Logistic Regression Classifier with Artificial Generated

In [1]:

```
import pandas as pd
```

In [2]:

```
import numpy as np
```

In [3]:

```
import matplotlib.pyplot as plt
```

In [4]:

```
from sklearn.datasets import make_classification
```

In [5]:

```
#without coefficient of underline model

X, y = make_classification(n_samples = 1000, n_features = 5 ,
                           n_clusters_per_class = 1, n_classes =2,
                           random_state = 2529)
```

In [6]:

```
X[0:5]
```

Out[6]:

```
array([[ 1.54701705,  0.84770596, -0.41725021, -0.62356778, -0.19388577],
       [ 0.80633556,  0.40985594, -0.45641095, -0.3052022 ,  0.50935923],
       [ 0.94390268,  0.70041038,  1.11385452, -0.49394417,  1.42305455],
       [ 1.92091517,  0.95815739, -1.2235022 , -0.71578154,  0.66588981],
       [ 1.45270369,  0.69035375, -1.18119669, -0.52009219, -0.22745417]])
```

In [7]:

```
y[0:5]
```

Out[7]:

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

In [8]:

```
X.shape, y.shape
```

Out[8]:

```
((1000, 5), (1000,))
```

In [9]:

```
from sklearn.model_selection import train_test_split
```

In [10]:

```
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 0.3,
                                                    random_state = 2529)
```

In [11]:

```
X_train.shape,X_test.shape,y_train.shape,y_test.shape
```

Out[11]:

```
((700, 5), (300, 5), (700,), (300,))
```

In [12]:

```
from sklearn.linear_model import LogisticRegression
```

In [13]:

```
model = LogisticRegression()
```

In [14]:

```
model.fit(X_train, y_train)
```

Out[14]:

```
LogisticRegression()
```

In [15]:

```
y_pred = model.predict(X_test)
```

In [16]:

```
y_pred.shape
```

Out[16]:

```
(300,)
```

In [17]:

```
y_pred
```

Out[17]:

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

In [18]:

```
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

In [19]:

```
accuracy_score(y_test, y_pred)
```

Out[19]:

```
0.99
```

In [20]:

```
confusion_matrix(y_test, y_pred)
```

Out[20]:

```
array([[156,   1],
       [  2, 141]], dtype=int64)
```

In [21]:

```
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99       157
           1       0.99      0.99      0.99       143

    accuracy                           0.99       300
   macro avg       0.99      0.99      0.99       300
weighted avg       0.99      0.99      0.99       300
```

In [22]:

```
from sklearn.model_selection import GridSearchCV
parameters = {'penalty': ['l1', 'l2'], 'C': [0.001,.009,0.01, .09,1,5,10,25],
              'solver': ['liblinear']}
gridsearch = GridSearchCV(LogisticRegression(),parameters)
gridsearch.fit(X_train, y_train)
```

Out[22]:

```
GridSearchCV(estimator=LogisticRegression(),
             param_grid={'C': [0.001, 0.009, 0.01, 0.09, 1, 5, 10, 25],
                         'penalty': ['l1', 'l2'], 'solver': ['liblinear']})
```

In [23]:

```
gridsearch.best_params_
```

Out[23]:

```
{'C': 1, 'penalty': 'l1', 'solver': 'liblinear'}
```

In [24]:

```
gridsearch.best_score_
```

Out[24]:

```
0.9914285714285714
```

In [25]:

```
gridsearch.best_estimator_
```

Out[25]:

```
LogisticRegression(C=1, penalty='l1', solver='liblinear')
```

In [26]:

```
gridsearch.best_index_
```

Out[26]:

```
8
```

In [27]:

```
y_pred_grid = gridsearch.predict(X_test)
```

In [28]:

```
confusion_matrix(y_test, y_pred_grid)
```

Out[28]:

```
array([[156,   1],
       [  2, 141]], dtype=int64)
```

In [29]:

```
print(classification_report(y_test, y_pred_grid))
```

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99       157
           1       0.99      0.99      0.99       143

    accuracy                           0.99       300
   macro avg       0.99      0.99      0.99       300
weighted avg       0.99      0.99      0.99       300
```

In [ ]: