Name: Matthew Seman

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import style
style.use('ggplot')
%matplotlib inline
import re
import pandas as pd
```

Numpy is library for scientific computing in Python. It has efficient implementation of n-dimensional array (tensor) manupulations, which is useful for machine learning applications.

In [2]:
```
import numpy as np
```

We can convert a list into numpy array (tensor)

In [3]:
```
b = [[1, 2, 4], [2, 6, 9]]
a = np.array(b)
a
```

Out[3]:
```
array([[1, 2, 4],
       [2, 6, 9]])
```

We can check the dimensions of the array

In [4]:
```
a.shape
```

Out[4]:
```
(2, 3)
```

We can apply simple arithmetic operation on all element of a tensor

In [5]:
```
a * 3
```

Out[5]:
```
array([[ 3,  6, 12],
       [ 6, 18, 27]])
```

You can transpose a tensor

In [6]:
```
print(a.T.shape)
a.T
```

Out[6]:
```
(3, 2)
array([[1, 2],
       [2, 6],
       [4, 9]])
```

You can apply aggregate functions on the whole tensor

In [7]:
```
np.sum(a)
```

Out[7]:
```
24
```

or on one dimension of it

In [8]:
```
np.sum(a, axis=0)
```

Out[8]:
```
array([ 3,  8, 13])
```

In [9]:
```
np.sum(a, axis=1)
```

Out[9]:
```
array([ 7, 17])
```

We can do element-wise arithmetic operation on two tensors (of the same size)

In [10]:
```
c1 = np.array([[1, 2, 4], [2, 6, 9]])
c2 = np.array([[2, 3, 5], [1, 2, 1]])
c1 * c2
```

Out[10]:
```
array([[ 2,  6, 20],
       [ 2, 12,  9]])
```

If you want to multiply all columns of a tensor by vector (for example if you want to multiply all data features by their lables) you need a

trick. This multiplication shows up in calculating the gradients.

In [11]:
```
a = np.array([[1, 2, 4], [2, 6, 9]])
b = np.array([1,-1])
print(a)
print(b)
```

```
[[1 2 4]
 [2 6 9]]
[ 1 -1]
```

Here we want to multiply the first row of a by 1 and the second row of a by -1. Simply multiplying a by b does not work because a and b do not have the same dimension

To do this multiplication we first have to assume b has one column and then repeat the column of b with the number of columns in a. We use tile function to do that

In [12]:
```
b_repeat = np.tile(b,  (a.shape[1],1)).T
print(b_repeat.shape)
b_repeat
```

Out[12]:
```
(2, 3)
array([[ 1,  1,  1],
       [-1, -1, -1]])
```

Now we can multiply each column of a by b:

In [13]:
```
a * b_repeat
```

Out[13]:
```
array([[ 1,  2,  4],
       [-2, -6, -9]])
```

You can create inital random vector using numpy (using N(0,1)):

In [14]:
```
mu = 0 #mean
sigma = 1 #standard deviation
r = np.random.normal(mu,sigma, 1000) #draws 1000 samples from a normal distribution
```

We can apply functions on tensors

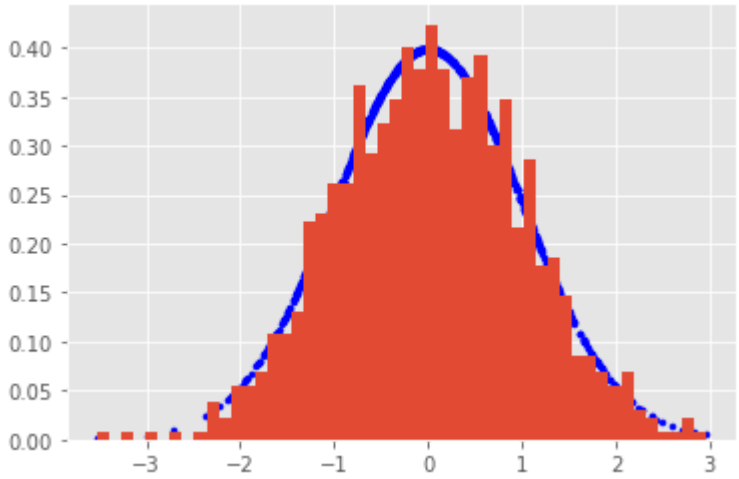
In [15]:
```
#implementation of Normal distribution
def normal(x, mu, sigma):
    return np.exp( -0.5 * ((x-mu)/sigma)**2)/np.sqrt(2.0*np.pi*sigma**2)

#probability of samples on the Normal distribution
probabilities = normal(r, mu, sigma)
```

Numpy has useful APIs for analysis. Here we plot the histogram of samples and also plot the probabilies to see if the samples follow the normal distribution.

In [16]:
```
counts, bins = np.histogram(r,50,density=True)
plt.hist(bins[:-1], bins, weights=counts)
plt.scatter(r, probabilities, c='b', marker='.')
```

Out[16]:
```
<matplotlib.collections.PathCollection at 0x2a8a9666370>
```



In [17]:
```
def read_data(filename):
    f = open(filename, 'r')
    p = re.compile(',')
    xdata = []
    ydata = []
    header = f.readline().strip()
    varnames = p.split(header)
```

```
    namehash = {}
    for l in f:
        li = p.split(l.strip())
        xdata.append([float(x) for x in li[:-1]])
        ydata.append(float(li[-1]))

    return np.array(xdata), np.array(ydata)
```

Assuming our data is x is available in numpy we use numpy to implement logistic regression

In [18]:
```
(xtrain_whole, ytrain_whole) = read_data('datasets/spambase-train.csv')
(xtest, ytest) = read_data('datasets/spambase-test.csv')
```

In [19]:
```
print("The shape of xtrain:", xtrain_whole.shape)
print("The shape of ytrain:", ytrain_whole.shape)
print("The shape of xtest:", xtest.shape)
print("The shape of ytest:", ytest.shape)
```

```
The shape of xtrain: (3601, 54)
The shape of ytrain: (3601,)
The shape of xtest: (1000, 54)
The shape of ytest: (1000,)
```

before training make we normalize the input data (features)

In [20]:
```
xmean = np.mean(xtrain_whole, axis=0)
xstd = np.std(xtrain_whole, axis=0)
xtrain_normal_whole = (xtrain_whole-xmean) / xstd
xtest_normal = (xtest-xmean) / xstd
```

We need to create a validation set. We create an array of indecies and permute it.

In [21]:
```
premute_indicies = np.random.permutation(np.arange(xtrain_whole.shape[0]))
```

We keep the first 2600 data points as the training data and rest as the validation data

In [22]:
```
xtrain_normal = xtrain_normal_whole[premute_indicies[:2600]]
ytrain = ytrain_whole[premute_indicies[:2600]]
```

```
xval_normal = xtrain_normal_whole[premute_indicies[2600:]]
yval = ytrain_whole[premute_indicies[2600:]]
```

Initiallizing the weights and bias with random values from N(0,1)

In [23]:
```
weights = np.random.normal(0, 1, xtrain_normal.shape[1]);
bias = np.random.normal(0,1,1)
```

In [24]:
```
#the sigmoid function
def sigmoid(v):
    #return np.exp(-np.logaddexp(0, -v)) #numerically stable implementation of sigmoid function
    return 1.0 / (1+np.exp(-v))
```

We can use dot-product from numpy to calculate the margin and pass it to the sigmoid function

In [25]:
```
#w: weight vector (numpy array of size n)
#b: numpy array of size 1
#returns p(y=1|x, w, b)
def prob(x, w, b):
    return sigmoid(np.dot(x,w) + b);
```

You can also calculate l_2 penalty using linalg library of numpy

In [26]:
```
np.linalg.norm(weights)
```

Out[26]:
```
8.790338846437223
```

$$\text{Cross Entropy Loss} = -\frac{1}{|D|}[\sum_{(y^i,\mathbf{x}^i)\in\mathcal{D}} y^i \log p(y=1|\mathbf{x}^i;\mathbf{w},b) + (1-y^i) \log (1 - p(y=1|\mathbf{x}^i;\mathbf{w},b))]+\frac{\lambda}{2} \|\mathbf{w}\|^2 $$

In [27]:
```
#w: weight vector (numpy array of size n)
#y_prob: p(y|x, w, b)
#y_true: class variable data
#lambda_: l2 penalty coefficient
#returns the cross entropy loss
def loss(w, y_prob, y_true, lambda_):
    loss = -(1/len(y_true)) * np.sum(y_true*np.log(y_prob) + (1-y_true)*np.log(y_prob)) + (lambda_/2)*(np.linalg.norm(w)
    return loss
```

In [28]:
```
#x: input variables (data of size m x n with m data point and n features)
#w: weight vector (numpy array of size n)
#y_prob: p(y|x, w, b)
#y_true: class variable data
#lambda_: l2 penalty coefficient
#returns tuple of gradient w.r.t w and w.r.t to bias

def grad_w_b(x, w, y_prob, y_true, lambda_):

    grad_w = np.matmul(np.transpose(y_prob-y_true), x) + lambda_ * w
    grad_b = np.mean(y_prob-y_true)

    return (grad_w,grad_b)
```

In [29]:
```
#lambda_ is the coeffienct of l2 norm penalty
#learning_rate is learning rate of gradient descent algorithm
#max_iter determines the maximum number of iterations if the gradients descent does not converge.
#continue the training while gradient > 0.1 or the number steps is less max_iter

#returns model as tuple of (weights,bias)

def fit(x, y_true, learning_rate, lambda_, max_iter, verbose=0):
    weights = np.random.normal(0, 1, x.shape[1]);
    bias = np.random.normal(0,1,1)

    i = 1
    loss_ = []

    #print("shape of w", weights.shape)
    #print("shape of x", x.shape)
    #print("shape of y", y_true.shape)

    #change the condition appropriately
    while i < max_iter:

        #Calculate loss
        y_prob = prob(x, weights, bias)

        loss_ = loss(weights, y_prob, y_true, lambda_)

        #Calculate Gradients
        (grad_w, grad_b) = grad_w_b(x, weights, y_prob, y_true, lambda_)
```

```
        #Update weights and bias
        weights = weights - learning_rate*grad_w
        bias = bias - learning_rate*grad_b

        if verbose: #verbose is used for debugging purposes
            #print iteration number, loss, l2 norm of gradients, l2 norm of weights
            print("Iteration Number: " + str(i) + " / Loss: " + str(loss_) +
                  " / Norm of grad_w, grad_b: " + str(np.linalg.norm(grad_w)) + " , " + str(np.linalg.norm(grad_b)) +
                  " / Norm of weights: " + str(np.linalg.norm(weights)))
            pass

        #End iterations if gradient is sufficiently small
        if np.linalg.norm(grad_w) < 0.1:
            break

        #increment while iteration
        i = i + 1

    return (weights, bias)
```

In [30]:

```
def accuracy(x, y_true, model):
    w, b = model
    return np.sum((prob(x, w, b)>0.5).astype(np.float64) == y_true)  / y_true.shape[0]
```

In [31]:

```
learning_rate = 0.001
lambda_ = 1.0

model = fit(xtrain_normal, ytrain, learning_rate, lambda_, 10000, verbose=1) #keep the verbose on here for your submissic
```

```
Iteration Number: 1 / Loss: 26.828504421539 / Norm of grad_w, grad_b: 2412.000756458321 , 0.07303917642375554 / Norm of w
eights: 6.578548044628662
Iteration Number: 2 / Loss: 24.733874464441648 / Norm of grad_w, grad_b: 776.7270387347985 , 0.08609004957603753 / Norm o
f weights: 6.352666294942999
Iteration Number: 3 / Loss: 23.436665174484546 / Norm of grad_w, grad_b: 564.9379722893176 , 0.08399822977758224 / Norm o
f weights: 6.179601685978154
Iteration Number: 4 / Loss: 22.466390140041987 / Norm of grad_w, grad_b: 434.7192290520903 , 0.07776788032482591 / Norm o
f weights: 6.0352977654402125
Iteration Number: 5 / Loss: 21.663422247930438 / Norm of grad_w, grad_b: 355.69750824040045 , 0.07086470897329152 / Norm
of weights: 5.908171111468392
Iteration Number: 6 / Loss: 20.949713033300956 / Norm of grad_w, grad_b: 303.19570663818695 , 0.06492974379368496 / Norm
of weights: 5.7924931448657695
```

