Logistic Regression

November 15, 2022

1 Logistic Regression

[30]:

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report,confusion_matrix
from sklearn.model_selection import train_test_split
```

[31]:

```
data = pd.read_csv('diabetes.csv')
data.head()
```

[31]:

```
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
0            6      148             72             35        0  33.6
1            1       85             66             29        0  26.6
2            8      183             64              0        0  23.3
3            1       89             66             23       94  28.1
4            0      137             40             35      168  43.1

   DiabetesPedigreeFunction  Age  Outcome
0                     0.627   50        1
1                     0.351   31        0
2                     0.672   32        1
3                     0.167   21        0
4                     2.288   33        1
```

[32]:

```
x=data.drop("Outcome",axis=1)
y=data[["Outcome"]]
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.
 ↪30,random_state=42)
```

[33]:

```
model = LogisticRegression()
model.fit(x_train, y_train)
y_predict=model.predict(x_test)
model_score=model.score(x_test,y_test)
```

```
/home/samuel-adirala/anaconda3/lib/python3.9/site-
packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
```

```
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
/home/samuel-adirala/anaconda3/lib/python3.9/site-
packages/sklearn/linear_model/_logistic.py:814: ConvergenceWarning: lbfgs failed
to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
```

[34]:

```python
from sklearn.metrics import accuracy_score
print(accuracy_score(y_test, model.predict(x_test))*100)
```

```
74.02597402597402
```

[35]:

```python
to_predict_list=[1,122,90,51,220,49.7,0.325,31]
# change the input data to a numpy array

input_data_as_numpy_array= np.asarray(to_predict_list)

# reshape the numpy array as we are predicting for only on instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

prediction = model.predict(input_data_reshaped)
```

```
/home/samuel-adirala/anaconda3/lib/python3.9/site-packages/sklearn/base.py:450:
UserWarning: X does not have valid feature names, but LogisticRegression was
fitted with feature names
  warnings.warn(
```

[36]:

```python
print(prediction[0])
```

```
1
```

[37]:

```python
if(prediction[0]==1):
    print("The Person is Effected By Diabetes")
else:
    print("The Person is Effected By Diabetes")
```

```
The Person is Effected By Diabetes
```