

In [232]:

```
# Tasks — Diagnostics
import numpy as np
import pandas as pd
import scipy.optimize
import random
import matplotlib.pyplot as plt
import sklearn.linear_model
from sklearn.utils import shuffle
data = open("5year.arff", 'r')
```

In [233]:

```
while not '@data' in data.readline():
    pass
dataset = []
for element in data:
    if '?' in element: # Missing entry
        continue
    element = element.split(',')
    values = [1] + [float(x) for x in element]
    values[-1] = values[-1] > 0
    dataset.append(values)
```

In [234]:

```

# Prepare our data variables and outcomes
x = [d[:-1] for d in dataset]
y = [d[-1] for d in dataset]
y = [1 if i == True else 0 for i in y ]

LogisticRegressor = linear_model.LogisticRegression(C=1)
LogisticRegressor.fit(x,y)
yhat = LogisticRegressor.predict(x)
Acc = sum(yhat == y)/len(y)
print('Accuracy: ',Acc)

TruePo = sum(np.logical_and(yhat, y))
FalsePo = sum(np.logical_and(yhat, np.logical_not(y)))
TrueNe = sum(np.logical_and(np.logical_not(yhat), np.logical_not(y)))
FalseNe = sum(np.logical_and(np.logical_not(yhat), y))

# True Positive Rate And True Negative Rate
TPR = TruePo/(TruePo + FalseNe)
TNR = TrueNe/(TrueNe + FalsePo)
# BER
BER = 1 - 0.5 * (TPR + TNR)
print('BER: ',BER)

# Answer to Problem 1:
# Accuracy: 0.9663477400197954, BER: 0.4810749837661251

```

```

//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
  FutureWarning)

```

```

Accuracy:  0.9663477400197954
BER:  0.4810749837661251

```

```

//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
  "the number of iterations.", ConvergenceWarning)

```

In [235]:

```

# Answer to Problem 1:
# Accuracy: 0.9663477400197954, BER: 0.4810749837661251

```

In [236]:

```
# Problem 3: Shuffle Data
datSet = list(zip(x,y))
random.shuffle(datSet)
# Split Data into Training, Validation, and Test
xP3 = [d[0] for d in datSet]
yP3 = [d[1] for d in datSet]

n1stBreak = np.int(len(yP3)*0.5)
n2ndBreak = np.int(len(yP3)*0.75)

xtraining = xP3[:n1stBreak]
xvalidation = xP3[n1stBreak:n2ndBreak]
xtest = xP3[n2ndBreak:]
ytraining = yP3[:n1stBreak]
yvalidation = yP3[n1stBreak:n2ndBreak]
ytest = yP3[n2ndBreak:]
```

In [237]:

```
LogisticModel1 = linear_model.LogisticRegression(C = 1.0, class_weight = 'balanced')
LogisticModel1.fit(xtraining, ytraining)

//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
"the number of iterations.", ConvergenceWarning)
```

Out[237]:

```
LogisticRegression(C=1.0, class_weight='balanced', dual=False,
                    fit_intercept=True, intercept_scaling=1, l1_ratio
=None,
                    max_iter=100, multi_class='warn', n_jobs=None, pe
nalty='l2',
                    random_state=None, solver='warn', tol=0.0001, ver
bose=0,
                    warm_start=False)
```

In [238]:

```
ytraininghat = LogisticModel1.predict(xtraining)
yvalidationhat = LogisticModel1.predict(xvalidation)
ytesthat = LogisticModel1.predict(xtest)
```

In [239]:

```
AccTrain = sum(ytraininghat == ytraining) / len(ytraining)
AccValid = sum(yvalidationhat == yvalidation) / len(yvalidation)
AccTest = sum(ytestthat == ytest) / len(ytest)
# Report Accuracy for 3 sets
print("Training accuracy: ", AccTrain)
print("Validation accuracy: ", AccValid)
print("Test accuracy: ", AccTest)
```

```
Training accuracy:  0.7722772277227723
Validation accuracy: 0.7427440633245382
Test accuracy:     0.7691292875989446
```