```
Iteration Number: 7 / Loss: 20.292143901855976 / Norm of grad_w, grad_b: 266.68754932776426 , 0.06011323126170147 / Norm
of weights: 5.685299702871647
Iteration Number: 8 / Loss: 19.677307211560336 / Norm of grad_w, grad_b: 240.22186697252525 , 0.05617734333400217 / Norm
of weights: 5.584959983779066
Iteration Number: 9 / Loss: 19.09913475493964 / Norm of grad_w, grad_b: 219.87179409604144 , 0.05293241009749725 / Norm o
f weights: 5.490487336007138
Iteration Number: 10 / Loss: 18.554109301197958 / Norm of grad_w, grad_b: 203.49038660760507 , 0.050248210394520966 / Nor
m of weights: 5.401200674684745
Iteration Number: 11 / Loss: 18.03958460192147 / Norm of grad_w, grad_b: 189.85545391669163 , 0.04803471849192545 / Norm
of weights: 5.316612020913012
Iteration Number: 12 / Loss: 17.55343828788403 / Norm of grad_w, grad_b: 177.98831893016865 , 0.04623126662253028 / Norm
of weights: 5.236371902530075
Iteration Number: 13 / Loss: 17.093909113966326 / Norm of grad_w, grad_b: 167.10218371289358 , 0.04479440224183891 / Norm
of weights: 5.160198751245082
Iteration Number: 14 / Loss: 16.65931911616824 / Norm of grad_w, grad_b: 156.95164272434735 , 0.04366742513310355 / Norm
of weights: 5.087816095106119
Iteration Number: 15 / Loss: 16.248004846342823 / Norm of grad_w, grad_b: 147.82332502087777 , 0.04275741975049913 / Norm
of weights: 5.018950016808706
Iteration Number: 16 / Loss: 15.858578252528353 / Norm of grad_w, grad_b: 139.9098933465216 , 0.0419662267936748 / Norm o
f weights: 4.953361147673817
Iteration Number: 17 / Loss: 15.48998840421346 / Norm of grad_w, grad_b: 133.0629491397179 , 0.041230748845108404 / Norm
of weights: 4.8908523487875275
Iteration Number: 18 / Loss: 15.14130438059723 / Norm of grad_w, grad_b: 127.01652533434886 , 0.04052538303730341 / Norm
of weights: 4.831257888114071
Iteration Number: 19 / Loss: 14.811585852596115 / Norm of grad_w, grad_b: 121.54591905087293 , 0.03984534658441914 / Norm
of weights: 4.774432562695725
Iteration Number: 20 / Loss: 14.499875537405636 / Norm of grad_w, grad_b: 116.49413028979524 , 0.039193572252776615 / Nor
m of weights: 4.720244397768117
Iteration Number: 21 / Loss: 14.205220193410627 / Norm of grad_w, grad_b: 111.75665591957697 , 0.03857472171863855 / Norm
of weights: 4.668570165177358
Iteration Number: 22 / Loss: 13.926686386325809 / Norm of grad_w, grad_b: 107.26502517514777 , 0.037993128200775425 / Nor
m of weights: 4.6192927518399225
Iteration Number: 23 / Loss: 13.663368919814335 / Norm of grad_w, grad_b: 102.97624757812552 , 0.03745217368094143 / Norm
of weights: 4.572299677555269
Iteration Number: 24 / Loss: 13.414394544408049 / Norm of grad_w, grad_b: 98.86625102627235 , 0.03695409738173552 / Norm
of weights: 4.527482423042125
Iteration Number: 25 / Loss: 13.178923551951787 / Norm of grad_w, grad_b: 94.9244193903804 , 0.03650000622918082 / Norm o
f weights: 4.484736461118743
Iteration Number: 26 / Loss: 12.956151749512397 / Norm of grad_w, grad_b: 91.14796440703019 , 0.036090060324134254 / Norm
of weights: 4.443961853771735
Iteration Number: 27 / Loss: 12.745314328863545 / Norm of grad_w, grad_b: 87.53682367294876 , 0.03572374651424744 / Norm
of weights: 4.405064103358961
Iteration Number: 28 / Loss: 12.545691277938626 / Norm of grad_w, grad_b: 84.09034437366411 , 0.03540008928458255 / Norm
of weights: 4.367954861795738
Iteration Number: 29 / Loss: 12.35661247047767 / Norm of grad_w, grad_b: 80.80608626101599 , 0.03511770537995423 / Norm o
```

```
f weights: 4.332552210407817
Iteration Number: 30 / Loss: 12.177460433955316 / Norm of grad_w, grad_b: 77.68000548923985 , 0.0348747336359573 / Norm o
f weights: 4.298780439700969
Iteration Number: 31 / Loss: 12.007669834007185 / Norm of grad_w, grad_b: 74.70701157711645 , 0.03466875028302907 / Norm
of weights: 4.2665694463049535
Iteration Number: 32 / Loss: 11.846724003120006 / Norm of grad_w, grad_b: 71.88127425341261 , 0.03449677200221476 / Norm
of weights: 4.235853952394346
Iteration Number: 33 / Loss: 11.694149660372297 / Norm of grad_w, grad_b: 69.19617748337161 , 0.03435538875263718 / Norm
of weights: 4.206572742901689
Iteration Number: 34 / Loss: 11.549511141582549 / Norm of grad_w, grad_b: 66.64414985313752 , 0.03424100172468062 / Norm
of weights: 4.178668045737596
Iteration Number: 35 / Loss: 11.412405162838533 / Norm of grad_w, grad_b: 64.21665093613464 , 0.03415009742182502 / Norm
of weights: 4.1520850938597
Iteration Number: 36 / Loss: 11.28245663598198 / Norm of grad_w, grad_b: 61.904437817635284 , 0.034079480588813583 / Norm
of weights: 4.126771839677131
Iteration Number: 37 / Loss: 11.15931556464448 / Norm of grad_w, grad_b: 59.69804361160793 , 0.03402641437702333 / Norm o
f weights: 4.102678759262133
Iteration Number: 38 / Loss: 11.042654727716855 / Norm of grad_w, grad_b: 57.58830443596664 , 0.033988657125171705 / Norm
of weights: 4.07975868511692
Iteration Number: 39 / Loss: 10.932167761490721 / Norm of grad_w, grad_b: 55.5667915472726 , 0.03396441839727968 / Norm o
f weights: 4.057966628018851
Iteration Number: 40 / Loss: 10.827567335123707 / Norm of grad_w, grad_b: 53.62608045945334 , 0.03395227048484977 / Norm
of weights: 4.037259574336073
Iteration Number: 41 / Loss: 10.72858327187636 / Norm of grad_w, grad_b: 51.75985713216132 , 0.03395104799913622 / Norm o
f weights: 4.017596264353617
Iteration Number: 42 / Loss: 10.6349606090022 / Norm of grad_w, grad_b: 49.9628976065916 , 0.03395975678989458 / Norm of
weights: 3.9989369664516285
Iteration Number: 43 / Loss: 10.546457670821399 / Norm of grad_w, grad_b: 48.23096575820565 , 0.03397750208836011 / Norm
of weights: 3.9812432633747252
Iteration Number: 44 / Loss: 10.462844254300796 / Norm of grad_w, grad_b: 46.560667376831866 , 0.03400343794771363 / Norm
of weights: 3.964477863784449
Iteration Number: 45 / Loss: 10.383900014861666 / Norm of grad_w, grad_b: 44.94928820382026 , 0.03403673591452461 / Norm
of weights: 3.948604447629219
Iteration Number: 46 / Loss: 10.309413112338682 / Norm of grad_w, grad_b: 43.394634137514856 , 0.03407656926935581 / Norm
of weights: 3.933587549236717
Iteration Number: 47 / Loss: 10.239179146007283 / Norm of grad_w, grad_b: 41.89488498034768 , 0.03412210893864026 / Norm
of weights: 3.919392478132142
Iteration Number: 48 / Loss: 10.173000379679666 / Norm of grad_w, grad_b: 40.44846845672286 , 0.03417252751398161 / Norm
of weights: 3.905985274602083
Iteration Number: 49 / Loss: 10.110685235383762 / Norm of grad_w, grad_b: 39.05395796223875 , 0.034227008320671035 / Norm
of weights: 3.8933326949464018
Iteration Number: 50 / Loss: 10.052048017720798 / Norm of grad_w, grad_b: 37.70999498913484 , 0.034284757005226485 / Norm
of weights: 3.8814022201243743
Iteration Number: 51 / Loss: 9.996908820708644 / Norm of grad_w, grad_b: 36.415235130911626 , 0.03434501362966948 / Norm
of weights: 3.8701620810259496
```

```
Iteration Number: 52 / Loss: 9.945093564553968 / Norm of grad_w, grad_b: 35.16831497889809 , 0.03440706376829913 / Norm o
f weights: 3.8595812937812384
Iteration Number: 53 / Loss: 9.896434110818475 / Norm of grad_w, grad_b: 33.967836165351116 , 0.03447024759505741 / Norm
of weights: 3.849629699224549
Iteration Number: 54 / Loss: 9.850768409897455 / Norm of grad_w, grad_b: 32.812362334390734 , 0.03453396640592294 / Norm
of weights: 3.8402780016861953
Iteration Number: 55 / Loss: 9.807940643283422 / Norm of grad_w, grad_b: 31.70042489593058 , 0.03459768641358978 / Norm o
f weights: 3.831497803516881
Iteration Number: 56 / Loss: 9.767801333227835 / Norm of grad_w, grad_b: 30.630533912075883 , 0.03466093995738576 / Norm
of weights: 3.82326163298972
Iteration Number: 57 / Loss: 9.730207402686426 / Norm of grad_w, grad_b: 29.60119120611944 , 0.03472332447842549 / Norm o
f weights: 3.8155429643431007
Iteration Number: 58 / Loss: 9.695022177670918 / Norm of grad_w, grad_b: 28.610903599884445 , 0.034784499721341976 / Norm
of weights: 3.808316229639998
Iteration Number: 59 / Loss: 9.662115331576848 / Norm of grad_w, grad_b: 27.658194943554754 , 0.034844183654489234 / Norm
of weights: 3.8015568227906953
Iteration Number: 60 / Loss: 9.631362776394418 / Norm of grad_w, grad_b: 26.741616225035838 , 0.03490214757184356 / Norm
of weights: 3.795241096520791
Iteration Number: 61 / Loss: 9.60264650898102 / Norm of grad_w, grad_b: 25.859753505064734 , 0.03495821077474611 / Norm o
f weights: 3.7893463532967133
Iteration Number: 62 / Loss: 9.575854422073535 / Norm of grad_w, grad_b: 25.01123372745774 , 0.035012235150073075 / Norm
of weights: 3.7838508312924413
Iteration Number: 63 / Loss: 9.55088008986424 / Norm of grad_w, grad_b: 24.19472862848879 , 0.035064119878314884 / Norm o
f weights: 3.7787336864420014
Iteration Number: 64 / Loss: 9.527622537195484 / Norm of grad_w, grad_b: 23.4089570493279 , 0.03511379642992779 / Norm of
weights: 3.7739749715162176
Iteration Number: 65 / Loss: 9.505986000138634 / Norm of grad_w, grad_b: 22.652685972550245 , 0.03516122394600184 / Norm
of weights: 3.7695556130238286
Iteration Number: 66 / Loss: 9.485879684221393 / Norm of grad_w, grad_b: 21.924730583626083 , 0.035206385051125055 / Norm
of weights: 3.7654573865915464
Iteration Number: 67 / Loss: 9.467217525073211 / Norm of grad_w, grad_b: 21.223953619542627 , 0.03524928211145902 / Norm
of weights: 3.7616628913408907
Iteration Number: 68 / Loss: 9.44991795490566 / Norm of grad_w, grad_b: 20.549264221452553 , 0.03528993392748172 / Norm o
f weights: 3.758155523660357
Iteration Number: 69 / Loss: 9.433903677101739 / Norm of grad_w, grad_b: 19.8996164636955 , 0.03532837283621764 / Norm of
weights: 3.754919450673138
Iteration Number: 70 / Loss: 9.419101450275766 / Norm of grad_w, grad_b: 19.27400769146235 , 0.03536464218976263 / Norm o
f weights: 3.7519395836230554
Iteration Number: 71 / Loss: 9.405441882473474 / Norm of grad_w, grad_b: 18.67147676530731 , 0.03539879417353776 / Norm o
f weights: 3.74920155134251
Iteration Number: 72 / Loss: 9.392859235683618 / Norm of grad_w, grad_b: 18.09110228290377 , 0.03543088792739509 / Norm o
f weights: 3.7466916739230864
Iteration Number: 73 / Loss: 9.38129124049364 / Norm of grad_w, grad_b: 17.532000826443433 , 0.035460987934264915 / Norm
of weights: 3.7443969366786938
Iteration Number: 74 / Loss: 9.370678920508066 / Norm of grad_w, grad_b: 16.99332526714548 , 0.035489162643638904 / Norm
```

```
of weights: 3.7423049644698234
Iteration Number: 75 / Loss: 9.360966426028153 / Norm of grad_w, grad_b: 16.474263145642045 , 0.03551548330025627 / Norm
of weights: 3.740403996442932
Iteration Number: 76 / Loss: 9.352100876437495 / Norm of grad_w, grad_b: 15.9740351377392 , 0.035540022951541926 / Norm o
f weights: 3.7386828612291616
Iteration Number: 77 / Loss: 9.344032210729903 / Norm of grad_w, grad_b: 15.491893608504622 , 0.03556285561041549 / Norm
of weights: 3.7371309526398333
Iteration Number: 78 / Loss: 9.33671304563581 / Norm of grad_w, grad_b: 15.027121253194698 , 0.03558405555293527 / Norm o
f weights: 3.735738205891318
Iteration Number: 79 / Loss: 9.330098540840098 / Norm of grad_w, grad_b: 14.579029820700516 , 0.0356036967328087 / Norm o
f weights: 3.734495074388117
Iteration Number: 80 / Loss: 9.324146270828576 / Norm of grad_w, grad_b: 14.146958913549684 , 0.03562185229707904 / Norm
of weights: 3.7333925070897123
Iteration Number: 81 / Loss: 9.31881610294663 / Norm of grad_w, grad_b: 13.730274857726476 , 0.035638594189298586 / Norm
of weights: 3.732421926483744
Iteration Number: 82 / Loss: 9.314070081298555 / Norm of grad_w, grad_b: 13.328369635402085 , 0.03565399282824476 / Norm
of weights: 3.73157520718499
Iteration Number: 83 / Loss: 9.309872316157042 / Norm of grad_w, grad_b: 12.940659873906206 , 0.03566811685175603 / Norm
of weights: 3.7308446551765577
Iteration Number: 84 / Loss: 9.306188878588689 / Norm of grad_w, grad_b: 12.566585884766074 , 0.03568103291658406 / Norm
of weights: 3.7302229877064916
Iteration Number: 85 / Loss: 9.302987700032595 / Norm of grad_w, grad_b: 12.205610747276845 , 0.03569280554631024 / Norm
of weights: 3.729703313849783
Iteration Number: 86 / Loss: 9.300238476595482 / Norm of grad_w, grad_b: 11.857219431768211 , 0.03570349702037559 / Norm
of weights: 3.7292791157425316
Iteration Number: 87 / Loss: 9.297912577848496 / Norm of grad_w, grad_b: 11.520917958436991 , 0.03571316729814826 / Norm
of weights: 3.7289442304918277
Iteration Number: 88 / Loss: 9.295982959928752 / Norm of grad_w, grad_b: 11.196232588286916 , 0.03572187397271997 / Norm
of weights: 3.728692832761847
Iteration Number: 89 / Loss: 9.294424082763017 / Norm of grad_w, grad_b: 10.8827090433314 , 0.03572967224979341 / Norm of
weights: 3.728519418033727
Iteration Number: 90 / Loss: 9.293211831242623 / Norm of grad_w, grad_b: 10.579911753758514 , 0.035736614947614535 / Norm
of weights: 3.72841878653408
Iteration Number: 91 / Loss: 9.2923234401881 / Norm of grad_w, grad_b: 10.287423130225836 , 0.035742752514423486 / Norm o
f weights: 3.7283860278244854
Iteration Number: 92 / Loss: 9.291737422949607 / Norm of grad_w, grad_b: 10.00484285984571 , 0.035748133060355076 / Norm
of weights: 3.7284165060420387
Iteration Number: 93 / Loss: 9.291433503495462 / Norm of grad_w, grad_b: 9.731787224742638 , 0.035752802401126216 / Norm
of weights: 3.7285058457790505
Iteration Number: 94 / Loss: 9.29139255184645 / Norm of grad_w, grad_b: 9.467888442321156 , 0.035756804111201725 / Norm o
f weights: 3.7286499185882156
Iteration Number: 95 / Loss: 9.291596522717963 / Norm of grad_w, grad_b: 9.212794026579967 , 0.035760179584445426 / Norm
of weights: 3.7288448300980983
Iteration Number: 96 / Loss: 9.292028397236171 / Norm of grad_w, grad_b: 8.96616616995751 , 0.03576296810054177 / Norm of
weights: 3.7290869077225164
```

```
Iteration Number: 97 / Loss: 9.29267212759798 / Norm of grad_w, grad_b: 8.727681145299497 , 0.03576520689571433 / Norm of
weights: 3.7293726889464116
Iteration Number: 98 / Loss: 9.293512584548221 / Norm of grad_w, grad_b: 8.497028727610768 , 0.035766931236485576 / Norm
of weights: 3.729698910169983
Iteration Number: 99 / Loss: 9.294535507550727 / Norm of grad_w, grad_b: 8.273911635297457 , 0.03576817449540788 / Norm o
f weights: 3.73006249609227
Iteration Number: 100 / Loss: 9.295727457533372 / Norm of grad_w, grad_b: 8.058044990626806 , 0.035768968227863525 / Norm
of weights: 3.7304605496149845
Iteration Number: 101 / Loss: 9.297075772090642 / Norm of grad_w, grad_b: 7.849155799137323 , 0.03576934224917366 / Norm
of weights: 3.73089034224711
Iteration Number: 102 / Loss: 9.298568523030466 / Norm of grad_w, grad_b: 7.646982447725145 , 0.03576932471138504 / Norm
of weights: 3.7313493049907382
Iteration Number: 103 / Loss: 9.300194476155786 / Norm of grad_w, grad_b: 7.451274221116014 , 0.03576894217921044 / Norm
of weights: 3.7318350196886154
Iteration Number: 104 / Loss: 9.3019430531746 / Norm of grad_w, grad_b: 7.261790836413044 , 0.03576821970469713 / Norm of
weights: 3.7323452108140356
Iteration Number: 105 / Loss: 9.303804295635967 / Norm of grad_w, grad_b: 7.078301995385077 , 0.03576718090027745 / Norm
of weights: 3.732877737683964
Iteration Number: 106 / Loss: 9.305768830792914 / Norm of grad_w, grad_b: 6.900586954137346 , 0.035765848009930136 / Norm
of weights: 3.7334305870766
Iteration Number: 107 / Loss: 9.307827839296918 / Norm of grad_w, grad_b: 6.728434109780647 , 0.03576424197823867 / Norm
of weights: 3.7340018662349967
Iteration Number: 108 / Loss: 9.309973024632173 / Norm of grad_w, grad_b: 6.5616406036935295 , 0.035762382517189585 / Nor
m of weights: 3.734589796238788
Iteration Number: 109 / Loss: 9.312196584201528 / Norm of grad_w, grad_b: 6.4000119409507805 , 0.035760288170594914 / Nor
m of weights: 3.735192705726605
Iteration Number: 110 / Loss: 9.31449118197959 / Norm of grad_w, grad_b: 6.243361625474794 , 0.03575797637606302 / Norm o
f weights: 3.735809024952274
Iteration Number: 111 / Loss: 9.316849922652013 / Norm of grad_w, grad_b: 6.091510810451197 , 0.03575546352447439 / Norm
of weights: 3.7364372801584618
Iteration Number: 112 / Loss: 9.319266327163504 / Norm of grad_w, grad_b: 5.944287963538315 , 0.035752765016943695 / Norm
of weights: 3.737076088252006
Iteration Number: 113 / Loss: 9.32173430960061 / Norm of grad_w, grad_b: 5.801528546393379 , 0.035749895319275055 / Norm
of weights: 3.737724151765781
Iteration Number: 114 / Loss: 9.32424815533862 / Norm of grad_w, grad_b: 5.663074708030746 , 0.035746868013932925 / Norm
of weights: 3.738380254092517
Iteration Number: 115 / Loss: 9.326802500385293 / Norm of grad_w, grad_b: 5.528774991527482 , 0.035743695849568144 / Norm
of weights: 3.739043254976617
Iteration Number: 116 / Loss: 9.329392311857273 / Norm of grad_w, grad_b: 5.398484053589975 , 0.03574039078814981 / Norm
of weights: 3.739712086250601
Iteration Number: 117 / Loss: 9.332012869528263 / Norm of grad_w, grad_b: 5.272062396499065 , 0.03573696404976394 / Norm
of weights: 3.7403857478033853
Iteration Number: 118 / Loss: 9.334659748390862 / Norm of grad_w, grad_b: 5.149376111955702 , 0.0357334261551465 / Norm o
f weights: 3.741063303768212
Iteration Number: 119 / Loss: 9.337328802177103 / Norm of grad_w, grad_b: 5.030296636355533 , 0.0357297869660255 / Norm o
```

