```python
import pandas as pd
import numpy as np
import math
import matplotlib.pyplot as plt
import matplotlib.colors as mcolors

df = pd.read_excel (r'C:\***\usable_data_final_train.xlsx')
tf = pd.read_excel (r'C:\***\usable_data_final_test.xlsx')
Data = df.values
Test = tf.values
print(Data.shape,Test.shape)
```

>> (15154, 43) (956, 43)

```python
def remove_outliers(arr, k):
    mu, sigma = np.mean(arr, axis=0), np.std(arr, axis=0, ddof=1)
    return arr[np.all(np.abs((arr - mu) / sigma) < k, axis=1)]

def remove_outliers_bis(arr, k):
    mask = np.ones((arr.shape[0],), dtype=np.bool)
    mu, sigma = np.mean(arr, axis=0), np.std(arr, axis=0, ddof=1)
    for j in range(arr.shape[1]-1):
        col = arr[:, j]
        mask[mask] &= np.abs((col[mask] - mu[j]) / sigma[j]) < k
    return arr[mask]

Clean_Data = remove_outliers(Data, 2)
Clean_Test = remove_outliers(Test, 2)
print(Clean_Data.shape,Clean_Test.shape)
```

>> (9007, 43) (524, 43)

```python
epsilon = 1e-12

neurons = np.array([10,5,1])
layers = neurons.size
print("number of layers: " + str(layers))
print("number of hidden layers: " + str(layers-1))
print(neurons)

input_features = len(Clean_Data[0])
examples = len(Clean_Data)
Y = Clean_Data[:,input_features-1]
Y = Y.reshape((examples,1))
X = np.delete(Clean_Data, input_features-1, axis=1)
X = (X - X.min(axis=0)) / (X.max(axis=0)- X.min(axis=0))
X = X.T


test_features = len(Clean_Test[0])
test_examples = len(Clean_Test)
Ytest = Clean_Test[:,test_features-1]
Ytest = Ytest.reshape((test_examples,1))
Xtest = np.delete(Clean_Test, test_features-1, axis=1)
Xtest = (Xtest - Xtest.min(axis=0)) / (Xtest.max(axis=0)- Xtest.min(axis=0))
Xtest = Xtest.T

print("number of input features + example column: ",input_features," and number of test features + example co
lumn: ",test_features)
print("number of usable examples: " ,examples," and number of usable test examples: ",test_examples)

print("shape of X and Xtest: ",X.shape,Xtest.shape)
print("shape of Y and Ytest: ",Y.shape,Ytest.shape)
```

>>
number of layers: 3
number of hidden layers: 2
[10 5 1]

```
number of input features + example column:  43  and number of test features + example column:  43
number of usable examples:  9007  and number of usable test examples:  524
shape of X and Xtest:  (42, 9007) (42, 524)
shape of Y and Ytest:  (9007, 1) (524, 1)
```

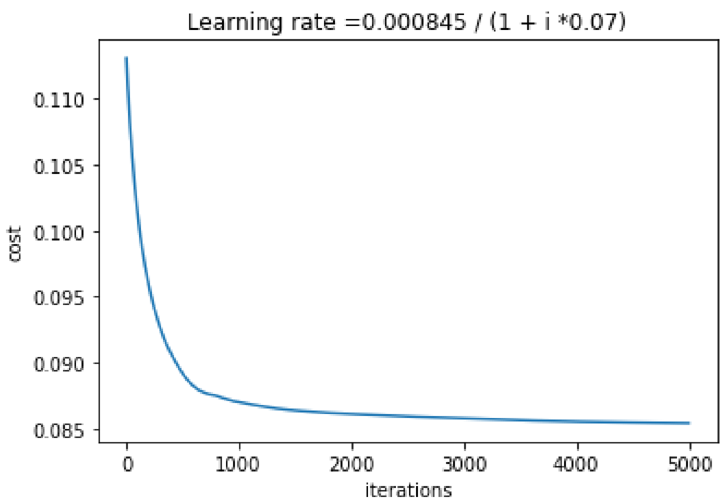
```
def sigmoid(Z):

    Z = np.round(Z,20)
    A = (1 / (1 + np.exp(-Z) + epsilon))
    activation_cache = Z

    return A, activation_cache

def leaky_relu(Z):

    Z = np.where(Z == 0, epsilon, Z)
    A = np.maximum(0.01*Z,Z)
    activation_cache = Z

    return A, activation_cache

def swish(Z):

    Z = np.where(Z ==np.nan, 0, Z)
    Z = np.round(Z,20)
    A = (Z / (1 + np.exp(-Z) + epsilon))
    activation_cache = Z

