

1. Load Dataset
 2. Divide in train_test_split
 3. Apply LogisticRegression liblinear and saga-solver
 4. Apply GridSearchCV to find best hyperparameter
 5. Evaluate it on trained model on test dataset
 6. Print Evaluation Metrics

texture (standard error): 0.36 4.885\n perimeter (standard error): 0.757 21.98\n area (standard error): 6.802 542.2\n smoothness (standard error): 0.002 0.031\n compactness (standard error):

X

```
array([[1.799e+01, 1.038e+01, 1.228e+02, ... , 2.654e-01, 4.601e-01,
       1.189e-01],
       [2.057e+01, 1.777e+01, 1.329e+02, ... , 1.860e-01, 2.750e-01,
       8.902e-02],
       [1.969e+01, 2.125e+01, 1.300e+02, ... , 2.430e-01, 3.613e-01,
       8.758e-02],
       ... ,
       [1.660e+01, 2.808e+01, 1.083e+02, ... , 1.418e-01, 2.218e-01,
       7.820e-02],
       [2.060e+01, 2.933e+01, 1.401e+02, ... , 2.650e-01, 4.087e-01,
       1.240e-01],
       [7.760e+00, 2.454e+01, 4.792e+01, ... , 0.000e+00, 2.871e-01,
       7.039e-02]])
```

y

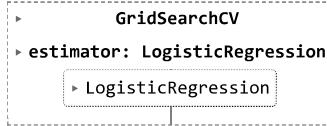
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

The `liblinear` solver is a library for large linear classification. It supports logistic regression and linear support vector machines.

```
log_reg = LogisticRegression(solver='liblinear')
```

Hyperparameters for GridSearch (GridSearchCV) are defined to find the best hyperparameters.

```
param_grid = {  
    'penalty': ['l2', 'l1'],  
    'C': [0.001, 0.01, 0.1, 1, 10, 100],  
    'fit_intercept': [True, False]  
}  
grid_search = GridSearchCV(log_reg, param_grid, cv=5)  
grid_search.fit(X_train, y_train)
```



```
# best hyperparameters  
grid_search.best_params_  
  
{'C': 100, 'fit_intercept': True, 'penalty': 'l1'}
```

C(Regularization Strength) decreases => Regularization increases => Simpler Model it comes

Regularization is used to avoid overfitting

```
for penalty in ['l1', 'l2']:
    for C in [0.001, 0.01, 0.1, 1, 10, 100]:
        for fit_intercept in [True, False]:
            log_reg = LogisticRegression(solver='liblinear', penalty=penalty, C=C, fit_intercept=fit_intercept)
            log_reg.fit(X_train, y_train)
            print("\n")
            print(f"Penalty: {penalty}, C: {C}, fit_intercept: {fit_intercept}")
            print("Coefficients:", log_reg.coef_)
            print("Intercept:", log_reg.intercept_)
```

```

0.      0.      0.      0.      0.      0.
0.      0.      0.      0.      0.      -0.009749
0.      0.      0.      0.      0.      0.      ]]
Intercept: 0.0

Penalty: l1, C: 0.01, fit_intercept: True
Coefficients: [[ 0.          0.          0.14091422  0.00724974  0.
0.          0.          0.          0.          0.
0.         -0.00651377  0.          0.          0.          0.
0.          0.          -0.00037397  0.          -0.02092428
0.          0.          0.          0.          0.          0.      ]]
Intercept: [0.]
```

```

Penalty: l1, C: 0.01, fit_intercept: False
Coefficients: [[ 0.          0.          0.14092924  0.00724521  0.
0.          0.          0.          0.          0.
0.         -0.00651707  0.          0.          0.          0.
0.          0.          -0.00039264  0.          -0.02092175
0.          0.          0.          0.          0.          0.      ]]
Intercept: 0.0
/usr/local/lib/python3.10/dist-packages/sklearn/_base.py:1244: ConvergenceWarning: Liblinear failed to converge, increase the num
warnings.warn()
```

```

Penalty: l1, C: 0.1, fit_intercept: True
Coefficients: [[ 0.30825519  0.          0.28994151 -0.00546104  0.
0.          0.          0.          0.          0.
0.         -0.04256436  0.          0.          0.          0.
0.          0.          0.32338781 -0.13823409 -0.12009132 -0.01825132
0.          0.          0.          0.          0.          0.      ]]
Intercept: [0.]
/usr/local/lib/python3.10/dist-packages/sklearn/_base.py:1244: ConvergenceWarning: Liblinear failed to converge, increase the num
warnings.warn()
```

```

Penalty: l1, C: 0.1, fit_intercept: False
Coefficients: [[ 0.37279899  0.          0.28079341 -0.00555956  0.
0.          0.          0.          0.          0.
0.         -0.04231862  0.          0.          0.          0.
0.          0.          0.2958589 -0.13816154 -0.11694135 -0.01814008
```

```
# It evaluate traind model on test dataset
y_pred = grid_search.predict(X_test)
```

```

# Evaluation Metrics
accuracy = accuracy_score(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)
classification_rep = classification_report(y_test, y_pred)

print("Evaluation Metrics:")
print("Accuracy:\n", accuracy)
print("Confusion Matrix:\n", conf_matrix)
print("Classification Report:\n", classification_rep)
```

```

Evaluation Metrics:
Accuracy:
0.9824561403508771
Confusion Matrix:
[[42  1]
 [ 1 70]]
Classification Report:
 precision    recall   f1-score   support
0          0.98     0.98     0.98      43
1          0.99     0.99     0.99      71

   accuracy           0.98      114
   macro avg       0.98     0.98     0.98      114
   weighted avg    0.98     0.98     0.98      114
```

LogisticRegression model with saga solver and elasticnet penalty.

Penalty: Tyoe of Regularization L1, L2 elasticnet

saga: Used for training large models like logistic regression for better efficiency & sometimes it overfits the model sometimes.

```
log_reg_elasticnet = LogisticRegression(solver='saga', penalty='elasticnet')
param_grid_elasticnet = {
    'C': [0.001, 0.01, 0.1, 1, 10, 100],
    'fit_intercept': [True, False],
    'l1_ratio': [0.1, 0.5, 0.9]
}
```

saga solver only supports elasticnet penalty and liblinear solver supports only L1,L2 regularization. Thus, GridSearch to find best hyperparameter is done separately for saga after liblinear.

```
grid_search_elasticnet = GridSearchCV(log_reg_elasticnet, param_grid_elasticnet, cv=5)
grid_search_elasticnet.fit(X_train, y_train)
```



```
> GridSearchCV
> estimator: LogisticRegression
    > LogisticRegression
```