LAB13

August 25, 2019

[2]:
```
import numpy as np
import pandas as pd
from sklearn.impute import SimpleImputer
```

```
/home/rahul/anaconda3/lib/python3.7/importlib/_bootstrap.py:219: RuntimeWarning:
numpy.ufunc size changed, may indicate binary incompatibility. Expected 216, got
192
  return f(*args, **kwds)
/home/rahul/anaconda3/lib/python3.7/importlib/_bootstrap.py:219: RuntimeWarning:
numpy.ufunc size changed, may indicate binary incompatibility. Expected 216, got
192
  return f(*args, **kwds)
```

[3]:
```
datasets = pd.read_csv('Data_for_Missing_Values.csv')
print("\n\nData :\n",datasets)
print("\n\nData Statistics :\n",datasets.describe())
```

```
Data :
     Country   Age   Salary Purchased
0    France  44.0  72000.0        No
1     Spain  27.0  48000.0       Yes
2   Germany  30.0  54000.0        No
3     Spain  38.0  61000.0        No
4       NaN   NaN      NaN       NaN
5   Germany  40.0      NaN       Yes
6    France  35.0  58000.0       Yes
7     Spain   NaN  52000.0        No
8    France  48.0  79000.0       Yes
9   Germany  50.0  83000.0        No
10   France  37.0  67000.0       Yes
11    Spain  45.0  55000.0        No


Data Statistics :
```

```
              Age        Salary
count   10.000000     10.000000
mean    39.400000  62900.000000
std      7.515909  11892.574714
min     27.000000  48000.000000
25%     35.500000  54250.000000
50%     39.000000  59500.000000
75%     44.750000  70750.000000
max     50.000000  83000.000000
```

[4]:
```
X = datasets.iloc[:, :-1].values
Y = datasets.iloc[:, -1].values
print("\n\nInput : \n", X)
print("\n\nOutput: \n", Y)
```

```
Input :
 [['France' 44.0 72000.0]
 ['Spain' 27.0 48000.0]
 ['Germany' 30.0 54000.0]
 ['Spain' 38.0 61000.0]
 [nan nan nan]
 ['Germany' 40.0 nan]
 ['France' 35.0 58000.0]
 ['Spain' nan 52000.0]
 ['France' 48.0 79000.0]
 ['Germany' 50.0 83000.0]
 ['France' 37.0 67000.0]
 ['Spain' 45.0 55000.0]]


Output:
 ['No' 'Yes' 'No' 'No' nan 'Yes' 'Yes' 'No' 'Yes' 'No' 'Yes' 'No']
```

[5]:
```
datasets.dropna(how='all',inplace=True)
print("\nNew Data :",datasets)
```

```
New Data :     Country   Age   Salary Purchased
0    France  44.0  72000.0        No
1     Spain  27.0  48000.0       Yes
2   Germany  30.0  54000.0        No
3     Spain  38.0  61000.0        No
5   Germany  40.0      NaN       Yes
6    France  35.0  58000.0       Yes
7     Spain   NaN  52000.0        No
```

```
8    France  48.0  79000.0       Yes
9   Germany  50.0  83000.0        No
10   France  37.0  67000.0       Yes
11    Spain  45.0  55000.0        No
```

[6]:
```
new_X = datasets.iloc[:, :-1].values
new_Y = datasets.iloc[:, -1].values
imputer = SimpleImputer(missing_values = np.nan,strategy = "mean")
imputer = imputer.fit(new_X[:, 1:3])
new_X[:, 1:3] = imputer.transform(new_X[:, 1:3])
print("\n\nNew Input with Mean Value for NaN : \n\n", new_X)
```

```
New Input with Mean Value for NaN :

 [['France' 44.0 72000.0]
 ['Spain' 27.0 48000.0]
 ['Germany' 30.0 54000.0]
 ['Spain' 38.0 61000.0]
 ['Germany' 40.0 62900.0]
 ['France' 35.0 58000.0]
 ['Spain' 39.4 52000.0]
 ['France' 48.0 79000.0]
 ['Germany' 50.0 83000.0]
 ['France' 37.0 67000.0]
 ['Spain' 45.0 55000.0]]

/home/rahul/anaconda3/lib/python3.7/importlib/_bootstrap.py:219: RuntimeWarning:
numpy.ufunc size changed, may indicate binary incompatibility. Expected 192 from
C header, got 216 from PyObject
  return f(*args, **kwds)
```

[ ]:
```
```