In [240]:

```
# Report BER for 3 sets
```

In [241]:

```
TPP3Tr = sum(np.logical_and(ytraininghat, ytraining))
FPP3Tr = sum(np.logical_and(ytraininghat, np.logical_not(ytraining)))
TNP3Tr = sum(np.logical_and(np.logical_not(ytraininghat), np.logical_not(ytraining)))
FNP3Tr = sum(np.logical_and(np.logical_not(ytraininghat), ytraining))
BERTr = 1 - 0.5 * (TPP3Tr / (TPP3Tr + FNP3Tr) + TNP3Tr / (TNP3Tr + FPP3Tr))
print('BER For Training: ', BERTr)
```

```
BER For Training:  0.2291918607708081
```

In [242]:

```
TPP3v = sum(np.logical_and(yvalidationhat, yvalidation))
FPP3v = sum(np.logical_and(yvalidationhat, np.logical_not(yvalidation)))
TNP3v = sum(np.logical_and(np.logical_not(yvalidationhat), np.logical_not(yvalidation)))
FNP3v = sum(np.logical_and(np.logical_not(yvalidationhat), yvalidation))
BERv = 1 - 0.5 * (TPP3v / (TPP3v + FNP3v) + TNP3v / (TNP3v + FPP3v))
print('BER For Validation: ', BERv)
```

```
BER For Validation: 0.2780487804878049
```

In [243]:

```
TPP3Te = sum(np.logical_and(ytestthat, ytest))
FPP3Te = sum(np.logical_and(ytestthat, np.logical_not(ytest)))
TNP3Te = sum(np.logical_and(np.logical_not(ytestthat), np.logical_not(ytest)))
FNP3Te = sum(np.logical_and(np.logical_not(ytestthat), ytest))
BERTe = 1 - 0.5 * (TPP3Te / (TPP3Te + FNP3Te) + TNP3Te / (TNP3Te + FPP3Te))
print('BER For Test: ', BERTe)
```

```
BER For Test: 0.24803113553113554
```

In [244]:

```
# Problem 4: Implement a complete regularization pipeline with the balanced classifier
def BERCulator(x,y,c):
    LogisticModel = linear_model.LogisticRegression(C = c, class_weight = 'balanced')
    LogisticModel.fit(x, y)
    yhat = LogisticModel.predict(x)
    TP = sum(np.logical_and(yhat, y))
    FP = sum(np.logical_and(yhat, np.logical_not(y)))
    TN = sum(np.logical_and(np.logical_not(yhat), np.logical_not(y)))
    FN = sum(np.logical_and(np.logical_not(yhat), y))
    BER = 1 - 0.5 * (TP / (TP + FN) + TN / (TN + FP))
    Acc = sum(yhat == y) / len(y)
    return (BER,Acc)
```

In [245]:

```
BERtrainlist = []
BERvalidationlist = []
BERtestlist = []
Acctrainlist = []
Accvalidationlist = []
Acctestlist = []

for c in (10**n for n in range(-4,5)):
    BERtrainlist.append(BERCalculator(xtraining, ytraining,c)[0])
    BERvalidationlist.append(BERCalculator(xvalidation, yvalidation,c)[0])
    BERtestlist.append(BERCalculator(xtest, ytest,c)[0])
    Acctrainlist.append(BERCalculator(xtraining, ytraining,c)[1])
    Accvalidationlist.append(BERCalculator(xvalidation, yvalidation,c)[1])
    Acctestlist.append(BERCalculator(xtest, ytest,c)[1])
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
"the number of iterations.", ConvergenceWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
"the number of iterations.", ConvergenceWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
```

```
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
```





10/22

11/22

```
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
"the number of iterations.", ConvergenceWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
"the number of iterations.", ConvergenceWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
"the number of iterations.", ConvergenceWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
"the number of iterations.", ConvergenceWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
"the number of iterations.", ConvergenceWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
"the number of iterations.", ConvergenceWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
```

```
f iterations.
    "the number of iterations.", ConvergenceWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
    FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
    FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
    "the number of iterations.", ConvergenceWarning)
```

In [246]:

```
# from matplotlib import pyplot as plt
# xAxis = [c for c in range(-4,5)]
# plt.plot(xAxis)
# plt.xlabel("c value: 10^-4 to 10^4")
# plt.ylabel("BER Value")
# plt.plot(xAxis, BERtrainlist, label='Training plot')
# plt.plot(xAxis, BERvalidationlist, label='Validation plot')
# plt.plot(xAxis, BERtestlist, label='Test plot')
# plt.legend()
outcome = {'c':[c for c in (10**n for n in range(-4,5))], 'Training BER':BERtrainlist,
           'Validation BER':BERvalidationlist, 'Test BER':BERtestlist, 'Train Acc':Acctrainlist,
           'Validation Acc':Accvalidationlist, 'Test Acc':Acctestlist}
outcome = pd.DataFrame(outcome)
print(outcome)
```

	c	Training BER	Validation BER	Test BER	Train Acc \
0	0.0001	0.252182	0.217818	0.265201	0.763696
1	0.0010	0.229921	0.202913	0.209386	0.788779
2	0.0100	0.242567	0.203591	0.161630	0.764356
3	0.1000	0.235048	0.126558	0.159570	0.778878
4	1.0000	0.229192	0.115041	0.161630	0.772277
5	10.0000	0.223040	0.175881	0.163690	0.784158
6	100.0000	0.229534	0.198848	0.159570	0.771617
7	1000.0000	0.237440	0.207656	0.156136	0.774257
8	10000.0000	0.243980	0.179268	0.162317	0.779538

  

	Validation Acc	Test Acc
0	0.812665	0.736148
1	0.841689	0.812665
2	0.840369	0.843008
3	0.848285	0.846966
4	0.870712	0.843008
5	0.846966	0.839050
6	0.849604	0.846966
7	0.832454	0.853562
8	0.840369	0.841689

In [247]:

```
# According to the output table, I would choose  $c = 0.1$  which gives the smallest
# BER in general. Also, the accuracy
# perform not bad under this C-value
```

In [248]:

```
# Using the given code, compute the  $F\beta$  score (on the test set) of your
# (unweighted) classifier, for  $\beta = 1$  and  $\beta = 10$ . Following this, identify weight
# vectors that yield better
# performance (compared to the unweighted vector) in terms of the F1 and F10 scores (2 marks).
```

```
# Equal Weight
weights = [1.0] * len(ytraining)
mod = linear_model.LogisticRegression(C=1, solver='lbfgs')
mod.fit(xtraining, ytraining, sample_weight=weights)
from sklearn.metrics import f1_score
ytestthat = mod.predict(xtest)
retr = sum(ytestthat)
rele = sum(ytest)
intersect = sum([y and yhat for y,yhat in zip(ytest,ytestthat)])

p = intersect / retr
r = intersect / rele

F1 = (2)*(p*r)/(p+r)
F10 = (1+100)*(p*r)/(100*p+r)
print(" $\beta = 1$ :  $F\beta =$ ", F1)
print(" $\beta = 10$ :  $F\beta =$ ", F10)
```

```
 $\beta = 1$ :  $F\beta =$  0.0625
 $\beta = 10$ :  $F\beta =$  0.03364423717521652
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:947: ConvergenceWarning: lbfgs failed to converge. Increase the
number of iterations.
"of iterations.", ConvergenceWarning)
```

In [249]:

```
# Diff Weight with weight = 1/0.1 for elements in ytraining
weights = [1 if e == 1 else 0.1 for e in ytraining]
modl = linear_model.LogisticRegression(C=1, solver='lbfgs')
modl.fit(xtraining, ytraining, sample_weight=weights)
from sklearn.metrics import f1_score
ytestthat1 = modl.predict(xtest)
retr1 = sum(ytestthat1)
rele1 = sum(ytest)
intersect1 = sum([y and yhat for y,yhat in zip(ytest,ytestthat1)])

p1 = intersect1 / retr1
r1 = intersect1 / rele1

F1 = (2)*(p1*r1)/(p1+r1)
F10 = (1+100)*(p1*r1)/(100*p1+r1)
print("β = 1: Fβ = ", F1)
print("β = 10: Fβ = ", F10)
```

```
β = 1: Fβ = 0.16393442622950818
β = 10: Fβ = 0.1666116793137578
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:947: ConvergenceWarning: lbfgs failed to converge. Increase the
number of iterations.
"of iterations.", ConvergenceWarning)
```

In [250]:

```
from sklearn.decomposition import PCA
pcamod = PCA(n_components=5)
pcamod.fit(xtraining)
print(pcamod.components_[0])
```

```
[ 4.49869740e-23 -1.74360868e-08 1.69429309e-07 1.08274698e-06
 6.08399702e-06 7.22707073e-04 -6.57326606e-07 1.60767757e-06
 6.92153239e-06 -7.54696399e-07 2.61353638e-08 2.43845215e-07
 1.25415476e-06 5.26433455e-07 1.60767757e-06 -7.91770162e-03
 1.17799013e-06 7.43299641e-06 1.60767757e-06 5.69716749e-07
 4.57842709e-05 -1.68986282e-07 2.25957198e-07 5.18769644e-07
 7.88227482e-07 8.33584857e-07 1.06521675e-06 -4.82127778e-05
 2.70762518e-05 3.85574775e-06 -1.83665254e-06 5.61949764e-07
 -4.54882891e-04 4.14435025e-06 -1.74550755e-06 2.01931669e-07
 -1.10811369e-06 5.92166223e-03 -4.93110833e-07 2.32084975e-07
 3.10649128e-06 -3.46365555e-06 4.77623850e-07 5.75592251e-05
 1.17737406e-05 -3.47460694e-06 4.57166043e-06 -6.74175825e-05
 2.66762839e-07 5.20134845e-07 5.14162157e-06 -7.83118623e-07
 -1.20901904e-06 2.72712040e-06 2.70856736e-05 9.99950703e-01
 2.38384662e-07 -7.89714288e-07 -3.43163171e-07 7.17012708e-07
 -1.77702849e-04 -1.11256991e-05 -2.44502741e-04 5.46613372e-06
 -1.24285974e-05]
```

In [251]:

```
Xpca_train = np.matmul(xtraining, pcamod.components_.T)
Xpca_valid = np.matmul(xvalidation, pcamod.components_.T)
Xpca_test = np.matmul(xtest, pcamod.components_.T)
```

In [252]:

```
def PCA_BER_Cal(component):
    pcamod = PCA(n_components = component)
    pcamod.fit(xtraining)
    print(pcamod.components_[0])
    Xpca_train = np.matmul(xtraining, pcamod.components_.T)
    Xpca_valid = np.matmul(xvalidation, pcamod.components_.T)
    Xpca_test = np.matmul(xtest, pcamod.components_.T)
    LogisticModelForPCA = linear_model.LogisticRegression(C = 1.0, class_weight
= 'balanced')
    LogisticModelForPCA.fit(Xpca_train, ytraining)
    BERValPCA, AccValPCA = BERCalculator(Xpca_valid, yvalidation, 1)
    BERTestPCA, AccTestPCA = BERCalculator(Xpca_test, ytest, 1)
    return (BERValPCA, BERTestPCA)
```



