

LINK: [https://colab.research.google.com/drive/1ZVd\\_gfTETVdqsD06d\\_mJyHRM66Hu0vd9?usp=sharing](https://colab.research.google.com/drive/1ZVd_gfTETVdqsD06d_mJyHRM66Hu0vd9?usp=sharing)

Activity 1.1 : Neural Networks

Objective(s):

This activity aims to demonstrate the concepts of neural networks

Intended Learning Outcomes (ILOs):

- Demonstrate how to use activation function in neural networks
- Demonstrate how to apply feedforward and backpropagation in neural networks

Resources:

- Jupyter Notebook

Procedure:

Import the libraries

```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

Define and plot an activation function

Sigmoid function:

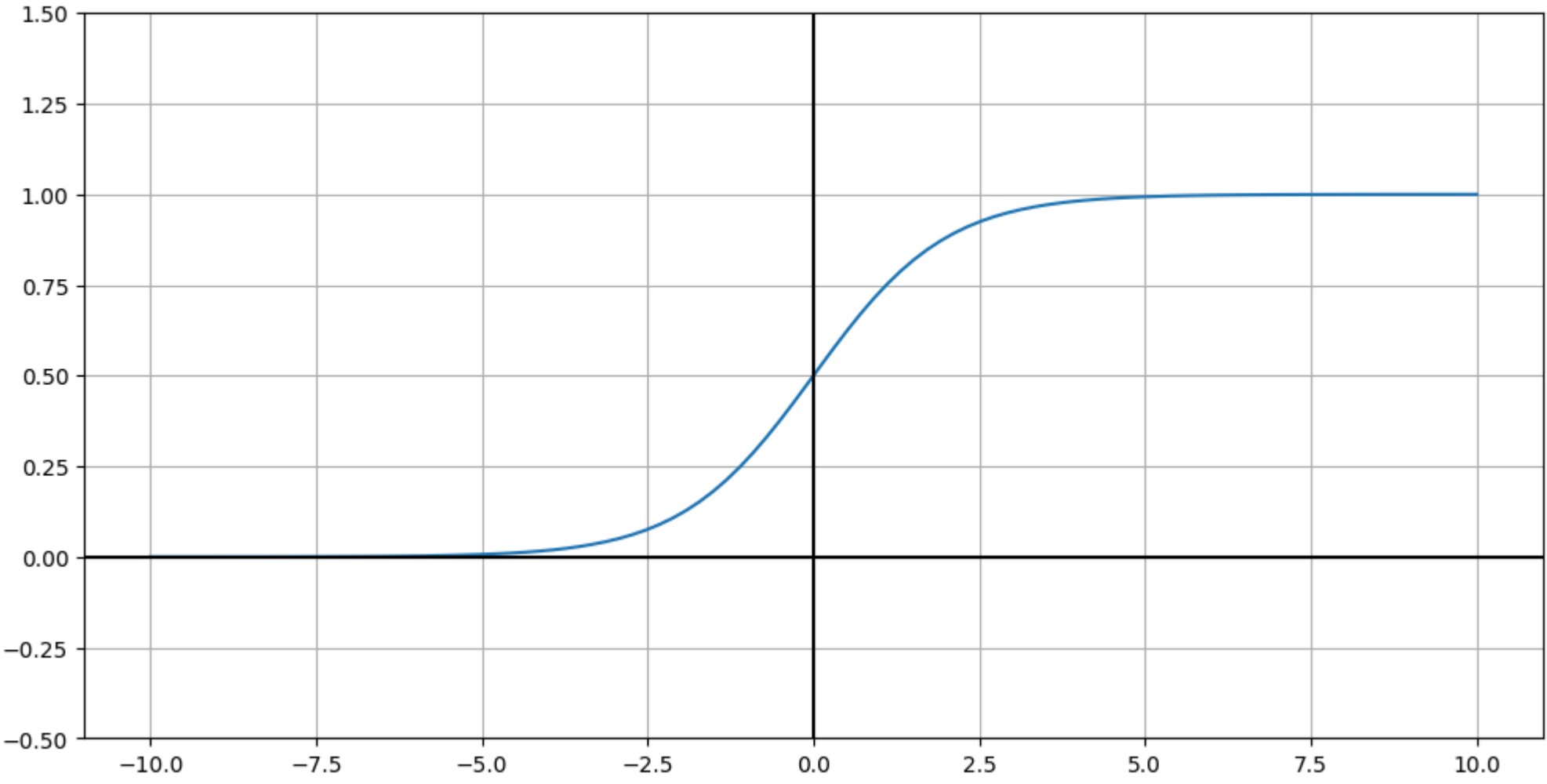
$$\sigma = \frac{1}{1 + e^{-x}}$$

$\sigma$  ranges from (0, 1). When the input  $x$  is negative,  $\sigma$  is close to 0. When  $x$  is positive,  $\sigma$  is close to 1. At  $x = 0, \sigma = 0.5$

```
## create a sigmoid function
def sigmoid(x):
    """Sigmoid function"""
    return 1.0 / (1.0 + np.exp(-x))
```

```
# Plot the sigmoid function
vals = np.linspace(-10, 10, num=100, dtype=np.float32)
activation = sigmoid(vals)
fig = plt.figure(figsize=(12,6))
fig.suptitle('Sigmoid function')
plt.plot(vals, activation)
plt.grid(True, which='both')
plt.axhline(y=0, color='k')
plt.axvline(x=0, color='k')
plt.yticks()
plt.ylim([-0.5, 1.5]);
```

Sigmoid function



Choose any activation function and create a method to define that function.

```
import math
def relu(x):
    return np.maximum(0,x)
def tanh(x):
    return np.tanh(x)
def smax(x):
    exp =[]
    for i in x:
        exp.append(math.exp(i))
    sum_exp = sum(exp)
    result = []
    for i in exp:
        result.append(i / sum_exp)
    return result
```

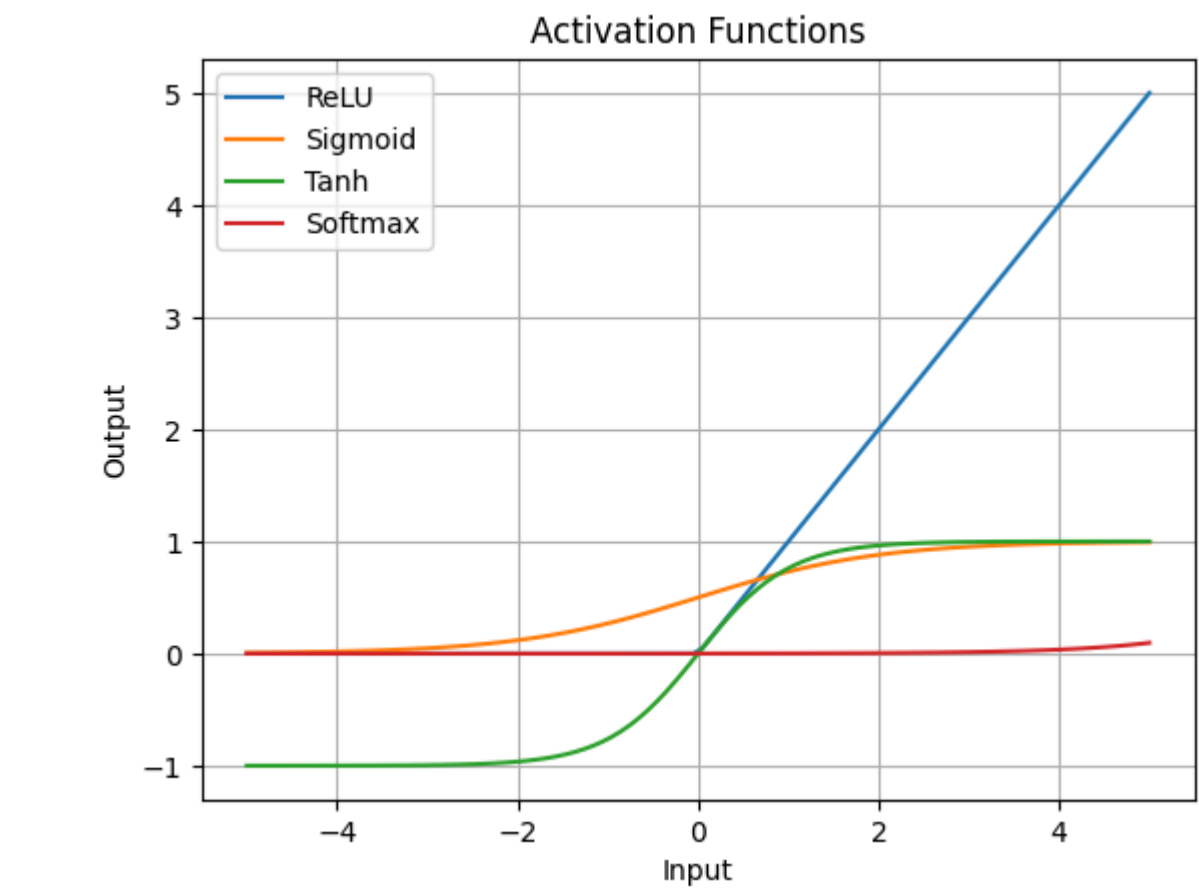
Plot the activation function

