

candy_chocolate_LogisticRegression

April 24, 2018

```
In [43]: # chocolate classification Logistic
%matplotlib inline
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
```

```
In [46]: df = pd.read_csv('candy-data.csv')
df.tail()
```

```
Out[46]:
```

	competitorname	chocolate	fruity	caramel	peanutyalmondy	\
80	Twizzlers	0	1	0	0	
81	Warheads	0	1	0	0	
82	Welch's Fruit Snacks	0	1	0	0	
83	Werther's Original Caramel	0	0	1	0	
84	Whoppers	1	0	0	0	

	nougat	crispedricewafer	hard	bar	pluribus	sugarpercent	pricepercent	\
80	0	0	0	0	0	0.220	0.116	
81	0	0	1	0	0	0.093	0.116	
82	0	0	0	0	1	0.313	0.313	
83	0	0	1	0	0	0.186	0.267	
84	0	1	0	0	1	0.872	0.848	

	winpercent
80	45.466282
81	39.011898
82	44.375519
83	41.904308
84	49.524113

```
In [27]: y = df['chocolate']
x = pd.DataFrame(df, columns=['pricepercent', 'sugarpercent', 'winpercent'])
x.head()
```

```
Out[27]:
```

	pricepercent	sugarpercent	winpercent
0	0.860	0.732	66.971725
1	0.511	0.604	67.602936

2	0.116	0.011	32.261086
3	0.511	0.011	46.116505
4	0.511	0.906	52.341465

```
In [31]: train_x, test_x, train_y, test_y = train_test_split(x, y)
```

```
In [34]: lgr = LogisticRegression()
         lgr.fit(train_x, train_y)
```

```
Out[34]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                             intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
                             penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
                             verbose=0, warm_start=False)
```

```
In [36]: lgr.score(test_x, test_y)
```

```
Out[36]: 0.6818181818181818
```

```
In [40]: lgr.predict(test_x)
```

```
Out[40]: array([0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0])
```

```
In [56]: # plt.plot(x['pricepercent'], y)
```