```
f weights: 3.7417438789185757
Iteration Number: 120 / Loss: 9.340016147785327 / Norm of grad_w, grad_b: 4.914700517029872 , 0.035726055723348714 / Norm
of weights: 3.7424266552610725
Iteration Number: 121 / Loss: 9.342718150563824 / Norm of grad_w, grad_b: 4.802469188999325 , 0.035722241083478705 / Norm
of weights: 3.743110868814626
Iteration Number: 122 / Loss: 9.345431410404242 / Norm of grad_w, grad_b: 4.69348876179846 , 0.03571835115243682 / Norm o
f weights: 3.7437958065660513
Iteration Number: 123 / Loss: 9.348152748600196 / Norm of grad_w, grad_b: 4.587649815939947 , 0.03571439351827903 / Norm
of weights: 3.744480803592443
Iteration Number: 124 / Loss: 9.350879195428924 / Norm of grad_w, grad_b: 4.4848472086013995 , 0.035710375281686274 / Nor
m of weights: 3.74516524034134
Iteration Number: 125 / Loss: 9.353607978416058 / Norm of grad_w, grad_b: 4.38497988812934 , 0.035706303084851565 / Norm
of weights: 3.745848540060074
Iteration Number: 126 / Loss: 9.3563365112457 / Norm of grad_w, grad_b: 4.287950716969085 , 0.035702183138744674 / Norm o
f weights: 3.7465301663661976
Iteration Number: 127 / Loss: 9.35906238328018 / Norm of grad_w, grad_b: 4.193666302643255 , 0.035698021248832805 / Norm
of weights: 3.747209620951244
Iteration Number: 128 / Loss: 9.36178334965556 / Norm of grad_w, grad_b: 4.102036836415257 , 0.0356938228393353 / Norm of
weights: 3.747886441410556
Iteration Number: 129 / Loss: 9.364497321921146 / Norm of grad_w, grad_b: 4.012975939287816 , 0.03568959297608636 / Norm
of weights: 3.748560199192246
Iteration Number: 130 / Loss: 9.367202359192703 / Norm of grad_w, grad_b: 3.9264005150008137 , 0.03568533638807858 / Norm
of weights: 3.749230497658755
Iteration Number: 131 / Loss: 9.369896659790927 / Norm of grad_w, grad_b: 3.8422306097067875 , 0.03568105748775683 / Norm
of weights: 3.749896970254828
Iteration Number: 132 / Loss: 9.37257855333831 / Norm of grad_w, grad_b: 3.760389278014397 , 0.03567676039012984 / Norm o
f weights: 3.750559278776041
Iteration Number: 133 / Loss: 9.37524649328887 / Norm of grad_w, grad_b: 3.680802455105693 , 0.035672448930763316 / Norm
of weights: 3.7512171117323554
Iteration Number: 134 / Loss: 9.377899049866885 / Norm of grad_w, grad_b: 3.603398834643712 , 0.03566812668271667 / Norm
of weights: 3.751870182801469
Iteration Number: 135 / Loss: 9.380534903391862 / Norm of grad_w, grad_b: 3.5281097522011104 , 0.035663796972481905 / Nor
m of weights: 3.7525182293670127
Iteration Number: 136 / Loss: 9.383152837968439 / Norm of grad_w, grad_b: 3.4548690739512606 , 0.03565946289498055 / Norm
of weights: 3.7531610111369322
Iteration Number: 137 / Loss: 9.385751735520998 / Norm of grad_w, grad_b: 3.3836130903770334 , 0.03565512732767271 / Norm
of weights: 3.753798308837633
Iteration Number: 138 / Loss: 9.388330570153954 / Norm of grad_w, grad_b: 3.3142804147612095 , 0.035650792943827926 / Nor
m of weights: 3.7544299229797264
Iteration Number: 139 / Loss: 9.390888402819732 / Norm of grad_w, grad_b: 3.246811886236145 , 0.03564646222500714 / Norm
of weights: 3.7550556726914364
Iteration Number: 140 / Loss: 9.393424376277494 / Norm of grad_w, grad_b: 3.181150477178431 , 0.035642137472800414 / Norm
of weights: 3.7556753946159493
Iteration Number: 141 / Loss: 9.395937710326574 / Norm of grad_w, grad_b: 3.1172412047464433 , 0.03563782081986485 / Norm
of weights: 3.7562889418691943
```

```
Iteration Number: 142 / Loss: 9.398427697299526 / Norm of grad_w, grad_b: 3.055031046366701 , 0.03563351424030235 / Norm
of weights: 3.756896183054742
Iteration Number: 143 / Loss: 9.400893697800552 / Norm of grad_w, grad_b: 2.994468858985202 , 0.0356292195594173 / Norm o
f weights: 3.7574970013326823
Iteration Number: 144 / Loss: 9.403335136675802 / Norm of grad_w, grad_b: 2.9355053019086834 , 0.035624938462889666 / Nor
m of weights: 3.758091293539536
Iteration Number: 145 / Loss: 9.405751499202898 / Norm of grad_w, grad_b: 2.8780927630685396 , 0.035620672505398895 / Nor
m of weights: 3.758678969356394
Iteration Number: 146 / Loss: 9.408142327487669 / Norm of grad_w, grad_b: 2.822185288548986 , 0.035616423118730846 / Norm
of weights: 3.759259950522662
Iteration Number: 147 / Loss: 9.410507217056786 / Norm of grad_w, grad_b: 2.767738515228458 , 0.03561219161939854 / Norm
of weights: 3.7598341700929097
Iteration Number: 148 / Loss: 9.41284581363564 / Norm of grad_w, grad_b: 2.7147096063902447 , 0.035607979215806024 / Norm
of weights: 3.7604015717344828
Iteration Number: 149 / Loss: 9.415157810101356 / Norm of grad_w, grad_b: 2.6630571901659756 , 0.03560378701498212 / Norm
of weights: 3.76096210906365
Iteration Number: 150 / Loss: 9.417442943601449 / Norm of grad_w, grad_b: 2.6127413006814355 , 0.03559961602891047 / Norm
of weights: 3.761515745018192
Iteration Number: 151 / Loss: 9.419700992829116 / Norm of grad_w, grad_b: 2.563723321781284 , 0.03559546718047974 / Norm
of weights: 3.7620624512644603
Iteration Number: 152 / Loss: 9.421931775446721 / Norm of grad_w, grad_b: 2.5159659332148463 , 0.03559134130907673 / Norm
of weights: 3.7626022076370145
Iteration Number: 153 / Loss: 9.424135145649412 / Norm of grad_w, grad_b: 2.469433059170286 , 0.03558723917584402 / Norm
of weights: 3.7631350016090996
Iteration Number: 154 / Loss: 9.426310991861353 / Norm of grad_w, grad_b: 2.424089819051587 , 0.035583161468622275 / Norm
of weights: 3.763660827792274
Iteration Number: 155 / Loss: 9.42845923455738 / Norm of grad_w, grad_b: 2.379902480395576 , 0.03557910880659637 / Norm o
f weights: 3.7641796874636153
Iteration Number: 156 / Loss: 9.430579824203377 / Norm of grad_w, grad_b: 2.3368384138333997 , 0.03557508174466251 / Norm
of weights: 3.7646915881190286
Iteration Number: 157 / Loss: 9.432672739308975 / Norm of grad_w, grad_b: 2.2948660500035807 , 0.035571080777534174 / Nor
m of weights: 3.765196543051232
Iteration Number: 158 / Loss: 9.434737984586535 / Norm of grad_w, grad_b: 2.253954838329611 , 0.035567106343601676 / Norm
of weights: 3.7656945709510987
Iteration Number: 159 / Loss: 9.436775589210756 / Norm of grad_w, grad_b: 2.214075207578422 , 0.035563158828560724 / Norm
of weights: 3.766185695531105
Iteration Number: 160 / Loss: 9.438785605173507 / Norm of grad_w, grad_b: 2.1751985281201502 , 0.035559238568823585 / Nor
m of weights: 3.7666699451696823
Iteration Number: 161 / Loss: 9.440768105728809 / Norm of grad_w, grad_b: 2.137297075813792 , 0.035555345854726035 / Norm
of weights: 3.767147352575367
Iteration Number: 162 / Loss: 9.44272318392315 / Norm of grad_w, grad_b: 2.100343997446168 , 0.0355514809335426 / Norm of
weights: 3.7676179544696726
Iteration Number: 163 / Loss: 9.444650951206597 / Norm of grad_w, grad_b: 2.064313277655879 , 0.03554764401232105 / Norm
of weights: 3.7680817912876923
Iteration Number: 164 / Loss: 9.446551536120431 / Norm of grad_w, grad_b: 2.0291797072763296 , 0.03554383526054754 / Norm
```

```
of weights: 3.7685389068954773
Iteration Number: 165 / Loss: 9.448425083057174 / Norm of grad_w, grad_b: 1.9949188530355781 , 0.03554005481265235 / Norm
of weights: 3.7689893483233
Iteration Number: 166 / Loss: 9.450271751089282 / Norm of grad_w, grad_b: 1.9615070285531013 , 0.03553630277036556 / Norm
of weights: 3.7694331655139433
Iteration Number: 167 / Loss: 9.452091712862739 / Norm of grad_w, grad_b: 1.928921266577034 , 0.03553257920493215 / Norm
of weights: 3.7698704110852304
Iteration Number: 168 / Loss: 9.453885153552227 / Norm of grad_w, grad_b: 1.8971392924071102 , 0.03552888415919427 / Norm
of weights: 3.770301140106017
Iteration Number: 169 / Loss: 9.455652269874536 / Norm of grad_w, grad_b: 1.8661394984518724 , 0.035525217649549266 / Nor
m of weights: 3.770725409884946
Iteration Number: 170 / Loss: 9.457393269157198 / Norm of grad_w, grad_b: 1.8359009198704452 , 0.03552157966779038 / Norm
of weights: 3.7711432797712745
Iteration Number: 171 / Loss: 9.459108368459395 / Norm of grad_w, grad_b: 1.8064032112517243 , 0.03551797018283743 / Norm
of weights: 3.771554810967135
Iteration Number: 172 / Loss: 9.460797793742415 / Norm of grad_w, grad_b: 1.7776266242858867 , 0.0355143891423637 / Norm
of weights: 3.771960066350621
Iteration Number: 173 / Loss: 9.462461779087004 / Norm of grad_w, grad_b: 1.7495519863850961 , 0.035510836474325604 / Nor
m of weights: 3.7723591103091234
Iteration Number: 174 / Loss: 9.464100565955192 / Norm of grad_w, grad_b: 1.7221606802117528 , 0.035507312088400494 / Nor
m of weights: 3.772752008582371
Iteration Number: 175 / Loss: 9.465714402494227 / Norm of grad_w, grad_b: 1.6954346240755813 , 0.035503815877338404 / Nor
m of weights: 3.7731388281146607
Iteration Number: 176 / Loss: 9.467303542880408 / Norm of grad_w, grad_b: 1.6693562531611452 , 0.03550034771823243 / Norm
of weights: 3.773519636915792
Iteration Number: 177 / Loss: 9.468868246700739 / Norm of grad_w, grad_b: 1.643908501550134 , 0.03549690747371303 / Norm
of weights: 3.773894503930234
Iteration Number: 178 / Loss: 9.470408778370354 / Norm of grad_w, grad_b: 1.6190747850036245 , 0.03549349499307028 / Norm
of weights: 3.774263498914102
Iteration Number: 179 / Loss: 9.471925406583942 / Norm of grad_w, grad_b: 1.5948389844712536 , 0.03549011011330839 / Norm
of weights: 3.7746266923195155
Iteration Number: 180 / Loss: 9.473418403799274 / Norm of grad_w, grad_b: 1.5711854302956925 , 0.03548675266013716 / Norm
of weights: 3.774984155185953
Iteration Number: 181 / Loss: 9.474888045751221 / Norm of grad_w, grad_b: 1.5480988870820807 , 0.035483422448902895 / Nor
m of weights: 3.7753359590382276
Iteration Number: 182 / Loss: 9.476334610994627 / Norm of grad_w, grad_b: 1.5255645392030714 , 0.035480119285463116 / Nor
m of weights: 3.7756821757907395
Iteration Number: 183 / Loss: 9.47775838047454 / Norm of grad_w, grad_b: 1.5035679769120414 , 0.035476842967008364 / Norm
of weights: 3.7760228776576548
Iteration Number: 184 / Loss: 9.479159637122347 / Norm of grad_w, grad_b: 1.4820951830371432 , 0.03547359328283388 / Norm
of weights: 3.7763581370687134
Iteration Number: 185 / Loss: 9.480538665476447 / Norm of grad_w, grad_b: 1.4611325202310539 , 0.03547037001506423 / Norm
of weights: 3.776688026590356
Iteration Number: 186 / Loss: 9.481895751326212 / Norm of grad_w, grad_b: 1.440666718751527 , 0.035467172939333526 / Norm
of weights: 3.7770126188518858
```

```
Iteration Number: 187 / Loss: 9.483231181377949 / Norm of grad_w, grad_b: 1.4206848647492427 , 0.035464001825424155 / Nor
m of weights: 3.777331986476401
Iteration Number: 188 / Loss: 9.48454524294176 / Norm of grad_w, grad_b: 1.4011743890401909 , 0.03546085643786578 / Norm
of weights: 3.7776462020162365
Iteration Number: 189 / Loss: 9.485838223638133 / Norm of grad_w, grad_b: 1.382123056341035 , 0.035457736536497704 / Norm
of weights: 3.7779553378926853
Iteration Number: 190 / Loss: 9.487110411123329 / Norm of grad_w, grad_b: 1.3635189549462665 , 0.03545464187699628 / Norm
of weights: 3.7782594663397493
Iteration Number: 191 / Loss: 9.488362092832421 / Norm of grad_w, grad_b: 1.345350486827541 , 0.03545157221136903 / Norm
of weights: 3.7785586593517273
Iteration Number: 192 / Loss: 9.489593555739207 / Norm of grad_w, grad_b: 1.3276063581355995 , 0.03544852728841852 / Norm
of weights: 3.778852988634413
Iteration Number: 193 / Loss: 9.490805086131974 / Norm of grad_w, grad_b: 1.3102755700864055 , 0.035445506854176426 / Nor
m of weights: 3.779142525559723
Iteration Number: 194 / Loss: 9.491996969404365 / Norm of grad_w, grad_b: 1.2933474102139562 , 0.03544251065231041 / Norm
of weights: 3.7794273411235584
Iteration Number: 195 / Loss: 9.49316948986047 / Norm of grad_w, grad_b: 1.276811443972526 , 0.03543953842450508 / Norm o
f weights: 3.779707505906731
Iteration Number: 196 / Loss: 9.49432293053345 / Norm of grad_w, grad_b: 1.2606575066722379 , 0.035436589910818564 / Norm
of weights: 3.7799830900387867
Iteration Number: 197 / Loss: 9.49545757301691 / Norm of grad_w, grad_b: 1.2448756957321268 , 0.03543366485001587 / Norm
of weights: 3.7802541631645674
Iteration Number: 198 / Loss: 9.496573697308381 / Norm of grad_w, grad_b: 1.2294563632357796 , 0.03543076297988078 / Norm
of weights: 3.780520794413358
Iteration Number: 199 / Loss: 9.49767158166425 / Norm of grad_w, grad_b: 1.2143901087752371 , 0.035427884037507386 / Norm
of weights: 3.7807830523704857
Iteration Number: 200 / Loss: 9.498751502465499 / Norm of grad_w, grad_b: 1.1996677725693627 , 0.035425027759572 / Norm o
f weights: 3.7810410050512235
Iteration Number: 201 / Loss: 9.499813734093683 / Norm of grad_w, grad_b: 1.1852804288430925 , 0.03542219388258736 / Norm
of weights: 3.7812947198768883
Iteration Number: 202 / Loss: 9.500858548816637 / Norm of grad_w, grad_b: 1.1712193794555716 , 0.035419382143139444 / Nor
m of weights: 3.7815442636529877
Iteration Number: 203 / Loss: 9.501886216683266 / Norm of grad_w, grad_b: 1.1574761477646243 , 0.035416592278108165 / Nor
m of weights: 3.7817897025493297
Iteration Number: 204 / Loss: 9.502897005427068 / Norm of grad_w, grad_b: 1.1440424727160303 , 0.0354138240248732 / Norm
of weights: 3.782031102081958
Iteration Number: 205 / Loss: 9.503891180377785 / Norm of grad_w, grad_b: 1.1309103031466181 , 0.035411077121504916 / Nor
m of weights: 3.7822685270968313
Iteration Number: 206 / Loss: 9.504869004380824 / Norm of grad_w, grad_b: 1.1180717922906829 , 0.035408351306942476 / Nor
m of weights: 3.7825020417551354
Iteration Number: 207 / Loss: 9.505830737723947 / Norm of grad_w, grad_b: 1.1055192924795194 , 0.03540564632115863 / Norm
of weights: 3.782731709520139
Iteration Number: 208 / Loss: 9.506776638070887 / Norm of grad_w, grad_b: 1.093245350024371 , 0.035402961905312946 / Norm
of weights: 3.7829575931455093
Iteration Number: 209 / Loss: 9.50770696040147 / Norm of grad_w, grad_b: 1.0812427002739458 , 0.03540029780189342 / Norm
```

```
of weights: 3.7831797546649915
Iteration Number: 210 / Loss: 9.508621956957857 / Norm of grad_w, grad_b: 1.0695042628373843 , 0.035397653754847845 / Nor
m of weights: 3.78339825538339
Iteration Number: 211 / Loss: 9.509521877196635 / Norm of grad_w, grad_b: 1.0580231369646527 , 0.03539502950970466 / Norm
of weights: 3.7836131558687582
Iteration Number: 212 / Loss: 9.51040696774632 / Norm of grad_w, grad_b: 1.0467925970763674 , 0.03539242481368474 / Norm
of weights: 3.7838245159457387
Iteration Number: 213 / Loss: 9.511277472370057 / Norm of grad_w, grad_b: 1.0358060884354936 , 0.03538983941580415 / Norm
of weights: 3.7840323946899783
Iteration Number: 214 / Loss: 9.51213363193316 / Norm of grad_w, grad_b: 1.0250572229540247 , 0.035387273066968504 / Norm
of weights: 3.784236850423558
Iteration Number: 215 / Loss: 9.512975684375226 / Norm of grad_w, grad_b: 1.0145397751276348 , 0.0353847255200594 / Norm
of weights: 3.7844379407113746
Iteration Number: 216 / Loss: 9.513803864686572 / Norm of grad_w, grad_b: 1.004247678092119 , 0.03538219653001349 / Norm
of weights: 3.784635722358413
Iteration Number: 217 / Loss: 9.514618404888708 / Norm of grad_w, grad_b: 0.9941750197955059 , 0.035379685853894664 / Nor
m of weights: 3.7848302514078647
Iteration Number: 218 / Loss: 9.515419534018637 / Norm of grad_w, grad_b: 0.9843160392802974 , 0.03537719325095922 / Norm
of weights: 3.7850215831400273
Iteration Number: 219 / Loss: 9.516207478116732 / Norm of grad_w, grad_b: 0.9746651230702906 , 0.03537471848271541 / Norm
of weights: 3.7852097720719446
Iteration Number: 220 / Loss: 9.51698246021798 / Norm of grad_w, grad_b: 0.9652168016569358 , 0.035372261312976744 / Norm
of weights: 3.7853948719577377
Iteration Number: 221 / Loss: 9.517744700346393 / Norm of grad_w, grad_b: 0.9559657460807583 , 0.03536982150790984 / Norm
of weights: 3.7855769357895794
Iteration Number: 222 / Loss: 9.518494415512363 / Norm of grad_w, grad_b: 0.9469067646028223 , 0.03536739883607739 / Norm
of weights: 3.785756015799282
Iteration Number: 223 / Loss: 9.519231819712843 / Norm of grad_w, grad_b: 0.9380347994628037 , 0.03536499306847594 / Norm
of weights: 3.7859321634604375
Iteration Number: 224 / Loss: 9.519957123934065 / Norm of grad_w, grad_b: 0.9293449237189831 , 0.0353626039785693 / Norm
of weights: 3.7861054294910956
Iteration Number: 225 / Loss: 9.520670536156757 / Norm of grad_w, grad_b: 0.9208323381668703 , 0.03536023134231766 / Norm
of weights: 3.7862758638569263
Iteration Number: 226 / Loss: 9.5213722613636 / Norm of grad_w, grad_b: 0.9124923683331065 , 0.03535787493820245 / Norm o
f weights: 3.7864435157748404
Iteration Number: 227 / Loss: 9.522062501548822 / Norm of grad_w, grad_b: 0.9043204615410756 , 0.03535553454724793 / Norm
of weights: 3.7866084337170407
Iteration Number: 228 / Loss: 9.52274145572979 / Norm of grad_w, grad_b: 0.8963121840456516 , 0.03535320995303867 / Norm
of weights: 3.7867706654154576
Iteration Number: 229 / Loss: 9.523409319960425 / Norm of grad_w, grad_b: 0.8884632182337756 , 0.03535090094173388 / Norm
of weights: 3.7869302578665622
Iteration Number: 230 / Loss: 9.524066287346374 / Norm of grad_w, grad_b: 0.8807693598887982 , 0.03534860730207902 / Norm
of weights: 3.7870872573365055
Iteration Number: 231 / Loss: 9.524712548061748 / Norm of grad_w, grad_b: 0.8732265155157115 , 0.0353463288254137 / Norm
of weights: 3.7872417093665827
```

```
Iteration Number: 232 / Loss: 9.525348289367377 / Norm of grad_w, grad_b: 0.86583069972534 , 0.03534406530567768 / Norm o
f weights: 3.7873936587789743
Iteration Number: 233 / Loss: 9.525973695630414 / Norm of grad_w, grad_b: 0.8585780326752356 , 0.03534181653941368 / Norm
of weights: 3.7875431496827603
Iteration Number: 234 / Loss: 9.526588948345237 / Norm of grad_w, grad_b: 0.8514647375653144 , 0.03533958232576808 / Norm
of weights: 3.7876902254801723
Iteration Number: 235 / Loss: 9.527194226155503 / Norm of grad_w, grad_b: 0.8444871381865481 , 0.03533736246648934 / Norm
of weights: 3.787834928873073
Iteration Number: 236 / Loss: 9.527789704877312 / Norm of grad_w, grad_b: 0.8376416565211023 , 0.035335156765924254 / Nor
m of weights: 3.7879773018696317
Iteration Number: 237 / Loss: 9.528375557523347 / Norm of grad_w, grad_b: 0.8309248103919462 , 0.03533296503101244 / Norm
of weights: 3.7881173857911903
Iteration Number: 238 / Loss: 9.52895195432792 / Norm of grad_w, grad_b: 0.8243332111612937 , 0.035330787071278576 / Norm
of weights: 3.7882552212792917
Iteration Number: 239 / Loss: 9.529519062772883 / Norm of grad_w, grad_b: 0.8178635614755794 , 0.03532862269882359 / Norm
of weights: 3.788390848302861
Iteration Number: 240 / Loss: 9.530077047614265 / Norm of grad_w, grad_b: 0.8115126530567626 , 0.03532647172831363 / Norm
of weights: 3.788524306165521
Iteration Number: 241 / Loss: 9.530626070909623 / Norm of grad_w, grad_b: 0.8052773645378176 , 0.035324333976967943 / Nor
m of weights: 3.788655633513025
Iteration Number: 242 / Loss: 9.531166292045999 / Norm of grad_w, grad_b: 0.7991546593422398 , 0.03532220926454535 / Norm
of weights: 3.7887848683407976
Iteration Number: 243 / Loss: 9.531697867768447 / Norm of grad_w, grad_b: 0.7931415836060319 , 0.03532009741332919 / Norm
of weights: 3.7889120480015688
Iteration Number: 244 / Loss: 9.532220952209059 / Norm of grad_w, grad_b: 0.7872352641412922 , 0.035317998248111604 / Nor
m of weights: 3.7890372092130877
Iteration Number: 245 / Loss: 9.532735696916442 / Norm of grad_w, grad_b: 0.781432906441061 , 0.03531591159617616 / Norm
of weights: 3.7891603880659024
Iteration Number: 246 / Loss: 9.533242250885563 / Norm of grad_w, grad_b: 0.7757317927238615 , 0.035313837287279826 / Nor
m of weights: 3.7892816200312036
Iteration Number: 247 / Loss: 9.533740760587973 / Norm of grad_w, grad_b: 0.7701292800178003 , 0.03531177515363408 / Norm
of weights: 3.7894009399687083
Iteration Number: 248 / Loss: 9.5342313700023 / Norm of grad_w, grad_b: 0.764622798283384 , 0.035309725029884824 / Norm o
f weights: 3.789518382134588
Iteration Number: 249 / Loss: 9.534714220644998 / Norm of grad_w, grad_b: 0.7592098485740637 , 0.03530768675309202 / Norm
of weights: 3.7896339801894197
Iteration Number: 250 / Loss: 9.535189451601312 / Norm of grad_w, grad_b: 0.7538880012345351 , 0.03530566016270805 / Norm
of weights: 3.78974776720616
Iteration Number: 251 / Loss: 9.5356571995564 / Norm of grad_w, grad_b: 0.748654894135548 , 0.0353036451005559 / Norm of
weights: 3.789859775678131
Iteration Number: 252 / Loss: 9.53611759882661 / Norm of grad_w, grad_b: 0.7435082309450665 , 0.035301641410806675 / Norm
of weights: 3.7899700375270085
Iteration Number: 253 / Loss: 9.536570781390827 / Norm of grad_w, grad_b: 0.7384457794350284 , 0.03529964893995617 / Norm
of weights: 3.79007858411081
Iteration Number: 254 / Loss: 9.537016876921903 / Norm of grad_w, grad_b: 0.7334653698236284 , 0.03529766753680167 / Norm
```