```
import matplotlib.pyplot as plt
import numpy as np

#range
x = np.linspace(-5, 5, 100)

# Plot activation functions
plt.plot(x, relu(x), label="ReLU")
plt.plot(x, sigmoid(x), label="Sigmoid")
plt.plot(x, tanh(x), label="Tanh")
plt.plot(x, smax(x), label="Softmax")
plt.legend()
plt.xlabel("Input")
plt.ylabel("Output")
plt.title("Activation Functions")

plt.grid(True)
plt.show()
```



Neurons as boolean logic gates

OR Gate

OR gate truth table		
Input		Output
0	0	0
0	1	1
1	0	1
1	1	1

A neuron that uses the sigmoid activation function outputs a value between (0, 1). This naturally leads us to think about boolean values.

By limiting the inputs of  $x_1$  and  $x_2$  to be in  $\{0, 1\}$ , we can simulate the effect of logic gates with our neuron. The goal is to find the weights, such that it returns an output close to 0 or 1 depending on the inputs.

What numbers for the weights would we need to fill in for this gate to output OR logic? Observe from the plot above that  $\sigma(z)$  is close to 0 when  $z$  is largely negative (around -10 or less), and is close to 1 when  $z$  is largely positive (around +10 or greater).

$$z = w_1x_1 + w_2x_2 + b$$

Let's think this through:

- When  $x_1$  and  $x_2$  are both 0, the only value affecting  $z$  is  $b$ . Because we want the result for (0, 0) to be close to zero,  $b$  should be negative (at least -10)
- If either  $x_1$  or  $x_2$  is 1, we want the output to be close to 1. That means the weights associated with  $x_1$  and  $x_2$  should be enough to offset  $b$  to the point of causing  $z$  to be at least 10.
- Let's give  $b$  a value of -10. How big do we need  $w_1$  and  $w_2$  to be?
  - At least +20
- So let's try out  $w_1 = 20, w_2 = 20$ , and  $b = -10$ !

```
def logic_gate(w1, w2, b):
    # Helper to create logic gate functions
    # Plug in values for weight_a, weight_b, and bias
    return lambda x1, x2: sigmoid(w1 * x1 + w2 * x2 + b)

def test(gate):
    # Helper function to test out our weight functions.
    for a, b in (0, 0), (0, 1), (1, 0), (1, 1):
        print("{} {}: {}".format(a, b, np.round(gate(a, b))))
```

```
or_gate = logic_gate(20, 20, -10)
test(or_gate)
```

```
0, 0: 0.0
0, 1: 1.0
1, 0: 1.0
1, 1: 1.0
```

OR gate truth table		
Input		Output
0	0	0
0	1	1
1	0	1
1	1	1

Try finding the appropriate weight values for each truth table.

AND Gate

AND gate truth table		
Input		Output
0	0	0
0	1	0
1	0	0
1	1	1

Try to figure out what values for the neurons would make this function as an AND gate.

```
# Fill in the w1, w2, and b parameters such that the truth table matches
w1 = 2
w2 = 2
b = -3
and_gate = logic_gate(w1, w2, b)

test(and_gate)
```

```
0, 0: 0.0
0, 1: 0.0
1, 0: 0.0
1, 1: 1.0
```

Do the same for the NOR gate and the NAND gate.

```
# Fill in the w1, w2, and b parameters such that the truth table matches
w1 = -2
w2 = -3
b = 2
nor_gate = logic_gate(w1, w2, b)

test(nor_gate)
```

```
0, 0: 1.0
0, 1: 0.0
1, 0: 0.0
1, 1: 0.0
```

```
# Fill in the w1, w2, and b parameters such that the truth table matches
w1 = -2
w2 = -2
b = 3
nand_gate = logic_gate(w1, w2, b)

test(nand_gate)
```

```
0, 0: 1.0
0, 1: 1.0
1, 0: 1.0
1, 1: 0.0
```

Limitation of single neuron

Here's the truth table for XOR:

XOR (Exclusive Or) Gate

XOR gate truth table		
Input		Output
0	0	0
0	1	1
1	0	1
1	1	0

Now the question is, can you create a set of weights such that a single neuron can output this property?

It turns out that you cannot. Single neurons can't correlate inputs, so it's just confused. So individual neurons are out. Can we still use neurons to somehow form an XOR gate?

```
# Make sure you have or_gate, nand_gate, and and_gate working from above!
def xor_gate(a, b):
    c = or_gate(a, b)
    d = nand_gate(a, b)
    return and_gate(c, d)
test(xor_gate)
```

```
0, 0: 0.0
0, 1: 1.0
1, 0: 1.0
1, 1: 0.0
```

Feedforward Networks

The feed-forward computation of a neural network can be thought of as matrix calculations and activation functions. We will do some actual computations with matrices to see this in action.

Exercise

Provided below are the following:

- Three weight matrices  $W_1$ ,  $W_2$  and  $W_3$  representing the weights in each layer. The convention for these matrices is that each  $W_{i,j}$  gives the weight from neuron  $i$  in the previous (left) layer to neuron  $j$  in the next (right) layer.
- A vector  $x_{in}$  representing a single input and a matrix  $x_{mat\_in}$  representing 7 different inputs.
- Two functions: `soft_max_vec` and `soft_max_mat` which apply the `soft_max` function to a single vector, and row-wise to a matrix.

The goals for this exercise are:

- For input  $x_{in}$  calculate the inputs and outputs to each layer (assuming sigmoid activations for the middle two layers and `soft_max` output for the final layer).
- Write a function that does the entire neural network calculation for a single input
- Write a function that does the entire neural network calculation for a matrix of inputs, where each row is a single input.
- Test your functions on  $x_{in}$  and  $x_{mat\_in}$ .

This illustrates what happens in a NN during one single forward pass. Roughly speaking, after this forward pass, it remains to compare the output of the network to the known truth values, compute the gradient of the loss function and adjust the weight matrices  $W_1$ ,  $W_2$  and  $W_3$  accordingly, and iterate. Hopefully this process will result in better weight matrices and our loss will be smaller afterwards

```
W_1 = np.array([[2,-1,1,4],[-1,2,-3,1],[3,-2,-1,5]])
W_2 = np.array([[3,1,-2,1],[-2,4,1,-4],[-1,-3,2,-5],[3,1,1,1]])
W_3 = np.array([[ -1,3,-2],[1,-1,-3],[3,-2,2],[1,2,1]])
x_in = np.array([.5,.8,.2])
x_mat_in = np.array([[.5,.8,.2],[.1,.9,.6],[.2,.2,.3],[.6,.1,.9],[.5,.5,.4],[.9,.1,.9],[.1,.8,.7]])
```

```
def soft_max_vec(vec):
    return np.exp(vec)/(np.sum(np.exp(vec)))

def soft_max_mat(mat):
    return np.exp(mat)/(np.sum(np.exp(mat),axis=1).reshape(-1,1))
```

```
print('the matrix W_1\n')
print(W_1)
print('-'*30)
print('vector input x_in\n')
print(x_in)
print ('-'*30)
print('matrix input x_mat_in -- starts with the vector `x_in`\n')
print(x_mat_in)
```

```
the matrix W_1

[[ 2 -1  1  4]
 [-1  2 -3  1]
 [ 3 -2 -1  5]]
-----
vector input x_in

[0.5 0.8 0.2]
-----
matrix input x_mat_in -- starts with the vector `x_in`

[[0.5 0.8 0.2]
 [0.1 0.9 0.6]
 [0.2 0.2 0.3]
 [0.6 0.1 0.9]
 [0.5 0.5 0.4]
 [0.9 0.1 0.9]
 [0.1 0.8 0.7]]
```

Exercise

- Get the product of array  $x_{in}$  and  $W_1$  ( $z_2$ )
- Apply sigmoid function to  $z_2$  that results to  $a_2$
- Get the product of  $a_2$  and  $z_2$  ( $z_3$ )
- Apply sigmoid function to  $z_3$  that results to  $a_3$
- Get the product of  $a_3$  and  $z_3$  that results to  $z_4$

```
#1. Get the product of array x_in and W_1 (z2)
z2 = np.dot(x_in, W_1)
```

```
print(z2)

[ 0.8  0.7 -2.1  3.8]
```

```
#2. Apply sigmoid function to z2 that results to a2
a2 = sigmoid(z2)
print(a2)
```

```
[0.68997448 0.66818777 0.10909682 0.97811873]
```

```
#3. Get the product of a2 and z2 (z3)
z3 = np.dot(a2, z2)
```

```
print(z3)

4.507458871351723
```

```
#4. Apply sigmoid function to z3 that results to a3
a3 = sigmoid(z3)
print(a3)
```

```
0.9890938122523221
```

```
#5. Get the product of a3 and z3 that results to z4
z4 = np.dot(a3, z3)
```

```
print(z4)

4.458299678635824
```

```
def soft_max_vec(vec):
    return np.exp(vec)/(np.sum(np.exp(vec)))

def soft_max_mat(mat):
    return np.exp(mat)/(np.sum(np.exp(mat),axis=1).reshape(-1,1))
```

- Apply `soft_max_vec` function to  $z_4$  that results to  $y_{out}$

```
y_out = soft_max_vec(z4)
print(y_out)
```

```
1.0
```

## A one-line function to do the entire neural net computation

```
def nn_comp_vec(x):
    return soft_max_vec(sigmoid(sigmoid(np.dot(x,W_1)).dot(W_2)).dot(W_3))

def nn_comp_mat(x):
    return soft_max_mat(sigmoid(sigmoid(np.dot(x,W_1)).dot(W_2)).dot(W_3))
```

```
nn_comp_vec(x_in)
```

```
array([[0.72780576, 0.26927918, 0.00291506]])
```

```
nn_comp_mat(x_mat_in)

array([[0.72780576, 0.26927918, 0.00291506],
       [0.62054212, 0.37682531, 0.00263257],
       [0.69267581, 0.30361576, 0.00370844],
       [0.36618794, 0.63016955, 0.00364252],
       [0.57199769, 0.4251982 , 0.00280411],
       [0.38373781, 0.61163804, 0.00462415],
       [0.52510443, 0.4725011 , 0.00239447]])
```

▼ Backpropagation

The backpropagation in this part will be used to train a multi-layer perceptron (with a single hidden layer). Different patterns will be used and the demonstration on how the weights will converge. The different parameters such as learning rate, number of iterations, and number of data points will be demonstrated

```
#Preliminaries
from __future__ import division, print_function
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

Fill out the code below so that it creates a multi-layer perceptron with a single hidden layer (with 4 nodes) and trains it via back-propagation. Specifically your code should:

1. Initialize the weights to random values between -1 and 1
2. Perform the feed-forward computation
3. Compute the loss function
4. Calculate the gradients for all the weights via back-propagation
5. Update the weight matrices (using a learning\_rate parameter)
6. Execute steps 2-5 for a fixed number of iterations
7. Plot the accuracies and log loss and observe how they change over time

Once your code is running, try it for the different patterns below.

- Which patterns was the neural network able to learn quickly and which took longer?
- What learning rates and numbers of iterations worked well?

```
## This code below generates two x values and a y value according to different patterns
## It also creates a "bias" term (a vector of 1s)
## The goal is then to learn the mapping from x to y using a neural network via back-propagation
```

```
num_obs = 500
x_mat_1 = np.random.uniform(-1,1,size = (num_obs,2))
x_mat_bias = np.ones((num_obs,1))
x_mat_full = np.concatenate( (x_mat_1,x_mat_bias), axis=1)
```

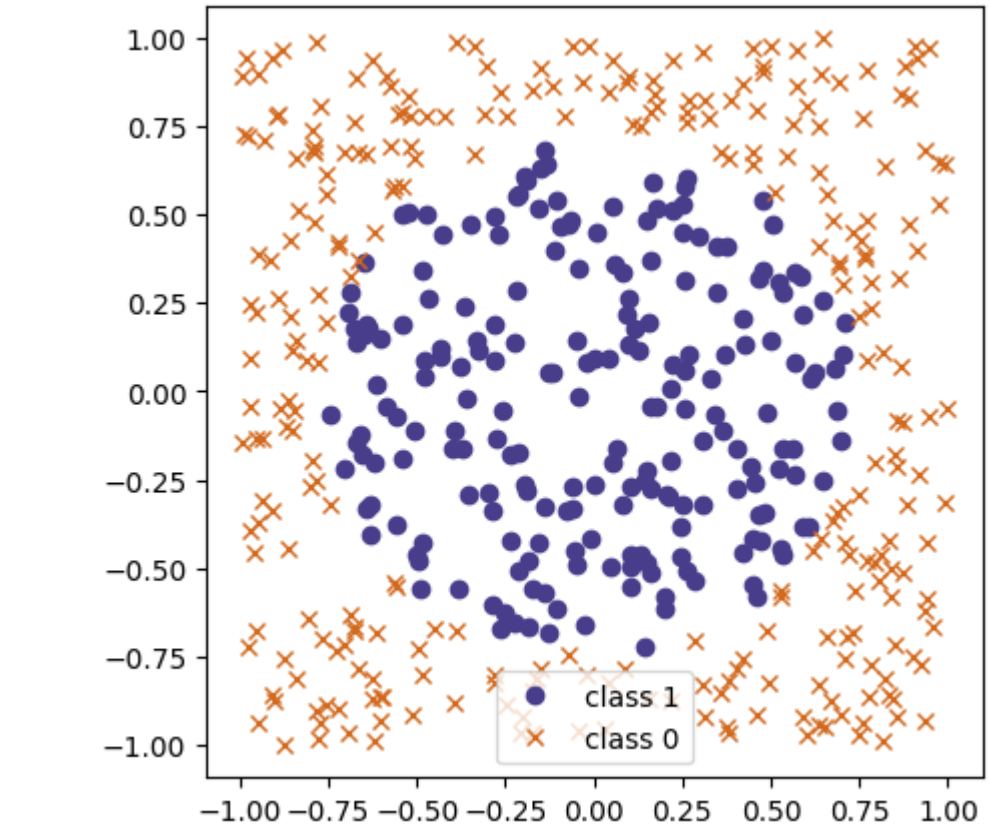
# PICK ONE PATTERN BELOW and comment out the rest.

```
# # Circle pattern
y = (np.sqrt(x_mat_full[:,0]**2 + x_mat_full[:,1]**2)<.75).astype(int)
```

```
print('shape of x_mat_full is {}'.format(x_mat_full.shape))
print('shape of y is {}'.format(y.shape))
```

```
fig, ax = plt.subplots(figsize=(5, 5))
ax.plot(x_mat_full[y==1, 0],x_mat_full[y==1, 1], 'ro', label='class 1', color='darkslateblue')
ax.plot(x_mat_full[y==0, 0],x_mat_full[y==0, 1], 'bx', label='class 0', color='chocolate')
# ax.grid(True)
ax.legend(loc='best')
ax.axis('equal');
```

```
shape of x_mat_full is (500, 3)
shape of y is (500,)
<ipython-input-55-b5d4fe162b68>:20: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "ro" (-> color='r'). The keyword argument will take precedence.
  ax.plot(x_mat_full[y==1, 0],x_mat_full[y==1, 1], 'ro', label='class 1', color='darkslateblue')
<ipython-input-55-b5d4fe162b68>:21: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "bx" (-> color='b'). The keyword argument will take precedence.
  ax.plot(x_mat_full[y==0, 0],x_mat_full[y==0, 1], 'bx', label='class 0', color='chocolate')
```



```
num_obs = 500
x_mat_1 = np.random.uniform(-1,1,size = (num_obs,2))
x_mat_bias = np.ones((num_obs,1))
x_mat_full = np.concatenate( (x_mat_1,x_mat_bias), axis=1)
```

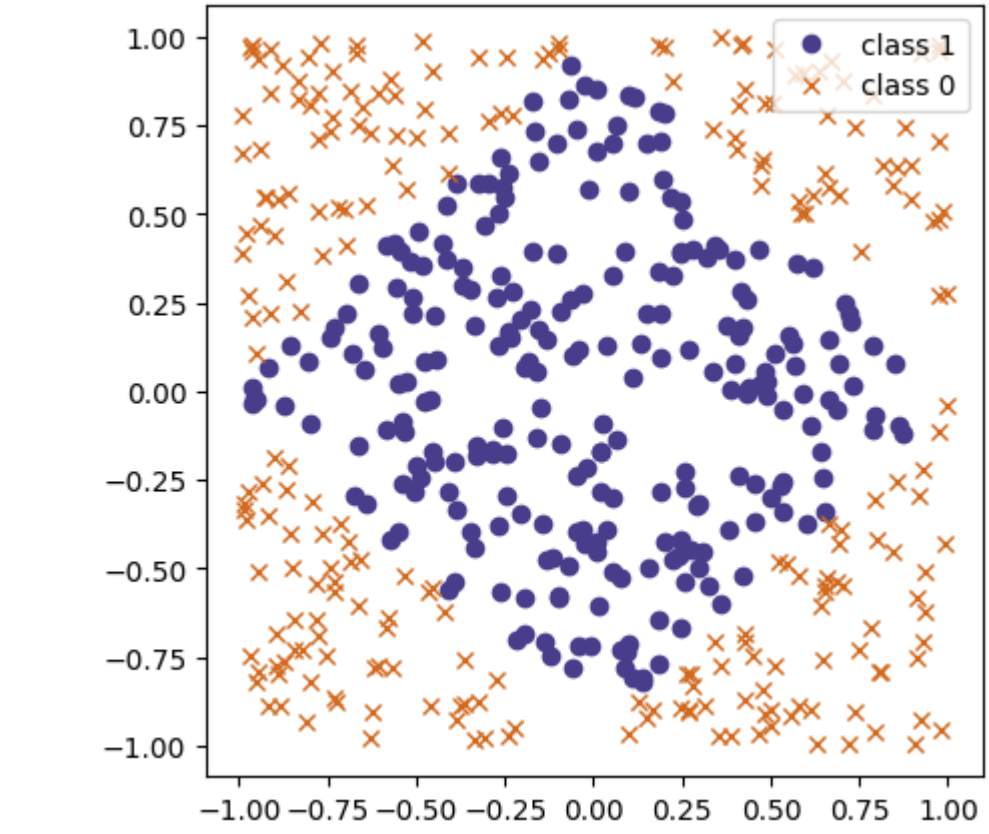
# PICK ONE PATTERN BELOW and comment out the rest.

```
# # Diamond Pattern
y = ((np.abs(x_mat_full[:,0]) + np.abs(x_mat_full[:,1]))<1).astype(int)
```

```
print('shape of x_mat_full is {}'.format(x_mat_full.shape))
print('shape of y is {}'.format(y.shape))
```

```
fig, ax = plt.subplots(figsize=(5, 5))
ax.plot(x_mat_full[y==1, 0],x_mat_full[y==1, 1], 'ro', label='class 1', color='darkslateblue')
ax.plot(x_mat_full[y==0, 0],x_mat_full[y==0, 1], 'bx', label='class 0', color='chocolate')
# ax.grid(True)
ax.legend(loc='best')
ax.axis('equal');
```

```
shape of x_mat_full is (500, 3)
shape of y is (500,)
<ipython-input-56-34db6cc40e36>:15: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "ro" (-> color='r'). The keyword argument will take precedence.
  ax.plot(x_mat_full[y==1, 0],x_mat_full[y==1, 1], 'ro', label='class 1', color='darkslateblue')
<ipython-input-56-34db6cc40e36>:16: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "bx" (-> color='b'). The keyword argument will take precedence.
  ax.plot(x_mat_full[y==0, 0],x_mat_full[y==0, 1], 'bx', label='class 0', color='chocolate')
```





```
num_obs = 500
x_mat_1 = np.random.uniform(-1,1,size = (num_obs,2))
x_mat_bias = np.ones((num_obs,1))
x_mat_full = np.concatenate( (x_mat_1,x_mat_bias), axis=1)

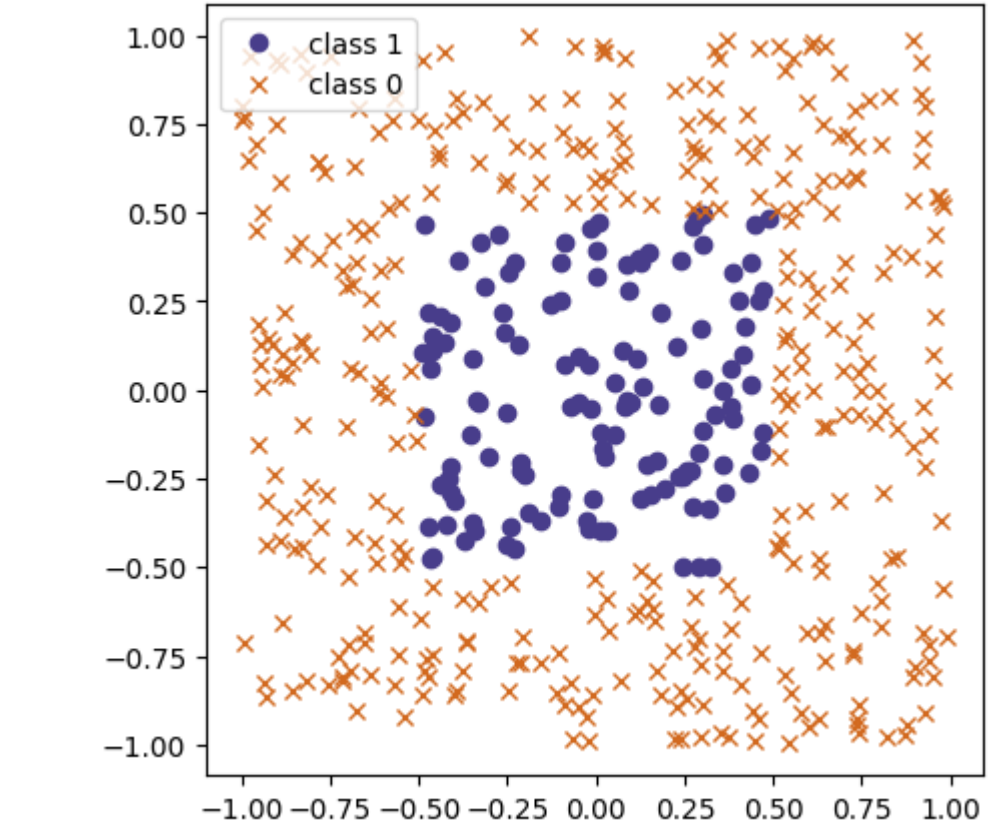
# PICK ONE PATTERN BELOW and comment out the rest.

# # Centered square
y = ((np.maximum(np.abs(x_mat_full[:,0]), np.abs(x_mat_full[:,1])))<.5).astype(int))
```

```
print('shape of x_mat_full is {}'.format(x_mat_full.shape))
print('shape of y is {}'.format(y.shape))
```

```
fig, ax = plt.subplots(figsize=(5, 5))
ax.plot(x_mat_full[y==1, 0],x_mat_full[y==1, 1], 'ro', label='class 1', color='darkslateblue')
ax.plot(x_mat_full[y==0, 0],x_mat_full[y==0, 1], 'bx', label='class 0', color='chocolate')
# ax.grid(True)
ax.legend(loc='best')
ax.axis('equal');
```

```
shape of x_mat_full is (500, 3)
shape of y is (500,)
<ipython-input-57-64d51d37595c>:16: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "ro" (-> color='r'). The keyword argument will take precedence.
  ax.plot(x_mat_full[y==1, 0],x_mat_full[y==1, 1], 'ro', label='class 1', color='darkslateblue')
<ipython-input-57-64d51d37595c>:17: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "bx" (-> color='b'). The keyword argument will take precedence.
  ax.plot(x_mat_full[y==0, 0],x_mat_full[y==0, 1], 'bx', label='class 0', color='chocolate')
```



```
num_obs = 500
x_mat_1 = np.random.uniform(-1,1,size = (num_obs,2))
x_mat_bias = np.ones((num_obs,1))
x_mat_full = np.concatenate( (x_mat_1,x_mat_bias), axis=1)

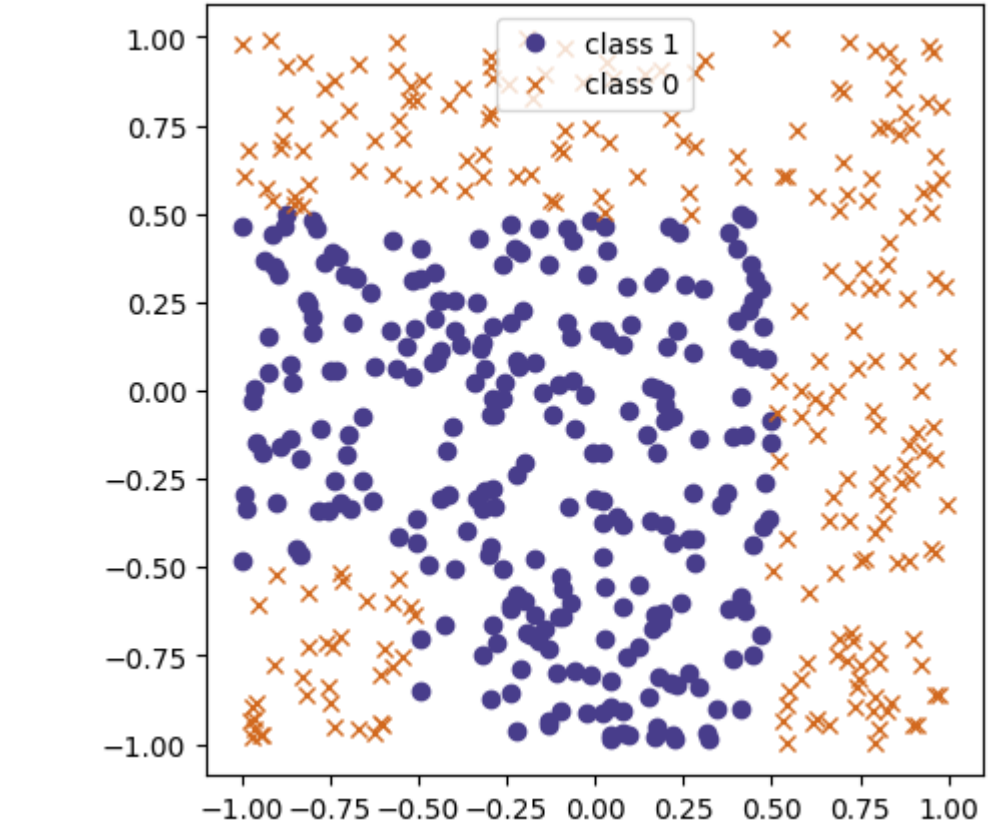
# PICK ONE PATTERN BELOW and comment out the rest.

# # Thick Right Angle pattern
y = (((np.maximum((x_mat_full[:,0]), (x_mat_full[:,1])))<.5) & ((np.maximum((x_mat_full[:,0]), (x_mat_full[:,1]))>-.5)).astype(int))
```

```
print('shape of x_mat_full is {}'.format(x_mat_full.shape))
print('shape of y is {}'.format(y.shape))
```

```
fig, ax = plt.subplots(figsize=(5, 5))
ax.plot(x_mat_full[y==1, 0],x_mat_full[y==1, 1], 'ro', label='class 1', color='darkslateblue')
ax.plot(x_mat_full[y==0, 0],x_mat_full[y==0, 1], 'bx', label='class 0', color='chocolate')
# ax.grid(True)
ax.legend(loc='best')
ax.axis('equal');
```

```
shape of x_mat_full is (500, 3)
shape of y is (500,)
<ipython-input-58-e46456d7e6db>:15: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "ro" (-> color='r'). The keyword argument will take precedence.
  ax.plot(x_mat_full[y==1, 0],x_mat_full[y==1, 1], 'ro', label='class 1', color='darkslateblue')
<ipython-input-58-e46456d7e6db>:16: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "bx" (-> color='b'). The keyword argument will take precedence.
  ax.plot(x_mat_full[y==0, 0],x_mat_full[y==0, 1], 'bx', label='class 0', color='chocolate')
```



