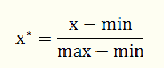
# [数据归一化和两种常用的归一化方法](http://www.cnblogs.com/chaosimple/p/3227271.html)

数据标准化（归一化）处理是数据挖掘的一项基础工作，不同评价指标往往具有不同的量纲和量纲单位，这样的情况会影响到数据分析的结果，为了消除指标之间的量纲影响，需要进行数据标准化处理，以解决数据指标之间的可比性。原始数据经过数据标准化处理后，各指标处于同一数量级，适合进行综合对比评价。以下是两种常用的归一化方法：

**一、min-max标准化（Min-Max Normalization）（归一化）**

也称为离差标准化（极差归一化），是对原始数据的线性变换，使结果值映射到[0 - 1]之间。转换函数如下：



其中max为样本数据的最大值，min为样本数据的最小值。这种方法有个缺陷就是当有新数据加入时，可能导致max和min的变化，需要重新定义。

**二、Z-score标准化方法（标准化）**

这种方法基于原始数据的均值（mean）和标准差（standard deviation）进行数据的标准化。经过处理的数据符合标准正态分布，即均值为0，标准差为1，转化函数为：



其中μ为所有样本数据的均值，σ为所有样本数据的标准差。

案例.matlab  mapminmax函数实现极差归一化

对数据{xn}，归一化为{x’n}落到区间[Ymin, Ymax]（默认是[-1, 1]）, 公式：

x’n =(Ymax-Ymin)\* (xn-Xmin)/(Xmax-Xmin) + Ymin

调用格式：[Y, PS] = mapminmax(X,Ymin,Ymax);  % PS记录该变换过程mapminmax('apply',Z,PS);     % 对矩阵Z实施PS变换过程mapminmax('reverse',Y,PS) ;   % 从Y变换回X

注：(1) Matlab7.1之前版本没有mapminmax函数，可以用premnmx函数。Matlab 2012版本之前mapminmax函数有bug，按标准格式调用得到的结果是错误的，需要特殊处理一下;(2) 若矩阵中有缺省数据，需要用fixunknowns函数。