    return A, activation_cache

def sigmoid_backward(dA, activation_cache):

    A, Z = sigmoid(activation_cache)

    Z = np.where(Z ==np.nan, 0, Z)
    Z = np.round(Z,20)
    dZ = dA / (A - np.square(A)+epsilon)
    #dZ = dA*((2*epsilon)/(sigmoid(Z+epsilon)-sigmoid(Z-epsilon)))
    dZ = np.round(dZ,20)

    return dZ

def leaky_relu_backward(dA, activation_cache):

    dZ = dA*((2*epsilon)/(leaky_relu(Z+epsilon)-leaky_relu(Z-epsilon)))

    return dZ

def swish_backward(dA, activation_cache):

    A, Z = swish(activation_cache)

    Z = np.where(Z ==np.nan, 0, Z)
    Z = np.round(Z,20)
    dZ = (Z*dA) / ((A*((A*np.exp(-Z))+1))+epsilon)
    #dZ = dA*((2*epsilon)/(swish(Z+epsilon)-swish(Z-epsilon)))
    dZ = np.round(dZ,20)

    return dZ

n_x = input_features - 1
n_y = examples
n_x_test = test_features - 1
n_y_test = test_examples
n_h = neurons
L = layers

def initialize_parameters(n_x,n_h,L,n_y):
```

```python
    parameters = {}

    for i in range(0,L):
        if i == 0:
            parameters['W' + str(1)] = np.random.randn(n_h[0],n_x)*(1/n_x)
            parameters['b' + str(1)] = np.zeros((n_h[0],1))
            assert (parameters['W' + str(1)].shape == (n_h[0],n_x))
            assert (parameters['b' + str(1)].shape == (n_h[0],1))
        else:
            parameters['W' + str(i+1)] = np.random.randn(n_h[i],n_h[i-1])*(1/n_h[i-1])
            parameters['b' + str(i+1)] = np.zeros((n_h[i],1))
            assert (parameters['W' + str(i+1)].shape == (n_h[i],n_h[i-1]))
            assert (parameters['b' + str(i+1)].shape == (n_h[i],1))

        #print("initialized layer ",i+1)

    return parameters

def linear_forward(A, W, b):

    Z = np.dot(W,A)+b
    assert(Z.shape == (W.shape[0], A.shape[1]))
    cache = (A, W, b)

    return Z, cache

def linear_activation_forward(A_prev, W, b, activation):

    Z, linear_cache = linear_forward(A_prev, W, b)

    if activation == "sigmoid":
        A, activation_cache = sigmoid(Z)

    elif activation == "relu":
        A, activation_cache = leaky_relu(Z)

    elif activation == "swish":
        A, activation_cache = swish(Z)

    else:
        A = Z
        activation_cache = Z

    assert (A.shape == (W.shape[0], A_prev.shape[1]))

    cache = (linear_cache, activation_cache)
    return A, cache

def forward_prop(X, parameters):

    caches = []

    A = X

    L = int(len(parameters)/2)

    for l in range(0, L):
        A_prev = A
        A, cache = linear_activation_forward(A_prev, parameters['W' + str(l+1)],parameters['b' + str(l+1)], '
swish')
        caches.append(cache)

    AL, cache = linear_activation_forward(A_prev, parameters['W' + str(L)],parameters['b' + str(L)], 'sigmoid
')

    assert(AL.shape == (1,X.shape[1]))

    return(AL,caches)
```

```python
def compute_cost(AL, Y):

    AL = np.round(AL,20)
    m = n_y