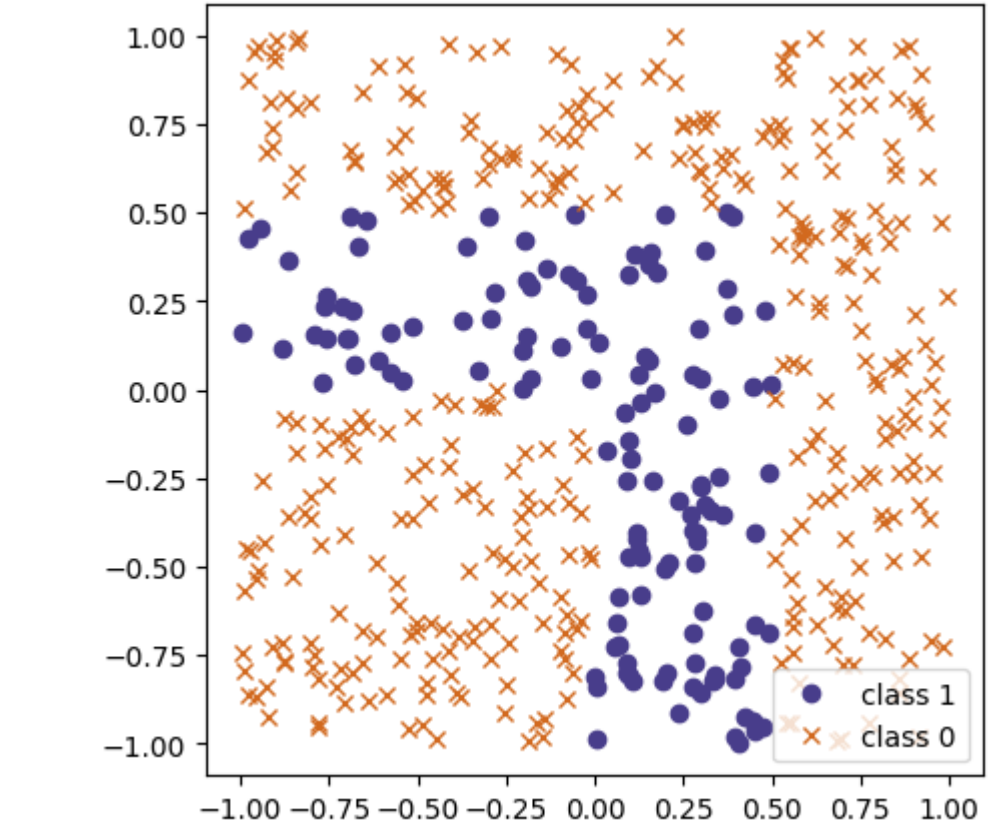
```
num_obs = 500
x_mat_1 = np.random.uniform(-1,1,size = (num_obs,2))
x_mat_bias = np.ones((num_obs,1))
x_mat_full = np.concatenate( (x_mat_1,x_mat_bias), axis=1)

# # Thin right angle pattern
y = (((np.maximum((x_mat_full[:,0]), (x_mat_full[:,1]))<.5) & ((np.maximum((x_mat_full[:,0]), (x_mat_full[:,1]))>0)).astype(int))
```

```
print('shape of x_mat_full is {}'.format(x_mat_full.shape))
print('shape of y is {}'.format(y.shape))
```

```
fig, ax = plt.subplots(figsize=(5, 5))
ax.plot(x_mat_full[y==1, 0],x_mat_full[y==1, 1], 'ro', label='class 1', color='darkslateblue')
ax.plot(x_mat_full[y==0, 0],x_mat_full[y==0, 1], 'bx', label='class 0', color='chocolate')
# ax.grid(True)
ax.legend(loc='best')
ax.axis('equal');
```

```
shape of x_mat_full is (500, 3)
shape of y is (500,)
<ipython-input-59-a43b15cdf0b0>:14: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "ro" (-> color='r'). The keyword argument will take precedence.
  ax.plot(x_mat_full[y==1, 0],x_mat_full[y==1, 1], 'ro', label='class 1', color='darkslateblue')
<ipython-input-59-a43b15cdf0b0>:15: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "bx" (-> color='b'). The keyword argument will take precedence.
  ax.plot(x_mat_full[y==0, 0],x_mat_full[y==0, 1], 'bx', label='class 0', color='chocolate')
```



```
def sigmoid(x):
    """
    Sigmoid function
    """
    return 1.0 / (1.0 + np.exp(-x))

def loss_fn(y_true, y_pred, eps=1e-16):
    """
    Loss function we would like to optimize (minimize)
    We are using Logarithmic Loss
    http://scikit-learn.org/stable/modules/model_evaluation.html#log-loss
    """
    y_pred = np.maximum(y_pred,eps)
    y_pred = np.minimum(y_pred,(1-eps))
    return -(np.sum(y_true * np.log(y_pred)) + np.sum((1-y_true)*np.log(1-y_pred)))/len(y_true)

def forward_pass(W1, W2):
    """
    Does a forward computation of the neural network
    Takes the input `x_mat` (global variable) and produces the output `y_pred`
    Also produces the gradient of the log loss function
    """
    global x_mat
    global y
    global num_
    # First, compute the new predictions `y_pred`
    z_2 = np.dot(x_mat, W_1)
    a_2 = sigmoid(z_2)
    z_3 = np.dot(a_2, W_2)
    y_pred = sigmoid(z_3).reshape((len(x_mat),))
    # Now compute the gradient
    J_z_3_grad = -y + y_pred
    J_W_2_grad = np.dot(J_z_3_grad, a_2)
    a_2_z_2_grad = sigmoid(z_2)*(1-sigmoid(z_2))
    J_W_1_grad = (np.dot((J_z_3_grad).reshape(-1,1), W_2.reshape(-1,1).T)*a_2_z_2_grad).T.dot(x_mat).T
    gradient = (J_W_1_grad, J_W_2_grad)

    # return
    return y_pred, gradient

def plot_loss_accuracy(loss_vals, accuracies):
    fig = plt.figure(figsize=(16, 8))
    fig.suptitle('Log Loss and Accuracy over iterations')

    ax = fig.add_subplot(1, 2, 1)
    ax.plot(loss_vals)
    ax.grid(True)
    ax.set(xlabel='iterations', title='Log Loss')

    ax = fig.add_subplot(1, 2, 2)
    ax.plot(accuracies)
    ax.grid(True)
    ax.set(xlabel='iterations', title='Accuracy');
```

