# 几种归一化方法

**1.Zero-mean normalization**

公式：



这就是均值方差归一化，这样处理后的数据将符合标准正太分布，常用在一些通过距离得出相似度的聚类算法中，比如 K-means。

**2.Min-max normalization**

公式：



min-max 归一化的手段是一种线性的归一化方法，它的特点是不会对数据分布产生影响。不过如果你的数据的最大最小值不是稳定的话，你的结果可能因此变得不稳定。min-max 归一化在图像处理上非常常用，因为大部分的像素值范围是 [0, 255]。

**3.Non-linear normalizations**

非线性的归一化函数包含 log，exp，arctan, sigmoid等等。用非线性归一化的函数取决于你的输入数据范围以及你期望的输出范围。比如 log() 函数在 [0, 1] 区间上有很强的区分度，arctan() 可以接收任意实数病转化到



区间，sigmoid 接收任意实数并映射到 (0, 1)。

**4.Length-one normalization**

公式：



将特征转为单位向量的形式，可以剔除特征的强度的影响。这种处理用在不考虑向量大小而需要考虑向量方向的问题中，比如在一些文本情感的分类中，我们可能并不需要知道情感表达的强弱，而只要知道情感的类型，比如开心，生气等等。