```
of weights: 3.790185446231868
Iteration Number: 255 / Loss: 9.53745601281812 / Norm of grad_w, grad_b: 0.7285648931516936 , 0.03529569705241772 / Norm
of weights: 3.7902906541447963
Iteration Number: 256 / Loss: 9.537888314234653 / Norm of grad_w, grad_b: 0.7237422996939005 , 0.03529373734013179 / Norm
of weights: 3.7903942375644246
Iteration Number: 257 / Loss: 9.53831390411501 / Norm of grad_w, grad_b: 0.7189955974033362 , 0.03529178825549959 / Norm
of weights: 3.7904962256737136
Iteration Number: 258 / Loss: 9.538732903222432 / Norm of grad_w, grad_b: 0.71432285038954 , 0.03528984965628005 / Norm o
f weights: 3.790596647131635
Iteration Number: 259 / Loss: 9.53914543017123 / Norm of grad_w, grad_b: 0.7097221774292358 , 0.03528792140240997 / Norm
of weights: 3.7906955300810097
Iteration Number: 260 / Loss: 9.53955160145798 / Norm of grad_w, grad_b: 0.7051917505096356 , 0.03528600335597867 / Norm
of weights: 3.7907929021563187
Iteration Number: 261 / Loss: 9.539951531492692 / Norm of grad_w, grad_b: 0.7007297934035903 , 0.0352840953812023 / Norm
of weights: 3.790888790491451
Iteration Number: 262 / Loss: 9.540345332629755 / Norm of grad_w, grad_b: 0.6963345802763338 , 0.03528219734439791 / Norm
of weights: 3.7909832217274153
Iteration Number: 263 / Loss: 9.540733115198787 / Norm of grad_w, grad_b: 0.6920044343232954 , 0.03528030911395767 / Norm
of weights: 3.7910762220199903
Iteration Number: 264 / Loss: 9.541114987535298 / Norm of grad_w, grad_b: 0.6877377264386363 , 0.035278430560322725 / Nor
m of weights: 3.7911678170473273
Iteration Number: 265 / Loss: 9.541491056011175 / Norm of grad_w, grad_b: 0.6835328739140343 , 0.035276561555957195 / Nor
m of weights: 3.791258032017486
Iteration Number: 266 / Loss: 9.541861425064951 / Norm of grad_w, grad_b: 0.6793883391672457 , 0.035274701975322065 / Nor
m of weights: 3.7913468916759117
Iteration Number: 267 / Loss: 9.542226197231892 / Norm of grad_w, grad_b: 0.6753026285001855 , 0.035272851694848825 / Nor
m of weights: 3.791434420312849
Iteration Number: 268 / Loss: 9.542585473173833 / Norm of grad_w, grad_b: 0.671274290885811 , 0.0352710105929138 / Norm o
f weights: 3.7915206417706866
Iteration Number: 269 / Loss: 9.542939351708787 / Norm of grad_w, grad_b: 0.6673019167838385 , 0.03526917854981154 / Norm
of weights: 3.791605579451234
Iteration Number: 270 / Loss: 9.543287929840318 / Norm of grad_w, grad_b: 0.6633841369842557 , 0.0352673554477291 / Norm
of weights: 3.7916892563229285
Iteration Number: 271 / Loss: 9.543631302786633 / Norm of grad_w, grad_b: 0.6595196214787151 , 0.03526554117071994 / Norm
of weights: 3.791771694927967
Iteration Number: 272 / Loss: 9.54396956400944 / Norm of grad_w, grad_b: 0.6557070783592965 , 0.03526373560467811 / Norm
of weights: 3.7918529173893654
Iteration Number: 273 / Loss: 9.544302805242506 / Norm of grad_w, grad_b: 0.6519452527437696 , 0.03526193863731231 / Norm
of weights: 3.7919329454179422
Iteration Number: 274 / Loss: 9.544631116519945 / Norm of grad_w, grad_b: 0.648232925727562 , 0.03526015015812035 / Norm
of weights: 3.7920118003192242
Iteration Number: 275 / Loss: 9.544954586204222 / Norm of grad_w, grad_b: 0.6445689133615472 , 0.03525837005836352 / Norm
of weights: 3.7920895030002746
Iteration Number: 276 / Loss: 9.545273301013854 / Norm of grad_w, grad_b: 0.640952065655366 , 0.035256598231041304 / Norm
of weights: 3.7921660739764427
```

```
Iteration Number: 277 / Loss: 9.545587346050826 / Norm of grad_w, grad_b: 0.6373812656057944 , 0.03525483457086587 / Norm
of weights: 3.7922415333780313
Iteration Number: 278 / Loss: 9.545896804827684 / Norm of grad_w, grad_b: 0.6338554282498673 , 0.03525307897423732 / Norm
of weights: 3.792315900956886
Iteration Number: 279 / Loss: 9.546201759294336 / Norm of grad_w, grad_b: 0.630373499742048 , 0.035251331339218456 / Norm
of weights: 3.7923891960929015
Iteration Number: 280 / Loss: 9.54650228986453 / Norm of grad_w, grad_b: 0.6269344564551846 , 0.03524959156551032 / Norm
of weights: 3.7924614378004464
Iteration Number: 281 / Loss: 9.54679847544203 / Norm of grad_w, grad_b: 0.6235373041050151 , 0.035247859554427455 / Norm
of weights: 3.792532644734705
Iteration Number: 282 / Loss: 9.547090393446451 / Norm of grad_w, grad_b: 0.6201810768970553 , 0.03524613520887378 / Norm
of weights: 3.7926028351979397
Iteration Number: 283 / Loss: 9.547378119838813 / Norm of grad_w, grad_b: 0.6168648366965173 , 0.035244418433318186 / Nor
m of weights: 3.7926720271456626
Iteration Number: 284 / Loss: 9.54766172914671 / Norm of grad_w, grad_b: 0.6135876722196721 , 0.03524270913377097 / Norm
of weights: 3.792740238192734
Iteration Number: 285 / Loss: 9.547941294489231 / Norm of grad_w, grad_b: 0.6103486982471631 , 0.035241007217759864 / Nor
m of weights: 3.79280748561937
Iteration Number: 286 / Loss: 9.548216887601484 / Norm of grad_w, grad_b: 0.6071470548582333 , 0.03523931259430683 / Norm
of weights: 3.7928737863770707
Iteration Number: 287 / Loss: 9.54848857885885 / Norm of grad_w, grad_b: 0.6039819066856491 , 0.03523762517390458 / Norm
of weights: 3.7929391570944646
Iteration Number: 288 / Loss: 9.54875643730087 / Norm of grad_w, grad_b: 0.6008524421909641 , 0.0352359448684938 / Norm o
f weights: 3.793003614083069
Iteration Number: 289 / Loss: 9.549020530654834 / Norm of grad_w, grad_b: 0.5977578729594661 , 0.035234271591440366 / Nor
m of weights: 3.7930671733429677
Iteration Number: 290 / Loss: 9.549280925359023 / Norm of grad_w, grad_b: 0.5946974330146705 , 0.035232605257512775 / Nor
m of weights: 3.7931298505684095
Iteration Number: 291 / Loss: 9.549537686585635 / Norm of grad_w, grad_b: 0.591670378151591 , 0.03523094578285992 / Norm
of weights: 3.7931916611533185
Iteration Number: 292 / Loss: 9.549790878263382 / Norm of grad_w, grad_b: 0.5886759852887115 , 0.03522929308498924 / Norm
of weights: 3.7932526201967276
Iteration Number: 293 / Loss: 9.55004056309976 / Norm of grad_w, grad_b: 0.5857135518380349 , 0.035227647082744735 / Norm
of weights: 3.7933127425081272
Iteration Number: 294 / Loss: 9.550286802602997 / Norm of grad_w, grad_b: 0.5827823950927679 , 0.035226007696285686 / Nor
m of weights: 3.7933720426127353
Iteration Number: 295 / Loss: 9.550529657103676 / Norm of grad_w, grad_b: 0.5798818516323371 , 0.03522437484706541 / Norm
of weights: 3.793430534756688
Iteration Number: 296 / Loss: 9.550769185776046 / Norm of grad_w, grad_b: 0.5770112767445291 , 0.035222748457810045 / Nor
m of weights: 3.793488232912144
Iteration Number: 297 / Loss: 9.551005446658996 / Norm of grad_w, grad_b: 0.574170043863572 , 0.03522112845249833 / Norm
of weights: 3.7935451507823164
Iteration Number: 298 / Loss: 9.55123849667672 / Norm of grad_w, grad_b: 0.5713575440250335 , 0.03521951475634068 / Norm
of weights: 3.7936013018064223
Iteration Number: 299 / Loss: 9.551468391659078 / Norm of grad_w, grad_b: 0.5685731853357918 , 0.03521790729575921 / Norm
```

```
of weights: 3.7936566991645537
Iteration Number: 300 / Loss: 9.551695186361615 / Norm of grad_w, grad_b: 0.5658163924597315 , 0.03521630599836787 / Norm
of weights: 3.793711355782475
Iteration Number: 301 / Loss: 9.551918934485292 / Norm of grad_w, grad_b: 0.5630866061182055 , 0.03521471079295273 / Norm
of weights: 3.7937652843363376
Iteration Number: 302 / Loss: 9.5521396886959 / Norm of grad_w, grad_b: 0.56038328260509 , 0.035213121609452606 / Norm of
weights: 3.793818497257322
Iteration Number: 303 / Loss: 9.552357500643172 / Norm of grad_w, grad_b: 0.557705893316128 , 0.035211538378939755 / Norm
of weights: 3.7938710067362034
Iteration Number: 304 / Loss: 9.552572420979565 / Norm of grad_w, grad_b: 0.5550539242919312 , 0.03520996103360141 / Norm
of weights: 3.7939228247278445
Iteration Number: 305 / Loss: 9.552784499378792 / Norm of grad_w, grad_b: 0.5524268757747068 , 0.03520838950672071 / Norm
of weights: 3.7939739629556093
Iteration Number: 306 / Loss: 9.552993784553989 / Norm of grad_w, grad_b: 0.5498242617778417 , 0.03520682373265869 / Norm
of weights: 3.7940244329157102
Iteration Number: 307 / Loss: 9.553200324275643 / Norm of grad_w, grad_b: 0.5472456096685143 , 0.035205263646836064 / Nor
m of weights: 3.794074245881479
Iteration Number: 308 / Loss: 9.553404165389209 / Norm of grad_w, grad_b: 0.5446904597626797 , 0.03520370918571533 / Norm
of weights: 3.7941234129075636
Iteration Number: 309 / Loss: 9.553605353832404 / Norm of grad_w, grad_b: 0.5421583649319661 , 0.035202160286783385 / Nor
m of weights: 3.7941719448340607
Iteration Number: 310 / Loss: 9.553803934652269 / Norm of grad_w, grad_b: 0.539648890222788 , 0.03520061688853399 / Norm
of weights: 3.794219852290575
Iteration Number: 311 / Loss: 9.553999952021911 / Norm of grad_w, grad_b: 0.5371616124864873 , 0.03519907893045091 / Norm
of weights: 3.7942671457002053
Iteration Number: 312 / Loss: 9.55419344925697 / Norm of grad_w, grad_b: 0.5346961200209446 , 0.035197546352990945 / Norm
of weights: 3.7943138352834684
Iteration Number: 313 / Loss: 9.554384468831802 / Norm of grad_w, grad_b: 0.5322520122228667 , 0.03519601909756744 / Norm
of weights: 3.794359931062156
Iteration Number: 314 / Loss: 9.554573052395426 / Norm of grad_w, grad_b: 0.5298288992507388 , 0.035194497106534044 / Nor
m of weights: 3.794405442863117
Iteration Number: 315 / Loss: 9.554759240787142 / Norm of grad_w, grad_b: 0.5274264016979491 , 0.03519298032316842 / Norm
of weights: 3.794450380321983
Iteration Number: 316 / Loss: 9.554943074051916 / Norm of grad_w, grad_b: 0.5250441502759381 , 0.0351914686916568 / Norm
of weights: 3.794494752886824
Iteration Number: 317 / Loss: 9.555124591455504 / Norm of grad_w, grad_b: 0.5226817855069857 , 0.03518996215707795 / Norm
of weights: 3.7945385698217406
Iteration Number: 318 / Loss: 9.555303831499298 / Norm of grad_w, grad_b: 0.5203389574265134 , 0.03518846066538829 / Norm
of weights: 3.7945818402103946
Iteration Number: 319 / Loss: 9.555480831934913 / Norm of grad_w, grad_b: 0.5180153252943642 , 0.035186964163406435 / Nor
m of weights: 3.7946245729594765
Iteration Number: 320 / Loss: 9.555655629778544 / Norm of grad_w, grad_b: 0.5157105573152133 , 0.03518547259879856 / Norm
of weights: 3.7946667768021105
Iteration Number: 321 / Loss: 9.555828261325043 / Norm of grad_w, grad_b: 0.5134243303673144 , 0.03518398592006365 / Norm
of weights: 3.7947084603012
```