Complete the pseudocode below

```
#### Initialize the network parameters

np.random.seed(1241)

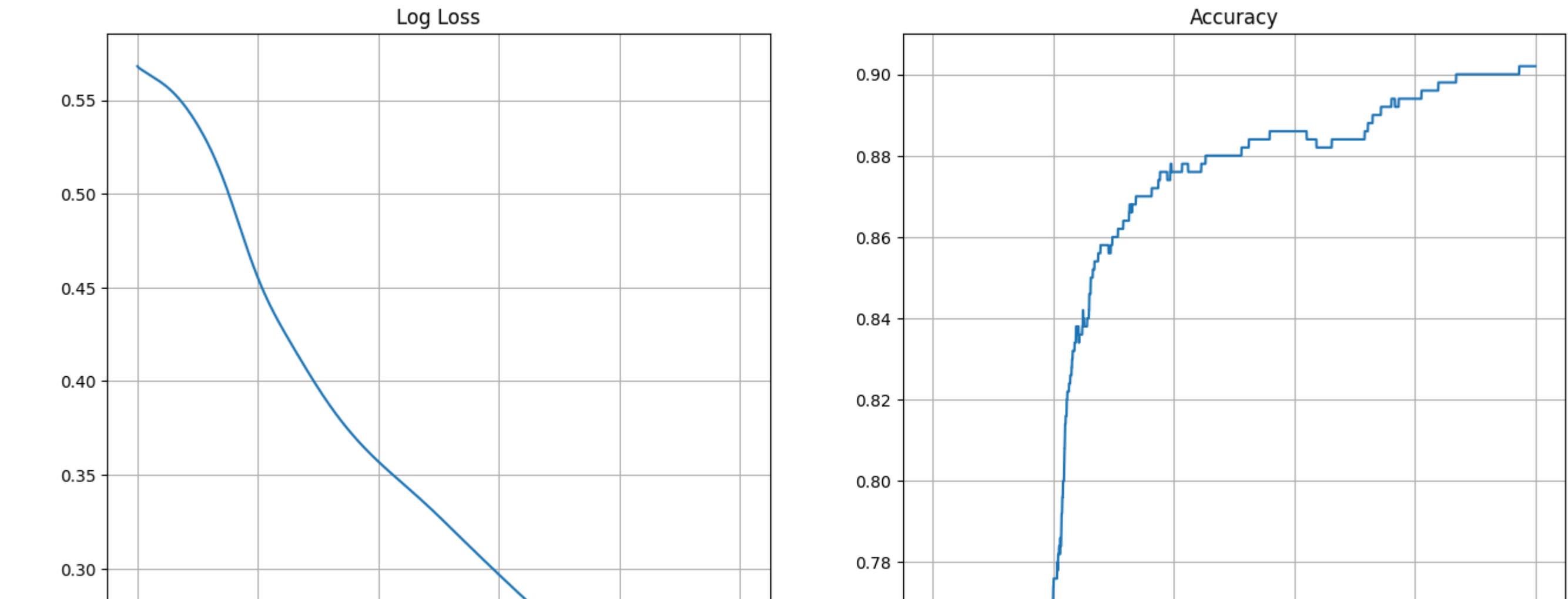
W_1 = np.random.uniform(-1,1,size = (3,4))
W_2 = np.random.uniform(-1,1,size = (4))
num_iter = 5000
learning_rate = 0.001
x_mat = x_mat_full

loss_vals, accuracies = [], []
for i in range(num_iter):
    ### Do a forward computation, and get the gradient
    y_pred, (J_W_1_grad, J_W_2_grad) = forward_pass(W_1, W_2)
    ## Update the weight matrices
    W_1 = W_1 - learning_rate*J_W_1_grad
    W_2 = W_2 - learning_rate*J_W_2_grad
    ### Compute the loss and accuracy
    loss = loss_fn(y, y_pred)
    loss_vals.append(loss)
    accuracy = np.sum((y_pred >= 0.5) == y) / num_obs
    accuracies.append(accuracy)
    ## Print the loss and accuracy for every 200th iteration
    for i, (loss, accuracy) in enumerate(zip(loss_vals, accuracies), 1):
        if ((i%200) == 0):
            print('log loss = {:.4f}, accuracy = {}'.format(
                i*200, loss, accuracy
            ))

plot_loss_accuracy(loss_vals, accuracies)

log loss = 40000.0000, accuracy = 0.5592968448265846
log loss = 80000.0000, accuracy = 0.5463256944022675
log loss = 120000.0000, accuracy = 0.5247450867763153
log loss = 160000.0000, accuracy = 0.4920078195062901
log loss = 200000.0000, accuracy = 0.4551154808237848
log loss = 240000.0000, accuracy = 0.42845468355793925
log loss = 280000.0000, accuracy = 0.406605109710444
log loss = 320000.0000, accuracy = 0.38675982188171304
log loss = 360000.0000, accuracy = 0.3703351491227839
log loss = 400000.0000, accuracy = 0.3570678394027878
log loss = 440000.0000, accuracy = 0.3455271042944421
log loss = 480000.0000, accuracy = 0.33402553061965873
log loss = 520000.0000, accuracy = 0.3217031048488217
log loss = 560000.0000, accuracy = 0.30912546120546225
log loss = 600000.0000, accuracy = 0.2968021800978604
log loss = 640000.0000, accuracy = 0.28473830809119965
log loss = 680000.0000, accuracy = 0.27334237311676407
log loss = 720000.0000, accuracy = 0.2628820023485311
log loss = 760000.0000, accuracy = 0.2533634268381629
log loss = 800000.0000, accuracy = 0.24486147989791002
log loss = 840000.0000, accuracy = 0.2374398812274377
log loss = 880000.0000, accuracy = 0.23104309715443921
log loss = 920000.0000, accuracy = 0.22553323656443525
log loss = 960000.0000, accuracy = 0.2207532526785259
log loss = 1000000.0000, accuracy = 0.21656139869873695
```

