min-max标准化（线性函数归一化）

- 本质： 对原始数据进行线性变换，将结果映射到[0,1]之间
- 转换函数

$$\frac{x - Min}{Max - Min}$$

  - Min为样本数据的最小值，Max为样本数据的最大值
- 映射到[-1,1], 则换成

$$\frac{x - Mean}{Max - Min}$$

- 当有**新数据加入**时，可能会影响到Min，Max

```
import numpy as np

arr = np.asarray([0, 10, 50, 80, 100])

for x in arr:
        x = float(x - np.min(arr)) / (np.max(arr) - np.min(arr))
        print(x)

# output
# 0.0
# 0.1
# 0.5
# 0.8
# 1.0
```

Z-Score（均值标准化）

- 处理后的数据，符合标准正态分布，即均值(Mean)为0，标准差(Std)为1
- 本质： 把有量纲表达式变成无量纲表达式
- 转换函数

$$\frac{x - Mean}{Std}$$

```
import numpy as np

arr = np.asarray([0, 10, 50, 80, 100])

for x in arr:
```

```
        x = float(x - arr.mean()) / arr.std()
        print(x)

# output
# -1.24101045599
# -0.982466610991
# 0.0517087689995
# 0.827340303992
# 1.34442799399
```

函数转换

1. log函数转换
   - 转换函数 $\log_{10} x$ / $\log_{10} max$
   - max为样本数据最大值，并且**所有数据都要大于等于1**

总结

1. 涉及距离度量（聚类分析）或协方差分析（PCA，LDA等），同时数据分布可以近似为正态分布，使用Z-Score效果更好。 因为第一种方法不能消除量纲对方差（协方差）的影响，对PCA分析影响巨大；
2. 在不涉及距离度量，协方差计算，数据不太符合正态分布的时候，可以使用第一种方法或其他归一化方法。