ML | Chi-square Test for feature selection

Difficulty Level : Easy ● Last Updated : 24 Dec, 2018

Feature selection is also known as attribute selection is a process of *extracting the most relevant features* from the dataset and then applying machine learning algorithms for the better performance of the model. A large number of irrelevant features increases the training time exponentially and increase the risk of overfitting.

Chi-square Test for Feature Extraction:
Chi-square test is used for categorical features in a dataset. We calculate Chi-square between each feature and the target and select the desired number of features with best Chi-square scores. It determines if the association between two categorical variables of the sample would reflect their real association in the population.
Chi- square score is given by :

$$|X^2 = \frac{(Observed\ frequency - Expected\ frequency)^2}{Expected\ frequency}$$

where –

Observed frequency = No. of observations of class

Python Implementation of Chi-Square feature selection:

Related Articles

```
# Load iris data
iris_dataset = load_iris()

# Create features and target
X = iris_dataset.data
y = iris_dataset.target

# Convert to categorical data by converting data to integers
X = X.astype(int)

# Two features with highest chi-squared statistics are selected
chi2_features = SelectKBest(chi2, k = 2)
X_kbest_features = chi2_features.fit_transform(X, y)

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

Output:

```
Original feature number: 4
Reduced feature number : 2
```

Attention reader! Don’t stop learning now. Get hold of all the important Machine Learning Concepts with the **Machine Learning Foundation Course** at a student-friendly price and become industry ready.

Previous Next

RECOMMENDED ARTICLES

Page : 1 2 3

01 numpy.random.chisquare() in Python
03, Jul 20

02 Chi-Square Test for Feature Selection - Mathematical Explanation
19, Jul 19

03 Parameters for Feature Selection
29, Apr 18

04 ML | Extra Tree Classifier for Feature Selection
21, Jul 19

05 Feature Selection using Branch and Bound Algorithm
24, Nov 20

06 Feature Selection Techniques in Machine Learning
19, Jan 21

07 Feature Subset Selection Process
09, Mar 21

08 ML | Feature Mapping
10, Jul 18

Article Contributed By :

aakarsha_chugh
@aakarsha_chugh

Vote for difficulty

Current difficulty : Easy

Easy | Normal | Medium | Hard | Expert

Article Tags : Machine Learning, Python

Practice Tags : Machine Learning

Writing code in comment? Please use ide.geeksforgeeks.org, generate link and share the link here.

Load Comments

GeeksforGeeks
5th Floor, A-118,
Sector-136, Noida, Uttar Pradesh - 201305

feedback@geeksforgeeks.org