CHRIS ALBON

ANOVA F-value For Feature Selection

20 Dec 2017

If the features are categorical, calculate a chi-square (χ^2) statistic between each feature and the target vector. However, if the features are quantitative, compute the ANOVA F-value between each feature and the target vector.

The F-value scores examine if, when we group the numerical feature by the target vector, the means for each group are significantly different.

Preliminaries

```
# Load libraries
from sklearn.datasets import load_iris
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import f_classif
```

Load Data

CHRIS ALBON

```
# Create features and target
X = iris.data
y = iris.target
```

Select Features With Best ANOVA F-Values

```
# Create an SelectKBest object to select features
with two best ANOVA F-Values
fvalue_selector = SelectKBest(f_classif, k=2)

# Apply the SelectKBest object to the features and
target
X_kbest = fvalue_selector.fit_transform(X, y)
```

View Results

```
# Show results
print('Original number of features:', X.shape[1])
print('Reduced number of features:',
X_kbest.shape[1])
```

```
Original number of features: 4
Reduced number of features: 2
```

Find an error or bug?

CHRIS ALBON

directly on Twitter.

This project contains 423 pages and is available on GitHub.