

```
# remove categorical label
X = np.delete(X, 1, axis=1)
# append encoded matrix
X = np.append(X, one_hot_encode_X.toarray(), axis=1)
X
'Output':
array([[ '5', '8', '1.0', '0.0', '0.0'],
       [ '9', '3', '0.0', '1.0', '0.0'],
       [ '8', '6', '0.0', '0.0', '1.0'],
       [ '0', '5', '0.0', '1.0', '0.0'],
       [ '2', '3', '1.0', '0.0', '0.0'],
       [ '0', '8', '1.0', '0.0', '0.0'],
       [ '1', '8', '0.0', '0.0', '1.0']], dtype='<U32')
```

## Input Missing Data

It is often the case that a dataset contains several missing observations. Scikit-learn implements the **Imputer** module for completing missing values.

```
# import packages
from sklearn. impute import SimpleImputer

# create dataset
data = np.array([[5,np.nan,8],[9,3,5],[8,6,4],
                 [np.nan,5,2],[2,3,9],[np.nan,8,7],
                 [1,np.nan,5]])

data
'Output':
array([[ 5., nan,  8.],
       [ 9.,  3.,  5.],
       [ 8.,  6.,  4.],
       [nan,  5.,  2.],
       [ 2.,  3.,  9.],
       [nan,  8.,  7.],
       [ 1., nan,  5.]])
```

```
# impute missing values - axis=0: impute along columns
imputer = SimpleImputer(missing_values=np.nan, strategy='mean')
imputer.fit_transform(data)
'Output':
array([[5., 5., 8.],
       [9., 3., 5.],
       [8., 6., 4.],
       [5., 5., 2.],
       [2., 3., 9.],
       [5., 8., 7.],
       [1., 5., 5.]])
```

## Generating Higher-Order Polynomial Features

Scikit-learn has a module called `PolynomialFeatures` for generating a new dataset containing high-order polynomial and interaction features based off the features in the original dataset. For example, if the original dataset has two dimensions  $[a, b]$ , the second-degree polynomial transformation of the features will result in  $[1, a, b, a^2, ab, b^2]$ .

```
# import packages
from sklearn.preprocessing import PolynomialFeatures

# create dataset
data = np.array([[5,8],[9,3],[8,6],
                 [5,2],[3,9],[8,7],
                 [1,5]])

data
'Output':
array([[5, 8],
       [9, 3],
       [8, 6],
       [5, 2],
       [3, 9],
       [8, 7],
       [1, 5]])
```