Implement SGD Classifier with Logloss and L2 regularization Using SGD without using sklearn

There will be some functions that start with the word "grader" ex: grader_weights(), grader_sigmoid(), grader_logloss() etc, you should not change those function definition.

Every Grader function has to return True.

Importing packages

In [1]:
```
import numpy as np
import pandas as pd
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn import linear_model
```

Creating custom dataset

In [2]:
```
# please don't change random_state
X, y = make_classification(n_samples=50000, n_features=15, n_informative=10, n_redundant=5,
                           n_classes=2, weights=[0.7], class_sep=0.7, random_state=15)
# make_classification is used to create custom dataset
# Please check this link (https://scikit-learn.org/stable/modules/generated/sklearn.dataset
s.make_classification.html) for more details
```

In [3]:
```
X.shape, y.shape
```

Out[3]:
```
((50000, 15), (50000,))
```

Splitting data into train and test

In [4]:
```
#please don't change random state
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=15)
```

In [5]:
```
# Standardizing the data.
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

In [6]:
```
X_train.shape, y_train.shape, X_test.shape, y_test.shape
```

Out[6]:
```
((37500, 15), (37500,), (12500, 15), (12500,))
```

SGD classifier

In [7]:
```
# alpha : float
# Constant that multiplies the regularization term.

# eta0 : double
# The initial learning rate for the ‘constant’, ‘invscaling’ or ‘adaptive’ schedules.

clf = linear_model.SGDClassifier(eta0=0.0001, alpha=0.0001, loss='log', random_state=15, pen
alty='l2', tol=1e-3, verbose=2, learning_rate='constant')
clf
# Please check this documentation (https://scikit-learn.org/stable/modules/generated/sklear
n.linear_model.SGDClassifier.html)
```

Out[7]:
```
SGDClassifier(alpha=0.0001, average=False, class_weight=None, epsilon=0.1,
       eta0=0.0001, fit_intercept=True, l1_ratio=0.15,
       learning_rate='constant', loss='log', max_iter=None, n_iter=None,
       n_jobs=1, penalty='l2', power_t=0.5, random_state=15, shuffle=True,
       tol=0.001, verbose=2, warm_start=False)
```

In [8]:
```
clf.fit(X=X_train, y=y_train) # fitting our model
```

```
-- Epoch 1
Norm: 0.70, NNZs: 15, Bias: -0.499391, T: 37500, Avg. loss: 0.552631
Total training time: 0.02 seconds.
-- Epoch 2
Norm: 1.04, NNZs: 15, Bias: -0.750277, T: 75000, Avg. loss: 0.448128
Total training time: 0.04 seconds.
-- Epoch 3
Norm: 1.26, NNZs: 15, Bias: -0.902777, T: 112500, Avg. loss: 0.415726
Total training time: 0.04 seconds.
-- Epoch 4
Norm: 1.42, NNZs: 15, Bias: -1.003874, T: 150000, Avg. loss: 0.400898
Total training time: 0.05 seconds.
-- Epoch 5
Norm: 1.55, NNZs: 15, Bias: -1.075094, T: 187500, Avg. loss: 0.392871
Total training time: 0.06 seconds.
-- Epoch 6
Norm: 1.65, NNZs: 15, Bias: -1.128728, T: 225000, Avg. loss: 0.388085
Total training time: 0.06 seconds.
-- Epoch 7
Norm: 1.73, NNZs: 15, Bias: -1.169943, T: 262500, Avg. loss: 0.385063
Total training time: 0.07 seconds.
-- Epoch 8
Norm: 1.80, NNZs: 15, Bias: -1.203552, T: 300000, Avg. loss: 0.383058
Total training time: 0.08 seconds.
-- Epoch 9
Norm: 1.86, NNZs: 15, Bias: -1.230411, T: 337500, Avg. loss: 0.381694
Total training time: 0.08 seconds.
-- Epoch 10
Norm: 1.91, NNZs: 15, Bias: -1.249466, T: 375000, Avg. loss: 0.380756
Total training time: 0.09 seconds.
Convergence after 10 epochs took 0.09 seconds
```

Out[8]:
```
SGDClassifier(alpha=0.0001, average=False, class_weight=None, epsilon=0.1,
       eta0=0.0001, fit_intercept=True, l1_ratio=0.15,
       learning_rate='constant', loss='log', max_iter=None, n_iter=None,
       n_jobs=1, penalty='l2', power_t=0.5, random_state=15, shuffle=True,
       tol=0.001, verbose=2, warm_start=False)
```

In [9]:
```
clf.coef_, clf.coef_.shape, clf.intercept_
#clf.coef_ will return the weights
#clf.coef_.shape will return the shape of weights
#clf.intercept_ will return the intercept term
```

Out[9]:
```
(array([[-0.83187476,  0.58935497, -0.05233948,  0.59159475, -0.33771644,
          0.87826212, -0.85798961, -0.06890856,  0.37968271,  0.3720168 ,
          0.22881296,  0.04398642, -0.08060734,  0.51274272,  0.07080401]]),
 (1, 15),
 array([-1.24946622]))
```