Log Loss and Accuracy over iterations

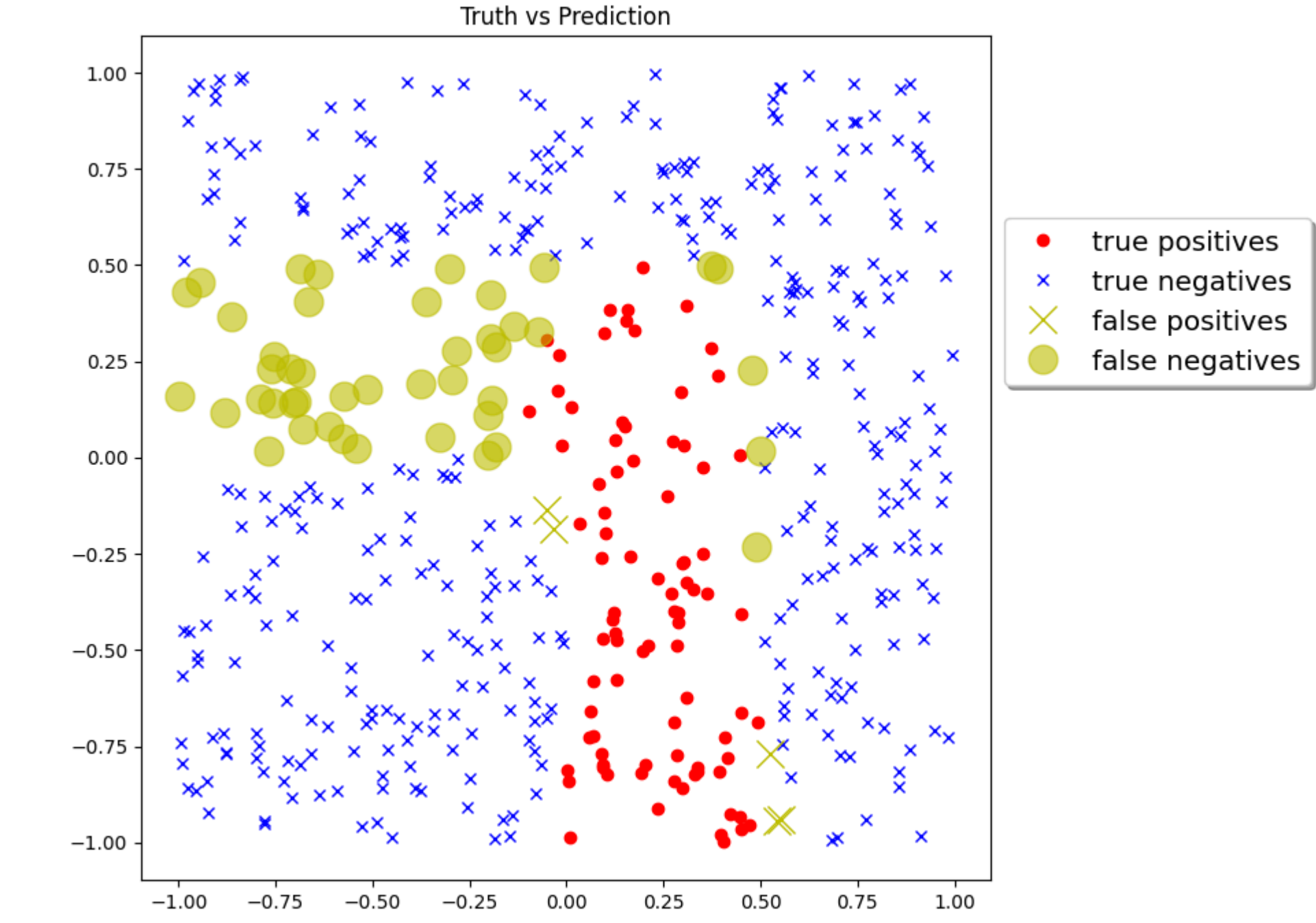


Plot the predicted answers, with mistakes in yellow

```
pred1 = (y_pred>=.5)
pred0 = (y_pred<.5)

fig, ax = plt.subplots(figsize=(8, 8))
# true predictions
ax.plot(x_mat[pred1 & (y==1),0],x_mat[pred1 & (y==1),1], 'ro', label='true positives')
ax.plot(x_mat[pred0 & (y==0),0],x_mat[pred0 & (y==0),1], 'bx', label='true negatives')
# false predictions
ax.plot(x_mat[pred1 & (y==0),0],x_mat[pred1 & (y==0),1], 'yx', label='false positives', markersize=15)
ax.plot(x_mat[pred0 & (y==1),0],x_mat[pred0 & (y==1),1], 'yo', label='false negatives', markersize=15, alpha=.6)
ax.set(title='Truth vs Prediction')
ax.legend(bbox_to_anchor=(1, 0.8), fancybox=True, shadow=True, fontsize='x-large');
```





Once your code is running, try it for the different patterns below.

- Which patterns was the neural network able to learn quickly and which took longer?
- What learning rates and numbers of iterations worked well?

**Answer:** The patterns that was able to quickly learn by the neural network was the circle and centered square since they have this shape that is easy to do since it the data would just pile each other or form a circle-like shape while the patterns that took longer are the diamond-shaped because it requires a detailed function and operations in order to achieve the shaped-like graph.

