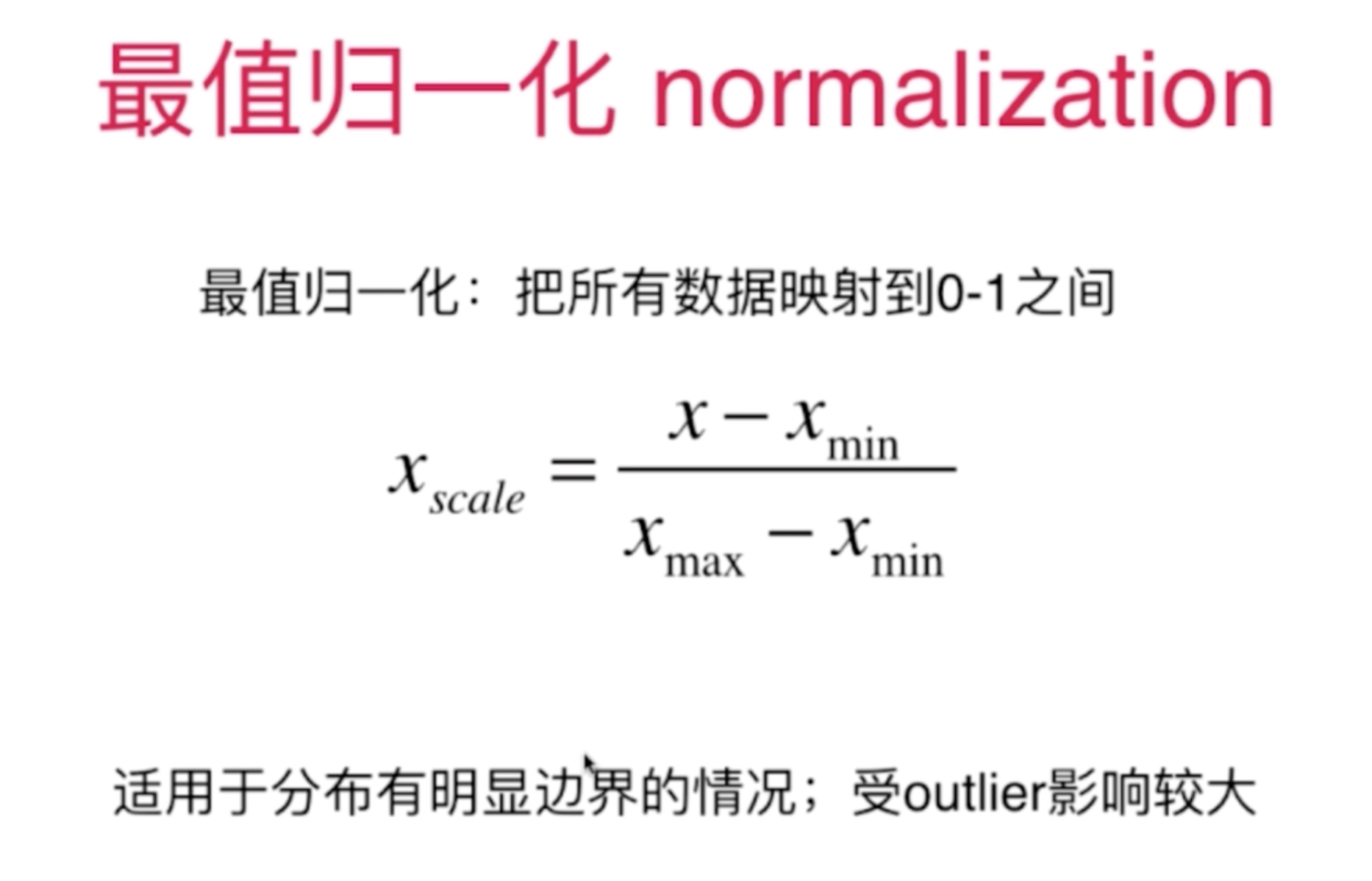
直接针对原始数据计算两个点之间的距离，发现时间特征的数量级明显大于肿瘤大小特征，那么计算出来的结果就会被发现时间特征主导