```
# This is formatted as code
```

Implement Logistic Regression with L2 regularization Using SGD: without using sklearn

1. We will be giving you some functions, please write code in that functions only.
2. After every function, we will be giving you expected output, please make sure that you get that output.

- Initialize the weight_vector and intercept term to zeros (Write your code in def initialize_weights())
- Create a loss function (Write your code in def logloss())

  $logloss = -1 * \frac{1}{n}\Sigma_{foreachYt,Y_{pred}}(Ytlog10(Y_{pred})+(1-Yt)log10(1-Y_{pred}))$
- for each epoch:
  - for each batch of data points in train: (keep batch size=1)
    - calculate the gradient of loss function w.r.t each weight in weight vector (write your code in def gradient_dw())

      $dw^{(t)} = x_n(y_n - \sigma((w^{(t)})^{T} x_n+b^{t}))- \frac{λ}{N}w^{(t)})$
    - Calculate the gradient of the intercept (write your code in def gradient_db()) check this

      $db^{(t)} = y_n- \sigma((w^{(t)})^{T} x_n+b^{t}))$
    - Update weights and intercept (check the equation number 32 in the above mentioned pdf):
      $w^{(t+1)}← w^{(t)}+α(dw^{(t)})$

      $b^{(t+1)}←b^{(t)}+α(db^{(t)})$
  - calculate the log loss for train and test with the updated weights (you can check the python assignment 10th question)
  - And if you wish, you can compare the previous loss and the current loss, if it is not updating, then you can stop the training
  - append this loss in the list (this will be used to see how loss is changing for each epoch after the training is over)

Initialize weights

In [10]:
```
def initialize_weights(dim):
    ''' In this function, we will initialize our weights and bias'''
    #initialize the weights to zeros array of (1,dim) dimensions
    #you use zeros_like function to initialize zero, check this link https://docs.scipy.org/
doc/numpy/reference/generated/numpy.zeros_like.html
    #initialize bias to zero
    w = np.zeros_like(dim)
    b = 0
    return w,b
```

In [11]:
```
dim=X_train[0]
w,b = initialize_weights(dim)
print('w =',(w))
print('b =',str(b))
```

```
w = [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
b = 0
```

Grader function - 1

In [12]:
```
dim=X_train[0]
w,b = initialize_weights(dim)
def grader_weights(w,b):
  assert((len(w)==len(dim)) and b==0 and np.sum(w)==0.0)
  return True
grader_weights(w,b)
```

Out[12]:
```
True
```

Compute sigmoid

$sigmoid(z)= 1/(1+exp(-z))$

In [13]:
```
import math
def sigmoid(z):
    ''' In this function, we will return sigmoid of z'''
    # compute sigmoid(z) and return
    sigmoid=1/(1+math.exp(-z))
    return sigmoid
```

Grader function - 2

In [14]:
```
def grader_sigmoid(z):
  val=sigmoid(z)
  assert(val==0.8807970779778823)
  return True
grader_sigmoid(2)
```

Out[14]:
```
True
```

Compute loss

$logloss = -1 * \frac{1}{n}\Sigma_{foreachYt,Y_{pred}}(Ytlog10(Y_{pred})+(1-Yt)log10(1-Y_{pred}))$

In [15]:
```
def logloss(y_true,y_pred):
    '''In this function, we will compute log loss '''
    n=len(y_true)
    s=0
    for i in range(n):
        s += (y_true[i] * np.log10(y_pred[i]))+((1-y_true[i]) * np.log10(1-y_pred[i]))
    loss= (-1 * s)/n
    return loss
```

Grader function - 3

In [16]:
```
def grader_logloss(true,pred):
  loss=logloss(true,pred)
  assert(loss==0.07644900402910389)
  return True
true=[1,1,0,1,0]
pred=[0.9,0.8,0.1,0.8,0.2]
grader_logloss(true,pred)
```

Out[16]:
```
True
```

Compute gradient w.r.to 'w'

$dw^{(t)} = x_n(y_n - \sigma((w^{(t)})^{T} x_n+b^{t}))- \frac{λ}{N}w^{(t)}$

In [17]:
```
def gradient_dw(x,y,w,b,alpha,N):
    '''In this function, we will compute the gardient w.r.to w '''