    cost = (-1/m)*(np.dot(Y.T,np.log(AL.T))+np.dot((1-Y.T),(np.log(1-AL.T))))

    cost = np.squeeze(cost)

    assert(cost.shape == ())

    return cost

def linear_backward(dZ, cache):
    A_prev, W, b = cache
    m = A_prev.shape[1]

    dW = (1/m)*(np.dot(dZ,A_prev.T))
    db = (1/m)*np.sum(dZ, axis = 1, keepdims = True)
    dA_prev = np.dot(W.T,dZ)

    assert (dA_prev.shape == A_prev.shape)
    assert (dW.shape == W.shape)
    assert (db.shape == b.shape)

    return dA_prev, dW, db

def linear_activation_backward(dA, cache, activation):
    linear_cache, activation_cache = cache

    if activation == "relu":
        dZ = relu_backward(dA, activation_cache)

    elif activation == "sigmoid":
        dZ = sigmoid_backward(dA, activation_cache)

    elif activation == "swish":
        dZ = swish_backward(dA, activation_cache)

    dA_prev, dW, db = linear_backward(dZ, linear_cache)

    return dA_prev, dW, db

def back_prop(AL, Y, caches):

    grads = {}
    L = len(caches)
    m = AL.shape[1]
    Y = Y.reshape(AL.shape)

    dAL = - (np.divide(Y,AL)) - np.divide(1-Y, 1-AL)

    current_cache = caches[L-1]
    grads["dA"+str(L)],grads["dW"+str(L)],grads["db"+str(L)] = linear_activation_backward(dAL, current_cache,
 "sigmoid")

    for l in reversed(range(L-1)):
        current_cache = caches[l]
        dA_prev_temp, dW_temp, db_temp = linear_activation_backward(grads["dA"+str(l+2)], current_cache, "swi
sh")
        grads["dA"+str(l+1)] = dA_prev_temp
        grads["dW"+str(l+1)] = dW_temp
        grads["db"+str(l+1)] = db_temp

    return grads

def initialize_adam(parameters):
```

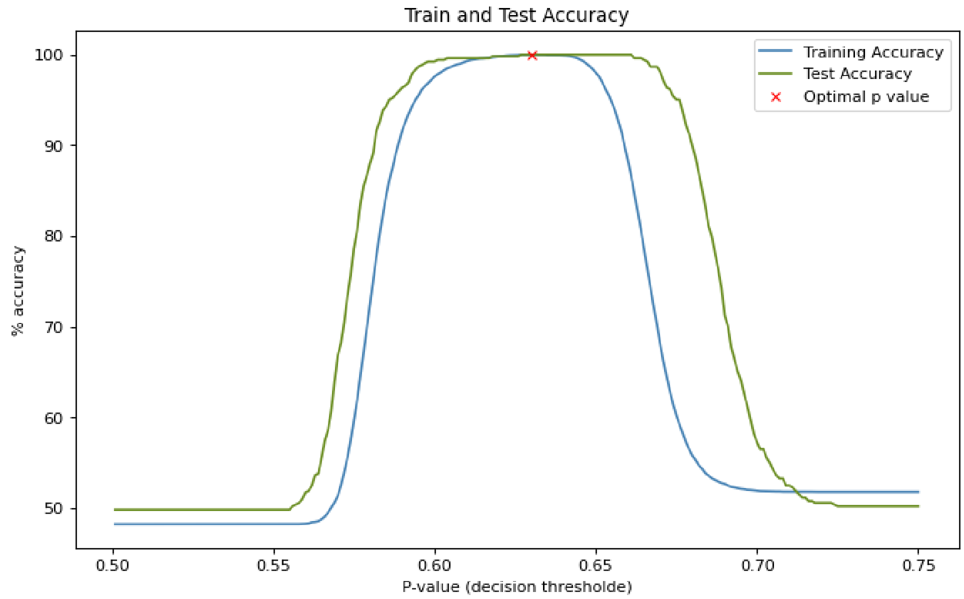
```
    L = int(len(parameters)/2)
    v = {}
    s = {}

    for l in range(L):
        v["dW" + str(l+1)] = np.zeros((parameters["W" + str(l+1)].shape))
        v["db" + str(l+1)] = np.zeros((parameters["b" + str(l+1)].shape))
        s["dW" + str(l+1)] = np.zeros((parameters["W" + str(l+1)].shape))
        s["db" + str(l+1)] = np.zeros((parameters["b" + str(l+1)].shape))

    return v, s

def update_parameters(parameters, grads, learning_rate):

    L = int(len(parameters)/2)

    for l in range(0, L):
        parameters["W" + str(l+1)] = parameters["W" + str(l+1)] - learning_rate*grads["dW" + str(l+1)]
        parameters["b" + str(l+1)] = parameters["b" + str(l+1)] - learning_rate*grads["db" + str(l+1)]

    return parameters

def update_parameters_adam(parameters, grads, v, s, t, learning_rate):

    beta1 = 0.9
    beta2 = 0.999
    L = int(len(parameters)/2)
    v_corrected = {}
    s_corrected = {}

    for l in range(0, L):
        v["dW" + str(l+1)] = beta1 * v["dW" + str(l+1)] + (1-beta1) * grads['dW' + str(l+1)]
        v["db" + str(l+1)] = beta1 * v["db" + str(l+1)] + (1-beta1) * grads['db' + str(l+1)]
        v_corrected["dW" + str(l+1)] = v["dW" + str(l+1)] / (1 - np.power(beta1,t))