**Answer:** Some of the learning that worked well for me is 0.01 and 0.001 while the number of iterations should be a lot like 2000 and above iteration

#### Supplementary Activity

- Use a different weights , input and activation function
- Apply feedforward and backpropagation
- Plot the loss and accuracy for every 300th iteration

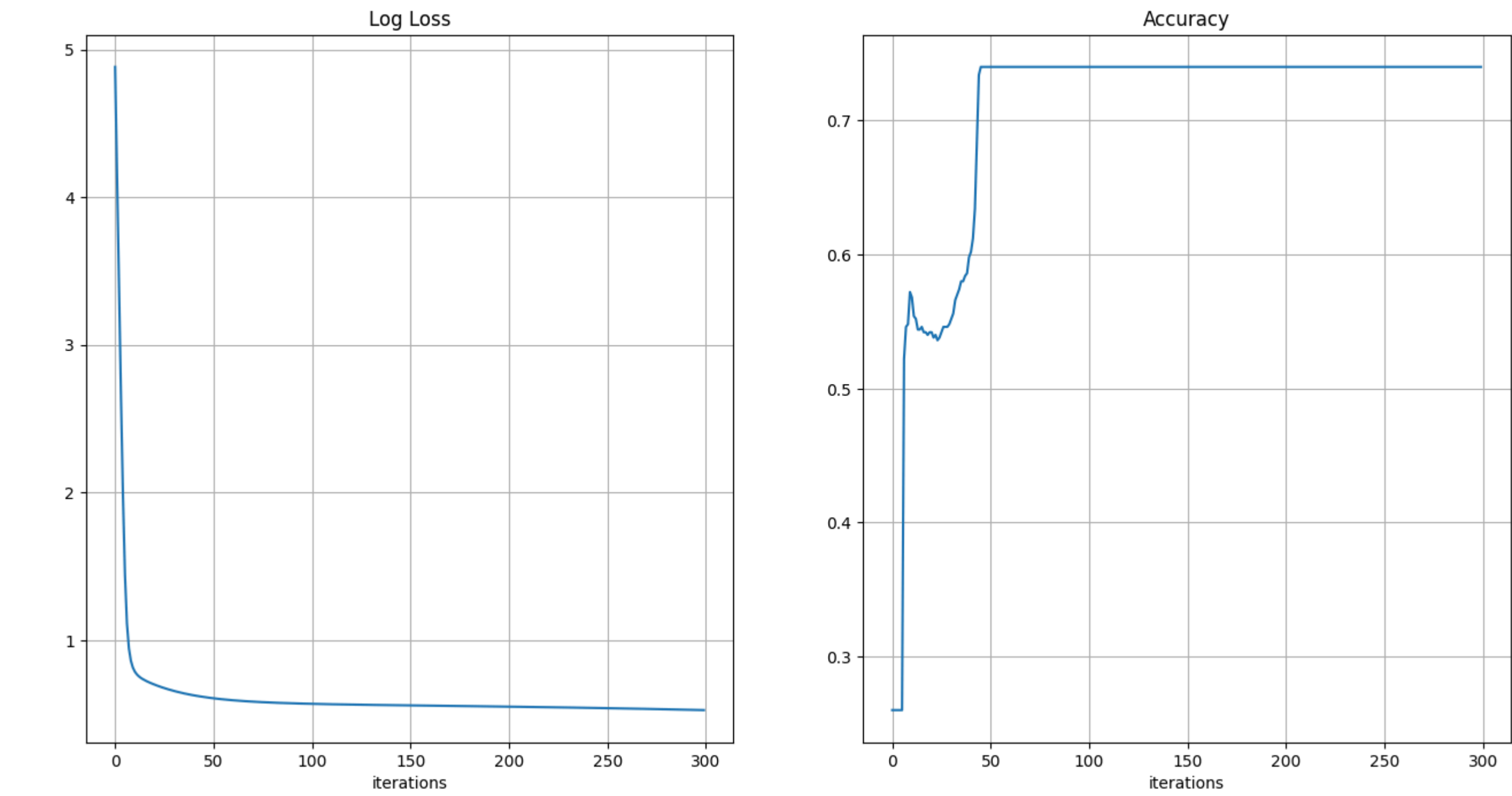
```
def sigmoid(x):  
    return 1.0 / (1.0 + np.exp(-x))  
  
def loss_fn(y_true, y_pred, eps=1e-16):  
    """  
    Loss function we would like to optimize (minimize)  
    We are using Logarithmic Loss  
    http://scikit-learn.org/stable/modules/model_evaluation.html#log-loss  
    """  
    y_pred = np.maximum(y_pred, eps)  
    y_pred = np.minimum(y_pred, (1-eps))  
    return -(np.sum(y_true * np.log(y_pred)) + np.sum((1-y_true)*np.log\  
        (1-y_pred)))/len(y_true)  
  
def forward_pass(W1, W2):  
    """  
    Does a forward computation of the neural network  
    Takes the input `x_mat` (global variable) and produces the output `y_pred`  
    Also produces the gradient of the log loss function  
    """  
    global x_mat  
    global y  
    global num_  
    # First, compute the new predictions `y_pred`  
    z_2 = np.dot(x_mat, W1)  
    a_2 = sigmoid(z_2)  
    z_3 = np.dot(a_2, W2)  
    y_pred = sigmoid(z_3).reshape((len(x_mat),))  
    # Now compute the gradient  
    J_z_3_grad = -y + y_pred  
    J_W_2_grad = np.dot(J_z_3_grad, a_2)  
    a_2_z_2_grad = sigmoid(z_2)*(1-sigmoid(z_2))  
    J_W_1_grad = (np.dot((J_z_3_grad).reshape(-1,1),  
        W_2.reshape(-1,1).T)*a_2_z_2_grad).T.dot(x_mat).T  
    gradient = (J_W_1_grad, J_W_2_grad)  
  
    # return  
    return y_pred, gradient
```

```
def plot_loss_accuracy(loss_vals, accuracies):  
    fig = plt.figure(figsize=(16, 8))  
    fig.suptitle('Log Loss and Accuracy over iterations')  
  
    ax = fig.add_subplot(1, 2, 1)  
    ax.plot(loss_vals)  
    ax.grid(True)  
    ax.set(xlabel='iterations', title='Log Loss')  
  
    ax = fig.add_subplot(1, 2, 2)  
    ax.plot(accuracies)  
    ax.grid(True)  
    ax.set(xlabel='iterations', title='Accuracy');
```

```
np.random.seed(1241)  
  
W_1 = np.random.uniform(-1,10,size=(3,4))  
W_2 = np.random.uniform(-1,10,size=(4))  
num_iter = 300  
learning_rate = .001  
x_mat = x_mat_full  
  
loss_vals, accuracies = [], []  
for i in range(num_iter):  
    ### Do a forward computation, and get the gradient  
    y_pred, (J_W_1_grad, J_W_2_grad) = forward_pass(W_1, W_2)  
  
    ## Update the weight matrices  
    W_1 = W_1 - learning_rate*J_W_1_grad  
    W_2 = W_2 - learning_rate*J_W_2_grad  
  
    ### Compute the loss and accuracy  
    curr_loss = loss_fn(y,y_pred)  
    loss_vals.append(curr_loss)  
    acc = np.sum((y_pred>=.5) == y)/num_obs  
    accuracies.append(acc)  
  
    ## Print the loss and accuracy for every 200th iteration  
    if((i%200) == 0):  
        print('iteration {}, log loss is {:.4f}, accuracy is {}'.format(  
            i, curr_loss, acc  
        ))  
plot_loss_accuracy(loss_vals, accuracies)
```

iteration 0, log loss is 4.8821, accuracy is 0.26  
iteration 200, log loss is 0.5490, accuracy is 0.74

Log Loss and Accuracy over iterations



```
pred1 = (y_pred>=.5)
pred0 = (y_pred<=.5)

fig, ax = plt.subplots(figsize=(8, 8))
# true predictions
ax.plot(x_mat[pred1 & (y==1),0],x_mat[pred1 & (y==1),1], 'ro',
        label='true positives')
ax.plot(x_mat[pred0 & (y==0),0],x_mat[pred0 & (y==0),1], 'bx',
        label='true negatives')
# false predictions
ax.plot(x_mat[pred1 & (y==0),0],x_mat[pred1 & (y==0),1], 'yx',
        label='false positives', markersize=15)
ax.plot(x_mat[pred0 & (y==1),0],x_mat[pred0 & (y==1),1], 'yo',
        label='false negatives', markersize=15, alpha=.6)
ax.set(title='Truth vs Prediction')
ax.legend(bbox_to_anchor=(1, 0.8), fancybox=True, shadow=True,
          fontsize='x-large');
```

