CSE 258, Fall 2019: Homework 2

Student ID: A53308934
Student Name: Deng Zhang

Tasks - Diagnostics

In [1]:

```
import gzip
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model
import sklearn
from random import shuffle
import random
```

In [2]:

```
f = open("./Data/hw2/5year.arff", 'r')
```

In [3]:

```
while not '@data' in f.readline():
    pass
```

In [4]:

```
dataset = []
for l in f:
    if '?' in l: # Missing entry
        continue
    l = l.split(',')
    values = [1] + [float(x) for x in l]
    values[-1] = values[-1] > 0 # Convert to bool
    dataset.append(values)
```

Question 1:

In [5]:

```
# use the last col as y, the reset as x
X = [values[:-1] for values in dataset]
y = [values[-1] for values in dataset]
```

In [6]:

```
model = linear_model.LogisticRegression(C=1.0)
```

In [7]:

```
model.fit(X, y)
```

```
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
```

Out[7]:

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=Tru
e,
                   intercept_scaling=1, l1_ratio=None, max_iter=100,
                   multi_class='warn', n_jobs=None, penalty='l2',
                   random_state=None, solver='warn', tol=0.0001, verbose=
0,
                   warm_start=False)
```

In [8]:

```
predictions = model.predict(X)

correct = predictions == y

print("Accuracy = " + str(sum(correct) / len(correct)))
```

```
Accuracy = 0.9663477400197954
```

In [9]:

```
TP = sum([(p and l) for (p,l) in zip(predictions, y)])
FP = sum([(p and not l) for (p,l) in zip(predictions, y)])
TN = sum([(not p and not l) for (p,l) in zip(predictions, y)])
FN = sum([(not p and l) for (p,l) in zip(predictions, y)])
```

In [10]:

```
TPR = TP / (TP + FN)
TNR = TN / (TN + FP)
```

In [11]:

```
BER = 1 - 1/2 * (TPR + TNR)
print("Balanced error rate = " + str(BER))
```

```
Balanced error rate = 0.48580623782459387
```

In [12]:

```
# Answer of Question 1:

# Accuracy = 0.9663477400197954
# Balanced error rate = 0.48580623782459387
```

Question 3:

In [13]:

```
random.shuffle(dataset)
```

In [14]:

```
X = [values[:-1] for values in dataset]
y = [values[-1] for values in dataset]
N = len(X)
X_train = X[:N//2]
X_valid = X[N//2:3*N//4]
X_test = X[3*N//4:]
y_train = y[:N//2]
y_valid = y[N//2:3*N//4]
y_test = y[3*N//4:]

len(X), len(X_train), len(X_test)
```

Out[14]:

```
(3031, 1515, 758)
```

In [15]:

```
model = linear_model.LogisticRegression(class_weight='balanced')
model.fit(X_train, y_train)
```

```
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
```

Out[15]:

```
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, penalty
='l2',
                   random_state=None, solver='warn', tol=0.0001, verbose=
0,
                   warm_start=False)
```

In [16]:

```
predictionsTrain = model.predict(X_train)
predictionsValid = model.predict(X_valid)
predictionsTest = model.predict(X_test)

correctPredictionsTrain = predictionsTrain == y_train
correctPredictionsValid = predictionsValid == y_valid
correctPredictionsTest = predictionsTest == y_test
```

In [17]:

```
print("Accuracy of Train = " + str(sum(correctPredictionsTrain) / len(correctPrediction
sTrain)))
print("Accuracy of Valid = " + str(sum(correctPredictionsValid) / len(correctPrediction
sValid)))
print("Accuracy of Test = " + str(sum(correctPredictionsTest) / len(correctPredictionsT
est)))
```

```
Accuracy of Train = 0.8085808580858086
Accuracy of Valid = 0.7823218997361477
Accuracy of Test = 0.7915567282321899
```

In [18]:

```
def countBer(predictions, Y):
    TP = sum([(p and l) for (p,l) in zip(predictions, Y)])
    FP = sum([(p and not l) for (p,l) in zip(predictions, Y)])
    TN = sum([(not p and not l) for (p,l) in zip(predictions, Y)])
    FN = sum([(not p and l) for (p,l) in zip(predictions, Y)])
    TPR = TP / (TP + FN)
    TNR = TN / (TN + FP)
    precision = TP / (TP + FP)
    recall = TP / (TP + FN)
    #F1 = 2 * (precision*recall) / (precision + recall)
    #print(F1)
    #F10 = 101 * (precision*recall) / (100 * precision + recall)
    #print(F10)
    return 1 - 1/2 * (TPR + TNR)
```

In [19]:

```
print("Balanced error rate of Train = " + str(countBer(predictionsTrain, y_train)))
print("Balanced error rate of Valid = " + str(countBer(predictionsValid, y_valid)))
print("Balanced error rate of Test = " + str(countBer(predictionsTest, y_test)))
```

```
Balanced error rate of Train = 0.22894211052105784
Balanced error rate of Valid = 0.31908023483365944
Balanced error rate of Test = 0.23962450592885376
```

In [20]:

```
# Answer of Question 3:

# Accuracy of Train = 0.7947194719471947
# Balanced error rate of Train = 0.22465886939571145
# Accuracy of Valid = 0.7968337730870713
# Balanced error rate of Valid = 0.16536103542234337
# Accuracy of Test = 0.7770448548812665
# Balanced error rate of Test = 0.27657168701944823
```

Question 4:

In [21]:

```
def getBerNAccu(c, X, y):
    model = linear_model.LogisticRegression(C=c, class_weight='balanced')
    model.fit(X, y)
    predictions = model.predict(X)
    correct = predictions == y
    accu = sum(correct) / len(correct)
    TP = sum([(p and l) for (p,l) in zip(predictions, y)])
    FP = sum([(p and not l) for (p,l) in zip(predictions, y)])
    TN = sum([(not p and not l) for (p,l) in zip(predictions, y)])
    FN = sum([(not p and l) for (p,l) in zip(predictions, y)])
    TPR = TP / (TP + FN)
    TNR = TN / (TN + FP)
    ber = 1 - 1/2 * (TPR + TNR)
    return ber, accu
```

In [22]:

```
C = [0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000]
berTrain = []
berValid = []
berTest = []
accuTrain = []
accuValid = []
accuTest = []
for c in C:
    ber_train, accu_train = getBerNAccu(c, X_train, y_train)
    berTrain.append(ber_train)
    accuTrain.append(accu_train)
    ber_valid, accu_valid = getBerNAccu(c, X_valid, y_valid)
    berValid.append(ber_valid)
    accuValid.append(accu_valid)
    ber_test, accu_test = getBerNAccu(c, X_test, y_test)
    berTest.append(ber_test)
    accuTest.append(accu_test)
```

```
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
```

```
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
```

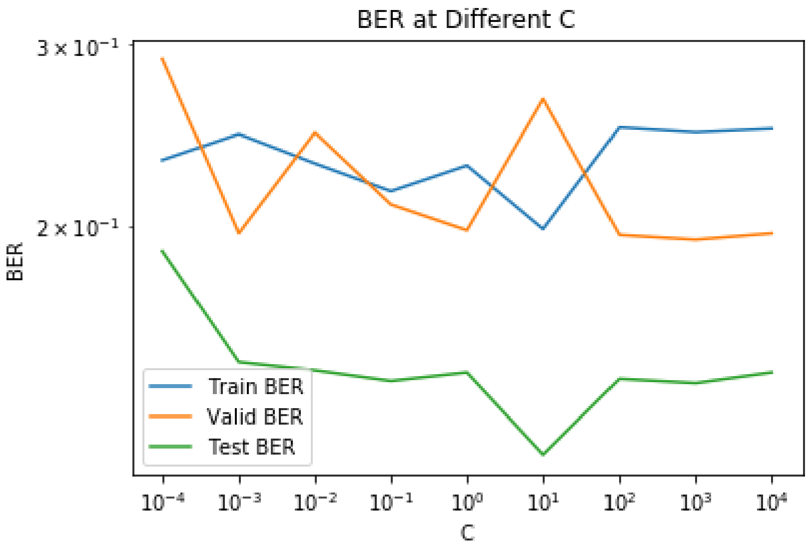
```
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
```

In [23]:

```
x = [10**i for i in range(-4, 5, 1)]
plt.loglog(x, berTrain, label = 'Train BER')
plt.loglog(x, berValid, label = 'Valid BER')
plt.loglog(x, berTest, label = 'Test BER')
plt.title("BER at Different C")
plt.xticks(x)
plt.xlabel("C")
plt.ylabel("BER")
plt.legend()
```

Out[23]:

```
<matplotlib.legend.Legend at 0x225297fc128>
```

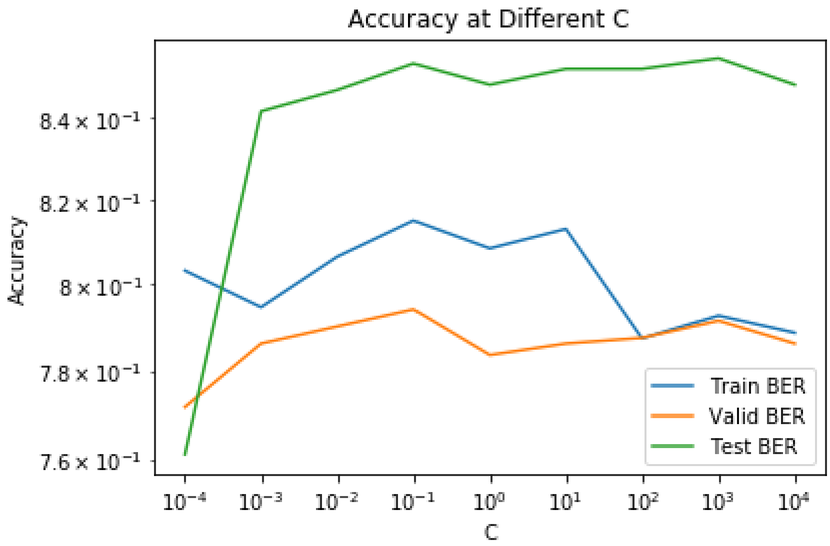


In [24]:

```
x = [10**i for i in range(-4, 5, 1)]
plt.loglog(x, accuTrain, label = 'Train BER')
plt.loglog(x, accuValid, label = 'Valid BER')
plt.loglog(x, accuTest, label = 'Test BER')
plt.title("Accuracy at Different C")
plt.xticks(x)
plt.xlabel("C")
plt.ylabel("Accuracy")
plt.legend()
```

Out[24]:

```
<matplotlib.legend.Legend at 0x22529c22f28>
```



In [25]:

```
# Answer to Question 4:
# BER at Different C is shown in the graph below
# I would choose 0.1 as my classifier, because the accuracy of 0.01
# is high and its BER is low comparably.
```

Question 6:

In [26]:

```
weights = [1.0] * len(y_train)
mod = linear_model.LogisticRegression(C=1, solver='lbfgs')
mod.fit(X_train, y_train, sample_weight=weights)
```

```
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:947: Conver
genceWarning: lbfgs failed to converge. Increase the number of iterations.
  "of iterations.", ConvergenceWarning)
```

Out[26]:

```
LogisticRegression(C=1, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, l1_ratio=None, max_iter=100,
                   multi_class='warn', n_jobs=None, penalty='l2',
                   random_state=None, solver='lbfgs', tol=0.0001, verbose=
0,
                   warm_start=False)
```

In [27]:

```
def countTF(predictions, Y):
    TP = sum([(p and l) for (p,l) in zip(predictions, Y)])
    FP = sum([(p and not l) for (p,l) in zip(predictions, Y)])
    TN = sum([(not p and not l) for (p,l) in zip(predictions, Y)])
    FN = sum([(not p and l) for (p,l) in zip(predictions, Y)])
    return TP, FP, TN, FN
```

In [28]:

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

In [29]:

```
TP, FP, TN, FN = countTF(predictionsTest, y_test)
```

In [30]:

```
precision = TP / (TP + FP)
recall = TP / (TP + FN)
precision, recall
```

Out[30]:

```
(0.09523809523809523, 0.7272727272727273)
```

In [31]:

```
F1 = 2 * (precision*recall) / (precision + recall)
print("Unweighted F1 = " + str(F1))
F10 = 101 * (precision*recall) / (100 * (precision + recall))
print("Unweighted F10 = " + str(F10))
```

```
Unweighted F1 = 0.16842105263157894
Unweighted F10 = 0.08505263157894737
```

In [32]:

```
weightPos = 1 - sum(d == True for d in y_train) / len(y_train)
weightNeg = 1 - weightPos

weights = [weightPos if i == True else weightNeg for i in y_train]
model = linear_model.LogisticRegression(C = 1, solver='lbfgs')
model.fit(X_train, y_train, sample_weight=weights);

predictionsTest = model.predict(X_test)
TP, FP, TN, FN = countTF(predictionsTest, y_test)

precision = TP / (TP + FP)
recall = TP / (TP + FN)

F1 = 2 * (precision*recall) / (precision + recall)
print("Weighted F1 = " + str(F1))
F10 = 101 * (precision*recall) / (100 * (precision + recall))
print("Weighted F10 = " + str(F10))
```

```
Weighted F1 = 0.14545454545454548
Weighted F10 = 0.07345454545454547

E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:947: Conver
genceWarning: lbfgs failed to converge. Increase the number of iterations.
  "of iterations.", ConvergenceWarning)
```

In [34]:

```
# Answer to Question 6:
# I choose the precentage of positive data at train_y dataset as the weight
# of positive value, and the precentage of negative value as the weight of
# negative value.
# The weight of positive is far more than negative, so I made it negative
# correlated with the result.
```

Tasks - Diagnostics

Question 7

In [35]:

```
from sklearn.decomposition import PCA
```

In [36]:

```
pca = PCA()
pca.fit(X_train)
print(pca.components_[0])
```

```
[ 3.92706411e-19 -1.85513201e-08 -9.24773520e-08 -1.17884013e-06
 -5.75127505e-06 -2.10289748e-03  8.03267837e-07 -1.53861883e-06
 -6.84769037e-06  5.26691130e-07 -8.82788171e-08 -2.88064149e-07
 -1.56892789e-06 -4.80255307e-07 -1.53861883e-06  2.72362045e-04
 -1.17237867e-06 -7.50212374e-06 -1.53861883e-06 -5.25686112e-07
 -5.84489530e-05  1.23039129e-05 -2.62666118e-07 -4.70013816e-07
 -5.26441317e-07 -7.47493553e-07 -1.02999438e-06  2.93237493e-05
 -2.80815572e-06 -3.05436166e-06  1.57240912e-06 -5.08126451e-07
  4.99328670e-04 -4.45263278e-06  6.01538345e-07 -2.32222189e-07
  1.37423155e-06  4.19977927e-03  3.20403647e-07 -2.14149773e-07
 -3.07369258e-06 -1.58221957e-06 -4.47914478e-07 -6.24210902e-05
 -3.97007587e-06  1.16137198e-06 -4.32045863e-06  2.88278856e-04
 -3.24611795e-07 -4.92469031e-07 -5.24521391e-06  8.67990222e-07
  1.33427150e-06 -4.25445050e-06 -2.81869185e-06 -9.99988728e-01
 -2.51644946e-07 -2.01930640e-07  2.99231007e-07  1.59404728e-07
  1.17950525e-04  1.11491313e-05  2.32758960e-04 -6.07822840e-06
  7.85583001e-06]
```

Question 8

In [37]:

```
def countComponent(X, y):
    berList = []
    for component in range(5, 31, 5):
        pca = PCA(n_components=component)
        pca.fit(X)
        Xpca = np.matmul(X, pca.components_.T)
        model = linear_model.LogisticRegression(C=1.0, class_weight='balanced')
        model.fit(Xpca, y)
        predictions = model.predict(Xpca)
        berList.append(countBer(predictions, y))
    return berList
```

In [39]:

```
ber_train = countComponent(X_train, y_train)
ber_valid = countComponent(X_valid, y_valid)
ber_test = countComponent(X_test, y_test)

plt.plot(range(5, 31, 5), ber_train, label='Train BER')
plt.plot(range(5, 31, 5), ber_valid, label='Valid BER')
plt.plot(range(5, 31, 5), ber_test, label='Test BER')
plt.title("BER of Collections")
plt.xlabel("Component")
plt.ylabel("BER")
plt.show()
```

```
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
```

```
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
E:\anaconda\lib\site-packages\sklearn\linear_model\logistic.py:432: Future
Warning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solv
er to silence this warning.
  FutureWarning)
E:\anaconda\lib\site-packages\sklearn\svm\base.py:929: ConvergenceWarning:
Liblinear failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
```