    dw = x*(y-sigmoid(np.dot(w.T,x)+b)) - ((alpha/N)*w)
    return dw
```

Grader function - 4

In [18]:
```
def grader_dw(x,y,w,b,alpha,N):
  grad_dw=gradient_dw(x,y,w,b,alpha,N)
  assert(np.sum(grad_dw)==2.613689585)
  return True
grad_x=np.array([-2.07864835,  3.31604252, -0.79104357, -3.87045546, -1.14783286,
       -2.81434437, -0.86771071, -0.04073287,  0.84827878,  1.99451725,
        3.67152472,  0.01451875,  2.01062888,  0.07373904, -5.54586092])
grad_y=0
grad_w,grad_b=initialize_weights(grad_x)
alpha=0.0001
N=len(X_train)
grader_dw(grad_x,grad_y,grad_w,grad_b,alpha,N)
```

Out[18]:
```
True
```

Compute gradient w.r.to 'b'

$db^{(t)} = y_n- \sigma((w^{(t)})^{T} x_n+b^{t})$

In [19]:
```
def gradient_db(x,y,w,b):
    '''In this function, we will compute gradient w.r.to b '''
    db = y - sigmoid(np.dot(w.T,x) + b)
    return db
```

Grader function - 5

In [20]:
```
def grader_db(x,y,w,b):
  grad_db=gradient_db(x,y,w,b)
  assert(grad_db==-0.5)
  return True
grad_x=np.array([-2.07864835,  3.31604252, -0.79104357, -3.87045546, -1.14783286,
       -2.81434437, -0.86771071, -0.04073287,  0.84827878,  1.99451725,
        3.67152472,  0.01451875,  2.01062888,  0.07373904, -5.54586092])
grad_y=0
grad_w,grad_b=initialize_weights(grad_x)
alpha=0.0001
N=len(X_train)
grader_db(grad_x,grad_y,grad_w,grad_b)
```

Out[20]:
```
True
```

Implementing logistic regression

In [21]:
```
def train(X_train,y_train,X_test,y_test,epochs,alpha,eta0):
    ''' In this function, we will implement logistic regression'''
    #Here eta0 is learning rate
    #implement the code as follows
    # initalize the weights (call the initialize_weights(X_train[0]) function)
    # for every epoch
        # for every data point(X_train,y_train)
           #compute gradient w.r.to w (call the gradient_dw() function)
           #compute gradient w.r.to b (call the gradient_db() function)
           #update w, b
        # predict the output of x_train[for all data points in X_train] using w,b
        #compute the loss between predicted and actual values (call the loss function)
        # store all the train loss values in a list
        # predict the output of x_test[for all data points in X_test] using w,b
        #compute the loss between predicted and actual values (call the loss function)
        # store all the test loss values in a list
        # you can also compare previous loss and current loss, if loss is not updating then
 stop the process and return w,b
    w,b = initialize_weights(dim)
    train_loss=[]
    test_loss=[]