In [253]:

```
pcaBERList = []  
  
for n in range(5,31,5):  
    pcaBERList.append(PCA_BER_Cal(n))
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.  
FutureWarning)  
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.  
FutureWarning)  
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.  
FutureWarning)
```

```
[ 4.82131165e-24 -1.74360868e-08 1.69429309e-07 1.08274698e-06
 6.08399702e-06 7.22707073e-04 -6.57326606e-07 1.60767757e-06
 6.92153239e-06 -7.54696399e-07 2.61353638e-08 2.43845215e-07
 1.25415476e-06 5.26433455e-07 1.60767757e-06 -7.91770162e-03
 1.17799013e-06 7.43299641e-06 1.60767757e-06 5.69716749e-07
 4.57842709e-05 -1.68986282e-07 2.25957198e-07 5.18769644e-07
 7.88227482e-07 8.33584857e-07 1.06521675e-06 -4.82127778e-05
 2.70762518e-05 3.85574775e-06 -1.83665254e-06 5.61949764e-07
 -4.54882891e-04 4.14435025e-06 -1.74550755e-06 2.01931669e-07
 -1.10811369e-06 5.92166223e-03 -4.93110833e-07 2.32084975e-07
 3.10649128e-06 -3.46365555e-06 4.77623850e-07 5.75592251e-05
 1.17737406e-05 -3.47460694e-06 4.57166043e-06 -6.74175825e-05
 2.66762839e-07 5.20134845e-07 5.14162157e-06 -7.83118623e-07
 -1.20901904e-06 2.72712040e-06 2.70856736e-05 9.99950703e-01
 2.38384662e-07 -7.89714288e-07 -3.43163171e-07 7.17012708e-07
 -1.77702849e-04 -1.11256991e-05 -2.44502741e-04 5.46613372e-06
 -1.24285974e-05]
[-6.18663533e-22 -1.74360868e-08 1.69429309e-07 1.08274698e-06
 6.08399702e-06 7.22707073e-04 -6.57326606e-07 1.60767757e-06
 6.92153239e-06 -7.54696399e-07 2.61353638e-08 2.43845215e-07
 1.25415476e-06 5.26433455e-07 1.60767757e-06 -7.91770162e-03
 1.17799013e-06 7.43299641e-06 1.60767757e-06 5.69716749e-07
 4.57842709e-05 -1.68986282e-07 2.25957198e-07 5.18769644e-07
 7.88227482e-07 8.33584857e-07 1.06521675e-06 -4.82127778e-05
 2.70762518e-05 3.85574775e-06 -1.83665254e-06 5.61949764e-07
 -4.54882891e-04 4.14435025e-06 -1.74550755e-06 2.01931669e-07
 -1.10811369e-06 5.92166223e-03 -4.93110833e-07 2.32084975e-07
 3.10649128e-06 -3.46365555e-06 4.77623850e-07 5.75592251e-05
 1.17737406e-05 -3.47460694e-06 4.57166043e-06 -6.74175825e-05
 2.66762839e-07 5.20134845e-07 5.14162157e-06 -7.83118623e-07
 -1.20901904e-06 2.72712040e-06 2.70856736e-05 9.99950703e-01
 2.38384662e-07 -7.89714288e-07 -3.43163171e-07 7.17012708e-07
 -1.77702849e-04 -1.11256991e-05 -2.44502741e-04 5.46613372e-06
 -1.24285974e-05]
[-1.16299715e-22 -1.74360868e-08 1.69429309e-07 1.08274698e-06
 6.08399702e-06 7.22707073e-04 -6.57326606e-07 1.60767757e-06
 6.92153239e-06 -7.54696399e-07 2.61353638e-08 2.43845215e-07
 1.25415476e-06 5.26433455e-07 1.60767757e-06 -7.91770162e-03
 1.17799013e-06 7.43299641e-06 1.60767757e-06 5.69716749e-07
 4.57842709e-05 -1.68986282e-07 2.25957198e-07 5.18769644e-07
 7.88227482e-07 8.33584857e-07 1.06521675e-06 -4.82127778e-05
 2.70762518e-05 3.85574775e-06 -1.83665254e-06 5.61949764e-07
 -4.54882891e-04 4.14435025e-06 -1.74550755e-06 2.01931669e-07
 -1.10811369e-06 5.92166223e-03 -4.93110833e-07 2.32084975e-07
 3.10649128e-06 -3.46365555e-06 4.77623850e-07 5.75592251e-05
 1.17737406e-05 -3.47460694e-06 4.57166043e-06 -6.74175825e-05
 2.66762839e-07 5.20134845e-07 5.14162157e-06 -7.83118623e-07
 -1.20901904e-06 2.72712040e-06 2.70856736e-05 9.99950703e-01
 2.38384662e-07 -7.89714288e-07 -3.43163171e-07 7.17012708e-07
 -1.77702849e-04 -1.11256991e-05 -2.44502741e-04 5.46613372e-06
 -1.24285974e-05]
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)

[-3.25806728e-22 -1.74360868e-08  1.69429309e-07  1.08274698e-06
 6.08399702e-06  7.22707073e-04 -6.57326606e-07  1.60767757e-06
 6.92153239e-06 -7.54696399e-07  2.61353638e-08  2.43845215e-07
 1.25415476e-06  5.26433455e-07  1.60767757e-06 -7.91770162e-03
 1.17799013e-06  7.43299641e-06  1.60767757e-06  5.69716749e-07
 4.57842709e-05 -1.68986282e-07  2.25957198e-07  5.18769644e-07
 7.88227482e-07  8.33584857e-07  1.06521675e-06 -4.82127778e-05
