v



# Convert Pandas Categorical Data For Scikit-Learn

30 November 2016 / Machine Learning / Preprocessing Structured Data

## Preliminaries

```
# Import required packages
from sklearn import preprocessing
import pandas as pd
```

## Create DataFrame

```
raw_data = {'patient': [1, 1, 1, 2, 2],
        'obs': [1, 2, 3, 1, 2],
        'treatment': [0, 1, 0, 1, 0],
        'score': ['strong', 'weak', 'normal', 'weak', 'strong']}
df = pd.DataFrame(raw_data, columns = ['patient', 'obs', 'treatment', 'score'])
```

## Fit The Label Encoder

```
# Create a label (category) encoder object
le = preprocessing.LabelEncoder()
```

```
# Fit the encoder to the pandas column
le.fit(df['score'])
```

```
LabelEncoder()
```

## View The Labels

```
# View the labels (if you want)
list(le.classes_)
```

```
['normal', 'strong', 'weak']
```

## Transform Categories Into Integers

```
# Apply the fitted encoder to the pandas column
le.transform(df['score'])
```

```
array([1, 2, 0, 2, 1])
```

## Transform Integers Into Categories

```
# Convert some integers into their category names
list(le.inverse_transform([2, 2, 1]))
```

```
['weak', 'weak', 'strong']
```

### Find an error or bug?

Everything on this site is available on GitHub. Head to and submit a suggested change (https://github.com/chrisalbon/notes_on_data_science_machine_learning_and_artificial_intelligence/issues/new). You can also message me directly on Twitter (https://twitter.com/chrisalbon).

---

This project contains 488 pages and is available on GitHub (https://github.com/chrisalbon/notes_on_data_science_machine_learning_and_artificial_intelligence).