```
Iteration Number: 322 / Loss: 9.55599876216176 / Norm of grad_w, grad_b: 0.5111563297398419 , 0.03518250407651923 / Norm
of weights: 3.794749631852713
Iteration Number: 323 / Loss: 9.556167167182153 / Norm of grad_w, grad_b: 0.5089062488783159 , 0.035181027018287 / Norm o
f weights: 3.7947902996889105
Iteration Number: 324 / Loss: 9.55633351059912 / Norm of grad_w, grad_b: 0.5066737891378281 , 0.035179554696279126 / Norm
of weights: 3.794830471881513
Iteration Number: 325 / Loss: 9.55649782595815 / Norm of grad_w, grad_b: 0.5044586595441012 , 0.035178087062184435 / Norm
of weights: 3.7948701563448126
Iteration Number: 326 / Loss: 9.556660146150177 / Norm of grad_w, grad_b: 0.502260576561743 , 0.035176624068454794 / Norm
of weights: 3.794909360838729
Iteration Number: 327 / Loss: 9.556820503424248 / Norm of grad_w, grad_b: 0.5000792638700731 , 0.03517516566829201 / Norm
of weights: 3.794948092971808
Iteration Number: 328 / Loss: 9.556978929399948 / Norm of grad_w, grad_b: 0.49791445214549734 , 0.03517371181563479 / Nor
m of weights: 3.794986360204168
Iteration Number: 329 / Loss: 9.557135455079605 / Norm of grad_w, grad_b: 0.4957658788510754 , 0.03517226246514572 / Norm
of weights: 3.79502416985039
Iteration Number: 330 / Loss: 9.557290110860256 / Norm of grad_w, grad_b: 0.49363328803249956 , 0.035170817572198705 / No
rm of weights: 3.7950615290823557
Iteration Number: 331 / Loss: 9.557442926545425 / Norm of grad_w, grad_b: 0.49151643012039825 , 0.035169377092866716 / No
rm of weights: 3.795098444932037
Iteration Number: 332 / Loss: 9.557593931356674 / Norm of grad_w, grad_b: 0.48941506173890664 , 0.03516794098390936 / Nor
m of weights: 3.7951349242942243
Iteration Number: 333 / Loss: 9.557743153944926 / Norm of grad_w, grad_b: 0.48732894552040645 , 0.035166509202760855 / No
rm of weights: 3.7951709739292183
Iteration Number: 334 / Loss: 9.557890622401608 / Norm of grad_w, grad_b: 0.48525784992574916 , 0.03516508170751845 / Nor
m of weights: 3.795206600465459
Iteration Number: 335 / Loss: 9.558036364269569 / Norm of grad_w, grad_b: 0.48320154907038787 , 0.035163658456930674 / No
rm of weights: 3.7952418104021146
Iteration Number: 336 / Loss: 9.558180406553817 / Norm of grad_w, grad_b: 0.48115982255590106 , 0.03516223941038571 / Nor
m of weights: 3.795276610111616
Iteration Number: 337 / Loss: 9.558322775732023 / Norm of grad_w, grad_b: 0.47913245530677034 , 0.03516082452790062 / Nor
m of weights: 3.795311005842152
Iteration Number: 338 / Loss: 9.558463497764887 / Norm of grad_w, grad_b: 0.4771192374124123 , 0.03515941377010988 / Norm
of weights: 3.7953450037201057
Iteration Number: 339 / Loss: 9.558602598106235 / Norm of grad_w, grad_b: 0.47511996397410905 , 0.03515800709825481 / Nor
m of weights: 3.79537860975246
Iteration Number: 340 / Loss: 9.558740101713008 / Norm of grad_w, grad_b: 0.47313443495674 , 0.03515660447417282 / Norm o
f weights: 3.795411829829144
Iteration Number: 341 / Loss: 9.558876033055006 / Norm of grad_w, grad_b: 0.47116245504529264 , 0.035155205860286885 / No
rm of weights: 3.7954446697253466
Iteration Number: 342 / Loss: 9.559010416124476 / Norm of grad_w, grad_b: 0.46920383350576583 , 0.0351538112195953 / Norm
of weights: 3.7954771351037775
Iteration Number: 343 / Loss: 9.55914327444552 / Norm of grad_w, grad_b: 0.46725838405062375 , 0.035152420051566159 / Norm
of weights: 3.795509231516891
Iteration Number: 344 / Loss: 9.55927463108331 / Norm of grad_w, grad_b: 0.4653259247083677 , 0.03515103371260451 / Norm
```

```
of weights: 3.795540964409066
Iteration Number: 345 / Loss: 9.559404508653152 / Norm of grad_w, grad_b: 0.4634062776972829 , 0.03514965077508823 / Norm
of weights: 3.7955723391187437
Iteration Number: 346 / Loss: 9.559532929329363 / Norm of grad_w, grad_b: 0.4614992693031241 , 0.03514827166831286 / Norm
of weights: 3.7956033608805266
Iteration Number: 347 / Loss: 9.559659914853981 / Norm of grad_w, grad_b: 0.45960472976080713 , 0.03514689635800493 / Nor
m of weights: 3.795634034827237
Iteration Number: 348 / Loss: 9.559785486545325 / Norm of grad_w, grad_b: 0.4577224931396282 , 0.03514552481040806 / Norm
of weights: 3.7956643659919322
Iteration Number: 349 / Loss: 9.559909665306371 / Norm of grad_w, grad_b: 0.455852397232273 , 0.03514415699227396 / Norm
of weights: 3.7956943593098904
Iteration Number: 350 / Loss: 9.560032471632985 / Norm of grad_w, grad_b: 0.45399428344723036 , 0.035142792870853444 / No
rm of weights: 3.795724019620546
Iteration Number: 351 / Loss: 9.560153925621986 / Norm of grad_w, grad_b: 0.4521479967047156 , 0.035141432413887605 / Nor
m of weights: 3.7957533516694
Iteration Number: 352 / Loss: 9.560274046979075 / Norm of grad_w, grad_b: 0.4503133853357286 , 0.03514007558959924 / Nor
m of weights: 3.7957823601098855
Iteration Number: 353 / Loss: 9.560392855026581 / Norm of grad_w, grad_b: 0.4484903009844064 , 0.035138722366684194 / Nor
m of weights: 3.795811049505201
Iteration Number: 354 / Loss: 9.560510368711107 / Norm of grad_w, grad_b: 0.44667859851341 , 0.03513737271430336 / Norm o
f weights: 3.795839424330105
Iteration Number: 355 / Loss: 9.56062660661096 / Norm of grad_w, grad_b: 0.44487813591240183 , 0.035136026602074065 / Nor
m of weights: 3.7958674889726822
Iteration Number: 356 / Loss: 9.56074158694351 / Norm of grad_w, grad_b: 0.4430887742091242 , 0.035134684000062415 / Norm
of weights: 3.7958952477360692
Iteration Number: 357 / Loss: 9.56085532757237 / Norm of grad_w, grad_b: 0.4413103773836972 , 0.035133344878775084 / Norm
of weights: 3.7959227048401494
Iteration Number: 358 / Loss: 9.56096784601442 / Norm of grad_w, grad_b: 0.4395428122851957 , 0.03513200920915177 / Norm
of weights: 3.795949864423214
Iteration Number: 359 / Loss: 9.561079159446749 / Norm of grad_w, grad_b: 0.437785948551122 , 0.03513067696255749 / Norm
of weights: 3.795976730543593
Iteration Number: 360 / Loss: 9.561189284713407 / Norm of grad_w, grad_b: 0.4360396585293414 , 0.035129348110775116 / Nor
m of weights: 3.796003307181254
Iteration Number: 361 / Loss: 9.561298238332075 / Norm of grad_w, grad_b: 0.43430381720241157 , 0.03512802262599792 / Nor
m of weights: 3.7960295982393646
Iteration Number: 362 / Loss: 9.561406036500552 / Norm of grad_w, grad_b: 0.43257830211433695 , 0.035126700480822566 / No
rm of weights: 3.7960556075458323
Iteration Number: 363 / Loss: 9.56151269510317 / Norm of grad_w, grad_b: 0.4308629932996732 , 0.0351253816482418 / Norm o
f weights: 3.796081338854811
Iteration Number: 364 / Loss: 9.561618229717052 / Norm of grad_w, grad_b: 0.4291577732146899 , 0.03512406610163759 / Norm
of weights: 3.7961067958481745
Iteration Number: 365 / Loss: 9.561722655618254 / Norm of grad_w, grad_b: 0.4274625266708252 , 0.03512275381477443 / Norm
of weights: 3.796131982136967
Iteration Number: 366 / Loss: 9.561825987787786 / Norm of grad_w, grad_b: 0.42577714077025597 , 0.035121444761792214 / No
rm of weights: 3.7961569012628207
```