 2.70762518e-05  3.85574775e-06 -1.83665254e-06  5.61949764e-07
-4.54882891e-04  4.14435025e-06 -1.74550755e-06  2.01931669e-07
-1.10811369e-06  5.92166223e-03 -4.93110833e-07  2.32084975e-07
 3.10649128e-06 -3.46365555e-06  4.77623850e-07  5.75592251e-05
 1.17737406e-05 -3.47460694e-06  4.57166043e-06 -6.74175825e-05
 2.66762839e-07  5.20134845e-07  5.14162157e-06 -7.83118623e-07
-1.20901904e-06  2.72712040e-06  2.70856736e-05  9.99950703e-01
 2.38384662e-07 -7.89714288e-07 -3.43163171e-07  7.17012708e-07
-1.77702849e-04 -1.11256991e-05 -2.44502741e-04  5.46613372e-06
-1.24285974e-05]

//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
```

```
[ 7.16010384e-22 -1.74360868e-08 1.69429309e-07 1.08274698e-06
 6.08399702e-06 7.22707073e-04 -6.57326606e-07 1.60767757e-06
 6.92153239e-06 -7.54696399e-07 2.61353638e-08 2.43845215e-07
 1.25415476e-06 5.26433455e-07 1.60767757e-06 -7.91770162e-03
 1.17799013e-06 7.43299641e-06 1.60767757e-06 5.69716749e-07
 4.57842709e-05 -1.68986282e-07 2.25957198e-07 5.18769644e-07
 7.88227482e-07 8.33584857e-07 1.06521675e-06 -4.82127778e-05
 2.70762518e-05 3.85574775e-06 -1.83665254e-06 5.61949764e-07
 -4.54882891e-04 4.14435025e-06 -1.74550755e-06 2.01931669e-07
 -1.10811369e-06 5.92166223e-03 -4.93110833e-07 2.32084975e-07
 3.10649128e-06 -3.46365555e-06 4.77623850e-07 5.75592251e-05
 1.17737406e-05 -3.47460694e-06 4.57166043e-06 -6.74175825e-05
 2.66762839e-07 5.20134845e-07 5.14162157e-06 -7.83118623e-07
 -1.20901904e-06 2.72712040e-06 2.70856736e-05 9.99950703e-01
 2.38384662e-07 -7.89714288e-07 -3.43163171e-07 7.17012708e-07
 -1.77702849e-04 -1.11256991e-05 -2.44502741e-04 5.46613372e-06
 -1.24285974e-05]
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
[ 4.81333019e-22 -1.74360868e-08 1.69429309e-07 1.08274698e-06
 6.08399702e-06 7.22707073e-04 -6.57326606e-07 1.60767757e-06
 6.92153239e-06 -7.54696399e-07 2.61353638e-08 2.43845215e-07
 1.25415476e-06 5.26433455e-07 1.60767757e-06 -7.91770162e-03
 1.17799013e-06 7.43299641e-06 1.60767757e-06 5.69716749e-07
 4.57842709e-05 -1.68986282e-07 2.25957198e-07 5.18769644e-07
 7.88227482e-07 8.33584857e-07 1.06521675e-06 -4.82127778e-05
 2.70762518e-05 3.85574775e-06 -1.83665254e-06 5.61949764e-07
 -4.54882891e-04 4.14435025e-06 -1.74550755e-06 2.01931669e-07
 -1.10811369e-06 5.92166223e-03 -4.93110833e-07 2.32084975e-07
 3.10649128e-06 -3.46365555e-06 4.77623850e-07 5.75592251e-05
 1.17737406e-05 -3.47460694e-06 4.57166043e-06 -6.74175825e-05
 2.66762839e-07 5.20134845e-07 5.14162157e-06 -7.83118623e-07
 -1.20901904e-06 2.72712040e-06 2.70856736e-05 9.99950703e-01
 2.38384662e-07 -7.89714288e-07 -3.43163171e-07 7.17012708e-07
 -1.77702849e-04 -1.11256991e-05 -2.44502741e-04 5.46613372e-06
 -1.24285974e-05]
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logisti
c.py:432: FutureWarning: Default solver will be changed to 'lbfgs' i
n 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
//anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:929: Con
vergenceWarning: Liblinear failed to converge, increase the number o
f iterations.
```

```
"the number of iterations.", ConvergenceWarning)
```

```
In [254]:
```

```
for i in range(0,6):
    print('PCA BER for Validation and Test for n = ', (i+1)*5 , pcaBERList[i])
```

```
PCA BER for Validation and Test for n = 5 (0.3394986449864499, 0.32
44505494505494)
```

```
PCA BER for Validation and Test for n = 10 (0.2279810298102981, 0.3
2138278388278385)
```

```
PCA BER for Validation and Test for n = 15 (0.21036585365853666, 0.
24821428571428572)
```

```
PCA BER for Validation and Test for n = 20 (0.2069783197831978, 0.1
9565018315018312)
```

```
PCA BER for Validation and Test for n = 25 (0.20630081300813008, 0.
17399267399267404)
```

```
PCA BER for Validation and Test for n = 30 (0.13130081300813012, 0.
16506410256410253)
```

```
In [ ]:
```