

Рубежный контроль

Гун Янь

Вариант:17

Задача1:17. Задача2:37.

```
import pandas as pd
import numpy as np
from sklearn.datasets import load_iris
from sklearn.preprocessing import PowerTransformer
from sklearn.feature_selection import SelectPercentile, mutual_info_classif
```

```
[2] # 加载Iris数据集
iris = load_iris()
X = iris.data
y = iris.target
```

```
[3] # 转换为DataFrame以便于操作
df = pd.DataFrame(X, columns=iris.feature_names)
df['target'] = y
```

```
[4] print("原始数据集: ")
print(df.head())
```

```
[3] # 转换为DataFrame以便于操作
df = pd.DataFrame(X, columns=iris.feature_names)
df['target'] = y
```

```
print("原始数据集: ")
print(df.head())
```

原始数据集:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	\
0	5.1	3.5	1.4	0.2	
1	4.9	3.0	1.4	0.2	
2	4.7	3.2	1.3	0.2	
3	4.6	3.1	1.5	0.2	
4	5.0	3.6	1.4	0.2	

	target
0	0
1	0
2	0
3	0
4	0

```
[5] pt = PowerTransformer(method='yeo-johnson')
df[iris.feature_names[0]] = pt.fit_transform(df[[iris.feature_names[0]]])
```

```
[5] pt = PowerTransformer(method='yeo-johnson')
df[iris.feature_names[0]] = pt.fit_transform(df[[iris.feature_names[0]]])
```

```
print("\nYeo-Johnson变换后的数据集：")
print(df.head())
```



Yeo-Johnson变换后的数据集：

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	\
0	-0.895690	3.5	1.4	0.2	
1	-1.185173	3.0	1.4	0.2	
2	-1.487921	3.2	1.3	0.2	
3	-1.644609	3.1	1.5	0.2	
4	-1.038838	3.6	1.4	0.2	

	target
0	0
1	0
2	0
3	0
4	0

```
[7] X = df.drop(columns=['target'])
y = df['target']
```

```
[8] # 使用SelectPercentile和互信息进行特征选择
selector = SelectPercentile(mutual_info_classif, percentile=5)
X_new = selector.fit_transform(X, y)
```

```
# 打印选择的特征
selected_features = np.array(iris.feature_names)[selector.get_support()]
print("\n选择的特征：")
print(selected_features)
```



选择的特征：
['petal width (cm)']

