

数据预处理——重复值处理

若样本中含有相同的重复数据，将会导致样本不唯一，使得机器学习算法对同一样本反复学习，容易导致模型过拟合。因此，去重是数据挖掘过程中的一个必备过程。MATLAB 提供了 `unique` 函数来去除重复值。

% 返回 A 中所有唯一值

```
C = unique(A)
```

% 添加参数 'rows'，返回 A 中所有不重复的唯一行

```
C = unique(A, 'rows')
```

公众号：数模加油站
QQ群：544457657