        v_corrected["db" + str(l+1)] = v["db" + str(l+1)] / (1 - np.power(beta1,t))

        s["dW" + str(l+1)] = beta2 * s["dW" + str(l+1)] + (1-beta2) * np.power(grads['dW' + str(l+1)],2)
        s["db" + str(l+1)] = beta2 * s["db" + str(l+1)] + (1-beta2) * np.power(grads['db' + str(l+1)],2)
        s_corrected["dW" + str(l+1)] = s["dW" + str(l+1)] / (1 - np.power(beta2,t))
        s_corrected["db" + str(l+1)] = s["db" + str(l+1)] / (1 - np.power(beta2,t))

        parameters["W" + str(l+1)] = parameters["W" + str(l+1)] - learning_rate * (v_corrected["dW" + str(l+
1)] / (np.sqrt(s_corrected["dW" + str(l+1)]) + epsilon))
        parameters["b" + str(l+1)] = parameters["b" + str(l+1)] - learning_rate * (v_corrected["db" + str(l+
1)] / (np.sqrt(s_corrected["db" + str(l+1)]) + epsilon))

    return parameters, v, s

def L_layer_model(X, Y, n_x,n_h,L,n_y, learning_rate = 0.0075, num_iterations = 3000, decay = 0, print_cost=F
alse,dky = False):

    costs = []
    parameters = initialize_parameters(n_x,n_h,L,n_y)
    v,s = initialize_adam(parameters)
    t = 0

    for i in range(0, num_iterations):

        t = t+1
        AL, caches = forward_prop(X, parameters)
        cost = compute_cost(AL, Y)-0.29
        grads = back_prop(AL, Y, caches)

        parameters, v, s = update_parameters_adam(parameters, grads,v,s,t, learning_rate/(1 + i*decay))

        costs.append(cost-0.29)

        if print_cost and i % 50 == 0:
```

```
            print ("Cost after iteration %i: %f" %(i, cost-0.29))

    plt.plot(np.squeeze(costs))
    plt.ylabel('cost')
    plt.xlabel('iterations')
    plt.title("Learning rate =" + str(learning_rate) +" / (1 + i *" + str(decay) + ")")
    plt.show()

    return parameters,AL

def model_run(X, parameters):

    AL, caches = forward_prop(X, parameters)

    return AL
parameters,AL = L_layer_model(X, Y, n_x,n_h,L,n_y, learning_rate = 0.000845, num_iterations = 5000,decay =0.0
7, print_cost=True,dky=True)
ALtest = model_run(Xtest, parameters)

>>
Cost after iteration 0: 0.113093
Cost after iteration 50: 0.105824
Cost after iteration 100: 0.101263
Cost after iteration 150: 0.098146
.

.

.

Cost after iteration 4850: 0.085431
Cost after iteration 4900: 0.085424
Cost after iteration 4950: 0.085422
```

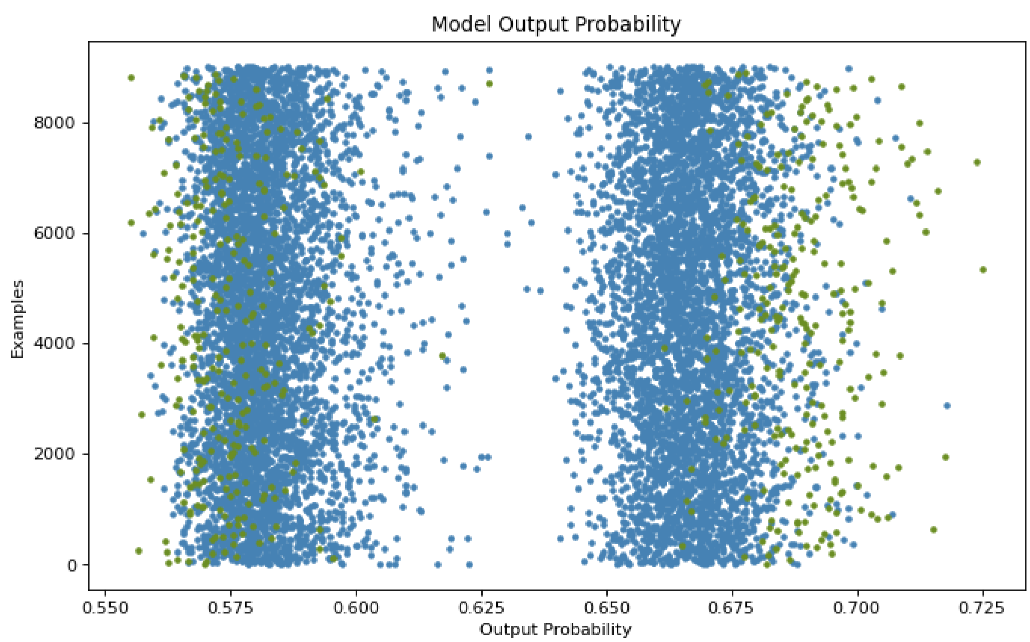



```
At,Atest = AL, ALtest


import csv
filename = "prediction_record.csv"

with open(filename, 'w') as csvfile:

    csvwriter = csv.writer(csvfile)

    csvwriter.writerows(At)
    csvwriter.writerows(Y.T)
    csvwriter.writerows(Atest)
    csvwriter.writerows(Ytest.T)
prediction = np.array(At.T)
prediction_test = np.array(Atest.T)
bias = 0
```

```python
p = 0.5
p_val = []
train_accuracy = []
test_accuracy = []

for acc in range(0,250):
    p = p+0.001
    p_val.append(p+bias)
    for i in range(0, n_y):
