Class_01.04.2023_Over_Under_Fitting

April 1, 2023

Classifying the object

- Shape
- Radius
- Weight
- Usage

Classifying the Ball

- Shape - Round
- Radius - 5.6 cm
- Weight - 150 gm to 250 gm
- Usage - Play

New Data

- Shape - Round
- Radius - 7.6 cm
- Weight - 150 gm to 250 gm
- Usage - Play

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
```

[3]:
```
df = pd.DataFrame(data = [['round',5.6,175,'play','ball'],
                          ['round',5.6,180,'play','ball'],
                          ['triange',8.6,189,'play','boomerang'],
                          ['round',5.6,176,'play','ball'],
                          ['triange',8.7,190,'play','boomerang'],
                          ['round',5.6,165,'play','ball'],
                          ['round',5.6,170,'play','ball'],
                          ['round',5.6,173,'play','ball'],
                          ['round',5.6,192,'play','ball'],
                          ['round',5.6,220,'play','ball'],
                          ['triange',7.6,188,'play','boomerang'],
```

```
                          ['round',5.6,205,'play','ball'],
                          ['round',5.6,193,'play','ball'],
                          ['round',5.6,171,'play','ball'],
                          ['round',5.6,165,'play','ball'],
                          ['round',5.6,185,'play','ball']],columns =␣
↪['shape','radius','weight','usage','label'])
```

[4]:
```
df.head()
```

[4]:

	shape	radius	weight	usage	label
0	round	5.6	175	play	ball
1	round	5.6	180	play	ball
2	triange	8.6	189	play	boomerang
3	round	5.6	176	play	ball
4	triange	8.7	190	play	boomerang

[8]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
```

[9]:
```
objlst = df.select_dtypes(include = 'object').columns
```

[10]:
```
for i in objlst:
    df[i] = le.fit_transform(df[i])
```

[11]:
```
df.head()
```

[11]:

	shape	radius	weight	usage	label
0	0	5.6	175	0	0
1	0	5.6	180	0	0
2	1	8.6	189	0	1
3	0	5.6	176	0	0
4	1	8.7	190	0	1

[12]:
```
x = df.iloc[:,0:-1]
y = df.iloc[:,-1]
```

[13]:
```
x
```

[13]:

	shape	radius	weight	usage
0	0	5.6	175	0
1	0	5.6	180	0
2	1	8.6	189	0
3	0	5.6	176	0
4	1	8.7	190	0
5	0	5.6	165	0
6	0	5.6	170	0
7	0	5.6	173	0
8	0	5.6	192	0

```
9       0     5.6     220      0
10      1     7.6     188      0
11      0     5.6     205      0
12      0     5.6     193      0
13      0     5.6     171      0
14      0     5.6     165      0
15      0     5.6     185      0
```

[14]:
```
y
```

[14]:
```
0     0
1     0
2     1
3     0
4     1
5     0
6     0
7     0
8     0
9     0
10    1
11    0
12    0
13    0
14    0
15    0
Name: label, dtype: int32
```

[ ]:
```

```