perform classification on a non-linear dataset

ML | Locally weighted Linear Regression

ML | Monte Carlo Tree Search (MCTS)

Applying Multinomial Naive Bayes to NLP Problems

Data Structure to Design a special social network

ML | Natural Language Processing using Deep Learning

MATLAB | Display histogram of a grayscale Image

Audio processing using Pydub and Google speechRecognition API

LSB based Image steganography using MATLAB

Elasticsearch Search Engine | An introduction

Draw Austria flag using Matlab

Introduction to Signals and Systems: Properties of systems

Python | Speech recognition on large audio files

Text extraction

from image using LSB based steganography

Draw Bangladesh Flag Using Matlab

MATLAB | Convert video into slow motion

Image Resizing in Matlab

Draw Switzerland Flag Using Matlab

ML | Label Encoding of datasets in Python

In machine learning, we usually deal with datasets which contains multiple labels in one or more than one columns. These labels can be in the form of words or numbers. To make the data understandable or in human readable form, the training data is often labeled in words.

Label Encoding refers to converting the labels into numeric form so as to convert it into the machine-readable form. Machine learning algorithms can then decide in a better way on how those labels must be operated. It is an important pre-processing step for the structured dataset in supervised learning.

Example :

Suppose we have a column *Height* in some dataset.

Height
Tall
Medium
Short

After applying label encoding, the Height column is converted into:

Height
0
1
2

where 0 is the label for tall, 1 is the label for medium and 2 is label for short height.

We apply *Label Encoding* on `iris dataset` on the target column which is Species. It contains three species *Iris-setosa, Iris-versicolor, Iris-virginica*.

```
# Import libraries
import numpy as np
import pandas as pd

# Import dataset
df = pd.read_csv('../../data/Iris.csv')

df['species'].unique()
```

Output:

```
array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

After applying Label Encoding –

```
# Import label encoder
from sklearn import preprocessing

# label_encoder object knows how to understand word labels.
label_encoder = preprocessing.LabelEncoder()

# Encode labels in column 'species'.
df['species']= label_encoder.fit_transform(df['species'])

df['species'].unique()
```

Output:

```
array([0, 1, 2], dtype=int64)
```


Recommended Posts:

ML | One Hot Encoding of datasets in Python
Python | Create Test DataSets using Sklearn
Python | Generate test datasets for Machine learning
Python | Add Label to a kivy window
Run Length Encoding in Python
Python | C Strings of Doubtful Encoding | Set-2
Python | C Strings of Doubtful Encoding | Set-1
Return the Index label if some condition is satisfied over a column in Pandas Dataframe
Encoding Methods in Genetic Algorithm
Reading Python File-Like Objects from C | Python
Important differences between Python 2.x and Python 3.x with examples
Python | Set 4 (Dictionary, Keywords in Python)
Python | Sort Python Dictionaries by Key or Value
Python | Add Logging to Python Libraries
Python | Add Logging to a Python Script

aakarsha_chugh
Check out this Author's contributed articles.

If you like GeeksforGeeks and would like to contribute, you can also write an article using contribute.geeksforgeeks.org or mail your article to contribute@geeksforgeeks.org. See your article appearing on the GeeksforGeeks main page and help other Geeks.

Please Improve this article if you find anything incorrect by clicking on the "Improve Article" button below.

Article Tags : Advanced Computer Subject, Machine Learning, Python

Practice Tags : Machine Learning

Be the First to upvote.

To-do Done

0

No votes yet.

Feedback/ Suggest Improvement | Add Notes | Improve Article

Please write to us at contribute@geeksforgeeks.org to report any issue with the above content.

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

Load Comments