如果把时间化作年，那么计算距离的结果又会被肿瘤大小特征主导。

所以，如果不对原始数据作一些前期的基础处理的话，后面计算出来的结果是会有很大的偏差，不能很好的同时反映每个特征的重要程度。

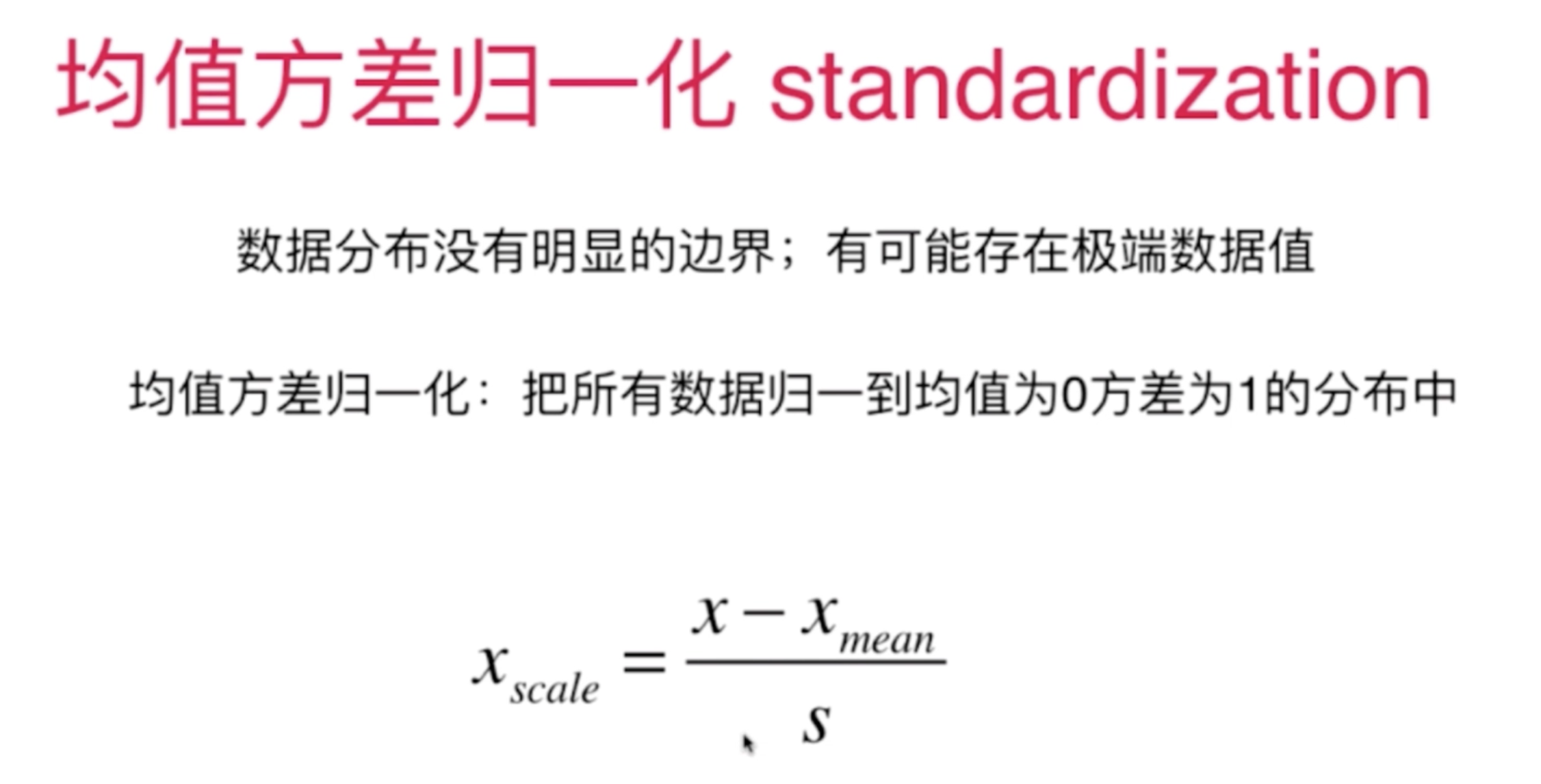




比如成绩分数，1~100，

再比如图像每一个像素点的颜色值，0~255

不适合的是比如人的收入，有些人的收入是极其高的，这些就是 outlier



用代码实现 两种归一化： [04-07 Feature-Scaling.ipynb]