Question 2

Note:

According to https://scikit-learn.org/stable/modules/linear_model.html (https://scikit-learn.org/stable/modules/linear_model.html), Ridge and LASSO Logistic Regression minimize the following objective functions respectively.

$$\min_{w,c} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \log(\exp(-y_i(X_i^T w + c)) + 1).$$

$$\min_{w,c} \|w\|_1 + C \sum_{i=1}^{n} \log(\exp(-y_i(X_i^T w + c)) + 1).$$

Tuning parameter C describes the inverse of regularization strength.

(a) Logistic Regression Model

Coefficients and confusion matrix for the test set are shown as follows.

```
Mean accuracy on test set is 0.9911504424778761

Coefficients are

[[ -4.63795182  15.51467914 -39.76227664 -27.97597849  72.39664305
   15.30168863   2.73402449 -13.49102549  26.58767053  22.79400755
   -1.58036096]]

Confusion matrix on test set is

[[103   1]
 [  0   9]]
```

(b) Ridge Logistic Regression Model

Optimal tuning parameter, optimal coefficients and the confusion matrix for the test set are shown as follows.

```
Mean accuracy on test set is 0.9911504424778761

Optimal Coefficients are

[[-0.44177597  0.52264659 -2.01860673 -1.83188506  3.18919965  1.13829431
   0.23507703 -0.63224381  1.43045919  1.73830259 -0.15760907]]

Optimal tuning parameter is 2.782559402207126

Confusion matrix on test set is

[[103   1]
 [  0   9]]
```

(c) LASSO Logistic Regression Model

Optimal tuning parameter, optimal coefficients and the confusion matrix for the test set are shown as follows.

```
Mean accuracy on test set is 0.9911504424778761

Optimal Coefficients are

[[-0.39758761  0.74330282 -2.53531895 -2.08584311  5.00152154  0.61646927
   0.05863434 -0.51958877  1.8041549   1.83936585  0.        ]]

Optimal tuning parameter is 2.782559402207126

Confusion matrix on test set is

[[103   1]
 [  0   9]]
```

(d) Adaptive LASSO Logistic Regression Model

Optimal tuning parameter, optimal coefficients and the confusion matrix for the test set are shown as follows.

```
Mean accuracy on test set is 0.9823008849557522

Optimal Coefficients are

[[-0.          0.         -1.2638265  -1.23438995  3.00867754  0.22642313
   0.         -0.          0.85101359  0.97340566 -0.        ]]

Optimal tuning parameter is 0.3593813663804626

Confusion matrix on test set is

[[103   1]
 [  1   8]]
```

In [42]:
```
import sklearn.linear_model as lm
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import confusion_matrix
from sklearn.linear_model import LogisticRegressionCV
```

In [100]:
```
#logistic regression

data = pd.read_csv('Question2-3.csv').to_numpy()

X_train = data[:250,:11]
X_test = data[250:,:11]
y_train = data[:250,11]
y_test = data[250:,11]

scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

clf = LogisticRegression(penalty='none').fit(X_train, y_train)
y_pred = clf.predict(X_test)
print(f'Mean accuracy on test set is {clf.score(X_test,y_test)}\n')
print(f'Coefficients are \n\n{clf.coef_}\n')
print(f'Confusion matrix on test set is \n\n{confusion_matrix(y_test,
y_pred)}')
```

```
Mean accuracy on test set is 0.9911504424778761

Coefficients are

[[ -4.63795182  15.51467914 -39.76227664 -27.97597849  72.39664305
   15.30168863   2.73402449 -13.49102549  26.58767053  22.79400755
   -1.58036096]]

Confusion matrix on test set is

[[103   1]
 [  0   9]]
```

In [108]:
```
#Ridge logistic regression

data = pd.read_csv('Question2-3.csv').to_numpy()

X_train = data[:250,:11]
X_test = data[250:,:11]
y_train = data[:250,11]
y_test = data[250:,11]

scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

clf = LogisticRegressionCV(penalty='l2',cv=5).fit(X_train, y_train)
y_pred = clf.predict(X_test)
print(f'Mean accuracy on test set is {clf.score(X_test,y_test)}\n')
print(f'Optimal Coefficients are \n\n{clf.coef_}\n')
print(f'Optimal tuning parameter is {clf.C_[0]}\n')
print(f'Confusion matrix on test set is \n\n{confusion_matrix(y_test,
y_pred)}')
```

```
Mean accuracy on test set is 0.9911504424778761

Optimal Coefficients are

[[-0.44177597  0.52264659 -2.01860673 -1.83188506  3.18919965  1.138
29431
   0.23507703 -0.63224381  1.43045919  1.73830259 -0.15760907]]

Optimal tuning parameter is 2.782559402207126

Confusion matrix on test set is

[[103   1]
 [  0   9]]
```

In [109]:
```
#Lasso logistic regression

data = pd.read_csv('Question2-3.csv').to_numpy()

X_train = data[:250,:11]
X_test = data[250:,:11]
y_train = data[:250,11]
y_test = data[250:,11]

scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

clf = LogisticRegressionCV(penalty='l1',solver='liblinear',cv=5).fit(X
_train, y_train)
y_pred = clf.predict(X_test)
print(f'Mean accuracy on test set is {clf.score(X_test,y_test)}\n')
print(f'Optimal Coefficients are \n\n{clf.coef_}\n')
print(f'Optimal tuning parameter is {clf.C_[0]}\n')
print(f'Confusion matrix on test set is \n\n{confusion_matrix(y_test,
y_pred)}')
```

```
Mean accuracy on test set is 0.9911504424778761

Optimal Coefficients are

[[-0.39758761  0.74330282 -2.53531895 -2.08584311  5.00152154  0.616
46927
   0.05863434 -0.51958877  1.8041549   1.83936585  0.        ]]

Optimal tuning parameter is 2.782559402207126

Confusion matrix on test set is

[[103   1]
 [  0   9]]
```

In [110]:
```
#adaptive lasso logistic regression

data = pd.read_csv('Question2-3.csv').to_numpy()

X_train = data[:250,:11]
X_test = data[250:,:11]
y_train = data[:250,11]
y_test = data[250:,11]

scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

gamma=1

ridge_betas = LogisticRegressionCV(penalty='l2',cv=5).fit(X_train, y_t
rain).coef_
w_ridge = ridge_betas**-gamma
X_ridge = X_train/w_ridge
X_ridge_test = X_test/w_ridge

clf = LogisticRegressionCV(penalty='l1',solver='liblinear',cv=5).fit(X
_ridge, y_train)
op_coef = clf.coef_/w_ridge
op_c = clf.C_[0]
y_pred = clf.predict(X_ridge_test)
print(f'Mean accuracy on test set is {clf.score(X_ridge_test,y_test)}\
n')
print(f'Optimal Coefficients are \n\n{op_coef}\n')
print(f'Optimal tuning parameter is {op_c}\n')
print(f'Confusion matrix on test set is \n\n{confusion_matrix(y_test,
y_pred)}')
```

```
Mean accuracy on test set is 0.9823008849557522

Optimal Coefficients are

[[-0.          0.         -1.2638265  -1.23438995  3.00867754  0.226
42313
   0.         -0.          0.85101359  0.97340566 -0.        ]]

Optimal tuning parameter is 0.3593813663804626

Confusion matrix on test set is

[[103   1]
 [  1   8]]
```

In [ ]: