homework_two

2018 年 10 月 24 日

0.1 1. 答案：PCA, Attribute Selection, Wavelet;

0.2 2. 答案：

1. [0, +inf];
2. [-1, +1];
3. [-1, +1];

```
In [1]: import numpy as np
        from sklearn import preprocessing

        # original data set
        dataset = np.array([[13.], [15.], [16.], [16.], [19.],
                            [20.], [20.], [21.], [22.], [22.],
                            [25.], [25.], [25.], [25.], [30.],
                            [33.], [33.], [35.], [35.], [35.],
                            [35.], [36.], [40.], [45.], [46.],
                            [52.], [70.]])

        print('数据均值: ', np.mean(dataset))
        print('数据方差: ', np.var(dataset))
```

```
数据均值:  29.962962962962962
数据方差:  161.29492455418384
```

0.3 3.(a) 答案：

```
In [25]: # min-max normalization with range [0. , 1.]
         minmax_normalized = preprocessing.minmax_scale(dataset)
         print('min-max normalized 35 =', minmax_normalized[17][0])
```

```
min-max normalized 35 = 0.3859649122807017
```

0.4 3.(b) 答案：

```
In [22]: # z-score normalization with std = 12.94
         scaler = preprocessing.scale(dataset, with_std=12.94)
         print('z-score normalized 35 =', scaler[17][0])
```

```
z-score normalized 35 = 0.3966110348537352
```

0.5 3.(c) 答案：

decimal normalized 35 = 0.35

0.6 3.(d) 答案：**z-score 规范化更好，因为它利用了数据的分布信息；**