    for i in range(epochs):
        pred_train=[]
        pred_test=[]
        for j in range(N):
            dw= gradient_dw(X_train[j],y_train[j],w,b,alpha,N)
            db= gradient_db(X_train[j],y_train[j],w,b)
            w= w + (eta0 * dw)
            b= b + (eta0 * db)
        for k in range(N):
            pred_train.append(sigmoid(np.dot(w,X_train[k])+b))
        LL=logloss(y_train,pred_train)
        train_loss.append(LL)
        for m in range(len(X_test)):
            pred_test.append(sigmoid(np.dot(w,X_test[m])+b))
        LL1=logloss(y_test,pred_test)
        test_loss.append(LL1)
    return w,b,train_loss,test_loss
```

In [22]:
```
alpha=0.0001
eta0=0.0001
N=len(X_train)
epochs=10
w,b,train_log_loss,test_log_loss=train(X_train,y_train,X_test,y_test,epochs,alpha,eta0)
```

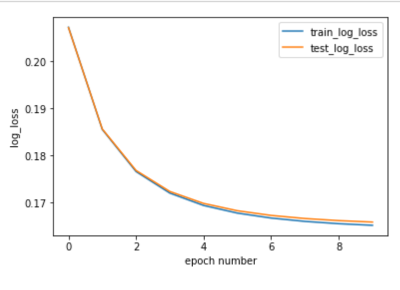
Goal of assignment

Compare your implementation and SGDClassifier's the weights and intercept, make sure they are as close as possible i.e difference should be in terms of 10^-3

Plot epoch number vs train , test loss

- epoch number on X-axis
- loss on Y-axis

In [24]:
```
from matplotlib import pyplot as plt
plt.plot(train_log_loss,label='train_log_loss')
plt.plot(test_log_loss,label='test_log_loss')
plt.xlabel('epoch number')
plt.ylabel('log_loss')
plt.legend()
plt.show()
```



In [25]:
```
def pred(w,b, X):
    N = len(X)
    predict = []
    for i in range(N):
        z=np.dot(w,X[i])+b
        if sigmoid(z) >= 0.5: # sigmoid(w,x,b) returns 1/(1+exp(-(dot(x,w)+b)))
            predict.append(1)
        else:
            predict.append(0)
    return np.array(predict)
print(1-np.sum(y_train - pred(w,b,X_train))/len(X_train))
print(1-np.sum(y_test  - pred(w,b,X_test))/len(X_test))
```

```
0.94944
0.94648
```

In [26]:
```
for v in range(epochs):
    train_loss=train_log_loss[v]
    test_loss=test_log_loss[v]
    print("Epochs",v)
    print("train_loss",train_log_loss[v],'\t','test_loss',test_log_loss[v],'\n')
```

```
Epochs 0
train_loss 0.20729781784140838   test_loss 0.2072221978118188

Epochs 1
train_loss 0.18556210141426166   test_loss 0.18565259434678275

Epochs 2
train_loss 0.17659652085620509   test_loss 0.17682567720849304

Epochs 3
train_loss 0.17201289496451905   test_loss 0.17235324848189568

Epochs 4
train_loss 0.16938000886115878   test_loss 0.16981009840800462

Epochs 5
train_loss 0.16775336575455      test_loss 0.16825663498220056

Epochs 6
train_loss 0.16669776297615663   test_loss 0.1672612889069227

Epochs 7
train_loss 0.16598837500432867   test_loss 0.16660192986644845

Epochs 8
train_loss 0.1654991822760498    test_loss 0.1661545712175774

Epochs 9
train_loss 0.16515513945496194   test_loss 0.16584572669386236
```

In [27]:
```
print('weights:',w,'\n')
print('intercept:',b)
```

```
weights: [-0.83543917  0.59554963 -0.05042441  0.59139957 -0.33675676  0.87822749
 -0.85715792 -0.068385    0.38429384  0.36811606  0.22870443  0.0472269
 -0.07976159  0.5131997   0.07063889]

intercept: -1.247904496218835
```

In [33]:
```
#Results we get after comparing our implementation and SGD classifier
print('weight difference:',w-clf.coef_)
print('Intercept difference:',b-clf.intercept_)
```

```
weight difference: [[-3.56441269e-03  6.19465370e-03  1.91506968e-03 -1.95178567e-04
   9.59670368e-04 -3.46279712e-05  8.31685290e-04  5.23557174e-04
   4.61113273e-03 -3.90074267e-03 -1.08531791e-04  3.24047785e-03
   8.45749627e-04  4.56977319e-04 -1.65124455e-04]]
Intercept difference: [0.00156173]
```

After implementing logistic regression with SGD and comparing with sklearn's SGD classifier, the weights and the intercept are similar to each other that is within a difference of 10^-3