        if(At.T[i] > p):
            prediction[i] = 1
        else:
            prediction[i] = 0
    count = 0
    for i in range(0, n_y):
        if(np.round(Y[i],0)==np.round(prediction[i],0)):
            count = count + 1
    accuracy = 100 * (count / n_y)
    #print(accuracy,"train accuracy for p value", p)
    #print(count)
    train_accuracy.append(accuracy)

    for i in range(0, n_y_test):
        if(Atest.T[i] > p):
            prediction_test[i] = 1
        else:
            prediction_test[i] = 0
    count = 0
    for i in range(0, n_y_test):
        if(np.round(Ytest[i],0)==np.round(prediction_test[i],0)):
            count = count + 1
    accuracy = 100 * (count / n_y_test)
    #print(accuracy,"% test accuracy for p value", p)
    #print(count)
    test_accuracy.append(accuracy)


from matplotlib.pyplot import figure
figure(num=None, figsize=(10, 6), dpi=80, facecolor='w', edgecolor='k')
plt.plot(p_val,train_accuracy,"-",label='Training Accuracy',linewidth=1.5,c=mcolors.CSS4_COLORS["steelblue"])
plt.plot(p_val,test_accuracy,"-",label='Test Accuracy',linewidth=1.5,c=mcolors.CSS4_COLORS["olivedrab"])
plt.plot(0.63,100,"rx",label='Optimal p value')
plt.ylabel('% accuracy')
plt.xlabel('P-value (decision thresholde)')
plt.title("Train and Test Accuracy")
plt.legend()
plt.show()
```




```
from matplotlib.pyplot import figure
figure(num=None, figsize=(10, 6), dpi=80, facecolor='w', edgecolor='k')

for i in range(9007):
    #plt.plot(0.63,i,'m.')
    plt.plot(At.T[i],i,'.',c=mcolors.CSS4_COLORS["steelblue"])

for i in range(524):
    plt.plot(Atest.T[i],i*17,'.',c=mcolors.CSS4_COLORS["olivedrab"])

plt.ylabel('Examples')
plt.xlabel('Output Probability')
plt.title("Model Output Probability")
plt.show()
plt.show()
```




```
true_train = At.T * Y
false_train = At.T - true_train
from matplotlib.pyplot import figure
figure(num=None, figsize=(10, 6), dpi=80, facecolor='w', edgecolor='k')

for i in range(9007):
    plt.plot(0.63,i,'.',c=mcolors.CSS4_COLORS["slategray"])
    if true_train[i] != 0:
```

```
        plt.plot(true_train[i],i,'.',c=mcolors.CSS4_COLORS["teal"])
    else:
        plt.plot(false_train[i],i,'.',c=mcolors.CSS4_COLORS["firebrick"])
plt.ylabel('Example #')
plt.xlabel('Output Probability')
plt.title("Training Decision Output")
plt.show()
```


```
true_test = Atest.T * Ytest
false_test = Atest.T - true_test
from matplotlib.pyplot import figure
figure(num=None, figsize=(10, 6), dpi=80, facecolor='w', edgecolor='k')

for i in range(524):
    plt.plot(0.63,i,'.',c=mcolors.CSS4_COLORS["slategray"])
    if true_test[i] != 0:
        plt.plot(true_test[i],i,'.',c=mcolors.CSS4_COLORS["teal"])
    else:
        plt.plot(false_test[i],i,'.',c=mcolors.CSS4_COLORS["firebrick"])
plt.ylabel('Example #')
plt.xlabel('Output Probability')
plt.title("Test Decision Output")
plt.show()
```