```
Iteration Number: 367 / Loss: 9.561928240917528 / Norm of grad_w, grad_b: 0.42410150484331155 , 0.035120138917200144 / No
rm of weights: 3.7961815566993495
Iteration Number: 368 / Loss: 9.56202942941601 / Norm of grad_w, grad_b: 0.4224355103879483 , 0.03511883625586998 / Norm
of weights: 3.7962059518535116
Iteration Number: 369 / Loss: 9.562129567414104 / Norm of grad_w, grad_b: 0.4207790510111292 , 0.03511753675302955 / Norm
of weights: 3.796230090066947
Iteration Number: 370 / Loss: 9.562228668770572 / Norm of grad_w, grad_b: 0.4191320223719791 , 0.03511624038425696 / Norm
of weights: 3.7962539746172874
Iteration Number: 371 / Loss: 9.562326747077535 / Norm of grad_w, grad_b: 0.4174943221267026 , 0.035114947125473944 / Nor
m of weights: 3.7962776087194468
Iteration Number: 372 / Loss: 9.562423815665827 / Norm of grad_w, grad_b: 0.41586584987521513 , 0.035113656952940255 / No
rm of weights: 3.7963009955268725
Iteration Number: 373 / Loss: 9.562519887610222 / Norm of grad_w, grad_b: 0.41424650710949135 , 0.03511236984324751 / Nor
m of weights: 3.796324138132788
Iteration Number: 374 / Loss: 9.562614975734594 / Norm of grad_w, grad_b: 0.4126361971634070 , 0.03511108577331353 / Norm
of weights: 3.7963470395714007
Iteration Number: 375 / Loss: 9.562709092616945 / Norm of grad_w, grad_b: 0.41103482516417034 , 0.035109804720376565 / No
rm of weights: 3.7963697028190873
Iteration Number: 376 / Loss: 9.562802250594359 / Norm of grad_w, grad_b: 0.4094422979853185 , 0.03510852666198973 / Norm
of weights: 3.796392130795558
Iteration Number: 377 / Loss: 9.562894461767822 / Norm of grad_w, grad_b: 0.4078585242010507 , 0.03510725157601548 / Norm
of weights: 3.7964143263649968
Iteration Number: 378 / Loss: 9.562985738006995 / Norm of grad_w, grad_b: 0.40628341404199114 , 0.03510597944062037 / Nor
m of weights: 3.796436292337179
Iteration Number: 379 / Loss: 9.563076090954851 / Norm of grad_w, grad_b: 0.40471687935244766 , 0.03510471023426953 / Nor
m of weights: 3.7964580314685663
Iteration Number: 380 / Loss: 9.563165532032244 / Norm of grad_w, grad_b: 0.4031588335486765 , 0.035103443935721707 / Nor
m of weights: 3.7964795464633787
Iteration Number: 381 / Loss: 9.563254072442371 / Norm of grad_w, grad_b: 0.4016091915788001 , 0.03510218052402394 / Norm
of weights: 3.7965008399746494
Iteration Number: 382 / Loss: 9.563341723175162 / Norm of grad_w, grad_b: 0.40006786988358933 , 0.03510091997850682 / Nor
m of weights: 3.796521914605255
Iteration Number: 383 / Loss: 9.56342849501158 / Norm of grad_w, grad_b: 0.3985347863587797 , 0.03509966227877923 / Norm
of weights: 3.7965427729089245
Iteration Number: 384 / Loss: 9.563514398527815 / Norm of grad_w, grad_b: 0.39700986031816005 , 0.03509840740472392 / Nor
m of weights: 3.7965634173912317
Iteration Number: 385 / Loss: 9.563599444099422 / Norm of grad_w, grad_b: 0.3954930124580963 , 0.035097155336492464 / Nor
m of weights: 3.7965838505105642
Iteration Number: 386 / Loss: 9.563683641905357 / Norm of grad_w, grad_b: 0.39398416482302245 , 0.03509590605450071 / Nor
m of weights: 3.7966040746790775
Iteration Number: 387 / Loss: 9.563767001931948 / Norm of grad_w, grad_b: 0.39248324077186314 , 0.03509465953942434 / Nor
m of weights: 3.796624092263625
Iteration Number: 388 / Loss: 9.563849533976784 / Norm of grad_w, grad_b: 0.3909901649456456 , 0.03509341577219412 / Norm
of weights: 3.796643905586671
Iteration Number: 389 / Loss: 9.563931247652494 / Norm of grad_w, grad_b: 0.38950486323592265 , 0.03509217473399182 / Nor
```

```
m of weights: 3.7966635169271883
Iteration Number: 390 / Loss: 9.56401215239051 / Norm of grad_w, grad_b: 0.3880272627543114 , 0.035090936406245665 / Norm
of weights: 3.796682928521535
Iteration Number: 391 / Loss: 9.564092257444711 / Norm of grad_w, grad_b: 0.3865572918027311 , 0.03508970077062616 / Norm
of weights: 3.7967021425643135
Iteration Number: 392 / Loss: 9.564171571894999 / Norm of grad_w, grad_b: 0.3850948798447365 , 0.03508846780904206 / Norm
of weights: 3.7967211612092133
Iteration Number: 393 / Loss: 9.564250104650815 / Norm of grad_w, grad_b: 0.38363995747761465 , 0.03508723750363601 / Nor
m of weights: 3.7967399865698366
Iteration Number: 394 / Loss: 9.564327864454581 / Norm of grad_w, grad_b: 0.3821924564051796 , 0.03508600983678088 / Norm
of weights: 3.79675862072051
Iteration Number: 395 / Loss: 9.564404859885082 / Norm of grad_w, grad_b: 0.38075230941163996 , 0.035084784791075595 / No
rm of weights: 3.7967770656970736
Iteration Number: 396 / Loss: 9.564481099360746 / Norm of grad_w, grad_b: 0.3793194503360163 , 0.0350835623493413 / Norm
of weights: 3.7967953234976584
Iteration Number: 397 / Loss: 9.564556591142898 / Norm of grad_w, grad_b: 0.3778938140474191 , 0.03508234249461766 / Norm
of weights: 3.7968133960834516
Iteration Number: 398 / Loss: 9.56463134333894 / Norm of grad_w, grad_b: 0.37647533642101255 , 0.03508112521015913 / Norm
of weights: 3.796831285379437
Iteration Number: 399 / Loss: 9.564705363905443 / Norm of grad_w, grad_b: 0.37506395431467293 , 0.03507991047943121 / Nor
m of weights: 3.79684899327513
Iteration Number: 400 / Loss: 9.564778660651198 / Norm of grad_w, grad_b: 0.3736596055462785 , 0.03507869828610696 / Norm
of weights: 3.796866521625293
Iteration Number: 401 / Loss: 9.564851241240218 / Norm of grad_w, grad_b: 0.3722622288719091 , 0.035077488614063415 / Nor
m of weights: 3.796883872250635
Iteration Number: 402 / Loss: 9.564923113194636 / Norm of grad_w, grad_b: 0.3708717639643054 , 0.035076281447378226 / Nor
m of weights: 3.7969010469385025
Iteration Number: 403 / Loss: 9.564994283897594 / Norm of grad_w, grad_b: 0.3694881513921511 , 0.035075076770326175 / Nor
m of weights: 3.796918047443551
Iteration Number: 404 / Loss: 9.565064760596043 / Norm of grad_w, grad_b: 0.3681113326000918 , 0.035073874567375825 / Nor
m of weights: 3.796934875488405
Iteration Number: 405 / Loss: 9.565134550403483 / Norm of grad_w, grad_b: 0.36674124988894097 , 0.03507267482318647 / Nor
m of weights: 3.7969515327643073
Iteration Number: 406 / Loss: 9.56520366030269 / Norm of grad_w, grad_b: 0.36537784639690174 , 0.03507147752260465 / Norm
of weights: 3.7969680209317502
Iteration Number: 407 / Loss: 9.565272097148327 / Norm of grad_w, grad_b: 0.3640210660809108 , 0.03507028265066122 / Norm
of weights: 3.7969843416210964
Iteration Number: 408 / Loss: 9.565339867669547 / Norm of grad_w, grad_b: 0.36267085369881935 , 0.03506909019256805 / Nor
m of weights: 3.79700049643319
Iteration Number: 409 / Loss: 9.565406978472527 / Norm of grad_w, grad_b: 0.3613271547919158 , 0.035067900133715295 / Nor
m of weights: 3.797016486939949
Iteration Number: 410 / Loss: 9.565473436042955 / Norm of grad_w, grad_b: 0.3599899156679752 , 0.035066712459668165 / Nor
m of weights: 3.797032314684952
Iteration Number: 411 / Loss: 9.565539246748463 / Norm of grad_w, grad_b: 0.35865908338486363 , 0.03506552715616406 / Nor
m of weights: 3.7970479811840105
```

```
Iteration Number: 412 / Loss: 9.565604416841015 / Norm of grad_w, grad_b: 0.35733460573442827 , 0.03506434420910982 / Nor
m of weights: 3.7970634879257266
Iteration Number: 413 / Loss: 9.565668952459236 / Norm of grad_w, grad_b: 0.35601643122703913 , 0.03506316360457874 / Nor
m of weights: 3.797078836372045
Iteration Number: 414 / Loss: 9.565732859630714 / Norm of grad_w, grad_b: 0.3547045090763778 , 0.035061985328808026 / Nor
m of weights: 3.797094027958789
Iteration Number: 415 / Loss: 9.56579614427423 / Norm of grad_w, grad_b: 0.35339878918476414 , 0.035060809368195826 / Nor
m of weights: 3.797109064096191
Iteration Number: 416 / Loss: 9.565858812201977 / Norm of grad_w, grad_b: 0.3520992221288802 , 0.035059635709298737 / Nor
m of weights: 3.797123946169404
Iteration Number: 417 / Loss: 9.565920869121687 / Norm of grad_w, grad_b: 0.3508057591457951 , 0.03505846433882927 / Norm
of weights: 3.797138675539013
Iteration Number: 418 / Loss: 9.565982320638758 / Norm of grad_w, grad_b: 0.3495183521195211 , 0.035057295243653065 / Nor
m of weights: 3.7971532535415253
Iteration Number: 419 / Loss: 9.566043172258322 / Norm of grad_w, grad_b: 0.3482369535677592 , 0.035056128410786494 / Nor
m of weights: 3.79716768148986
Iteration Number: 420 / Loss: 9.566103429387255 / Norm of grad_w, grad_b: 0.3469615166291111 , 0.035054963827394266 / Nor
m of weights: 3.797181960673823
Iteration Number: 421 / Loss: 9.566163097336183 / Norm of grad_w, grad_b: 0.3456919950507345 , 0.03505380148078684 / Norm
of weights: 3.7971960923605717
Iteration Number: 422 / Loss: 9.566222181321406 / Norm of grad_w, grad_b: 0.344428343175981 , 0.03505264135841823 / Norm
of weights: 3.797210077795074
Iteration Number: 423 / Loss: 9.566280686466815 / Norm of grad_w, grad_b: 0.3431705159328516 , 0.03505148344788348 / Norm
of weights: 3.797223918200552
Iteration Number: 424 / Loss: 9.566338617805746 / Norm of grad_w, grad_b: 0.3419184688223176 , 0.03505032773691653 / Norm
of weights: 3.797237614778924
Iteration Number: 425 / Loss: 9.56639598028282 / Norm of grad_w, grad_b: 0.3406721579072372 , 0.0350491742133879 / Norm o
f weights: 3.7972511687112314
Iteration Number: 426 / Loss: 9.566452778755723 / Norm of grad_w, grad_b: 0.3394315398014573 , 0.035048022865302376 / Nor
m of weights: 3.7972645811580605
Iteration Number: 427 / Loss: 9.566509017996967 / Norm of grad_w, grad_b: 0.3381965716591417 , 0.03504687368079713 / Norm
of weights: 3.797277853259953
Iteration Number: 428 / Loss: 9.566564702695606 / Norm of grad_w, grad_b: 0.3369672111645056 , 0.035045726648139264 / Nor
m of weights: 3.7972909861378126
Iteration Number: 429 / Loss: 9.566619837458916 / Norm of grad_w, grad_b: 0.3357434165217669 , 0.03504458175572394 / Norm
of weights: 3.7973039808932962
Iteration Number: 430 / Loss: 9.566674426814048 / Norm of grad_w, grad_b: 0.3345251464453139 , 0.03504343899207226 / Norm
of weights: 3.797316838609205
Iteration Number: 431 / Loss: 9.566728475209645 / Norm of grad_w, grad_b: 0.3333123601501971 , 0.0350422983458292 / Norm
 of weights: 3.7973295603498616
Iteration Number: 432 / Loss: 9.56678198701742 / Norm of grad_w, grad_b: 0.33210501734277353 , 0.035041159805761724 / Nor
m of weights: 3.797342147161483
Iteration Number: 433 / Loss: 9.566834966533705 / Norm of grad_w, grad_b: 0.3309030782117162 , 0.03504002336075677 / Norm
of weights: 3.797354600072542
Iteration Number: 434 / Loss: 9.566887417980976 / Norm of grad_w, grad_b: 0.3297065034190732 , 0.03503888899981944 / Norm
```

```
of weights: 3.7973669200941296
Iteration Number: 435 / Loss: 9.566939345509333 / Norm of grad_w, grad_b: 0.3285152540918926 , 0.03503775671207103 / Norm
of weights: 3.7973791082202983
Iteration Number: 436 / Loss: 9.56699075319797 / Norm of grad_w, grad_b: 0.32732929181345144 , 0.03503662648674717 / Norm
of weights: 3.7973911654284067
Iteration Number: 437 / Loss: 9.56704164505658 / Norm of grad_w, grad_b: 0.32614857861541974 , 0.03503549831319625 / Norm
of weights: 3.797403092679458
Iteration Number: 438 / Loss: 9.567092025026792 / Norm of grad_w, grad_b: 0.32497307696978434 , 0.03503437218087732 / Nor
m of weights: 3.7974148909184207
Iteration Number: 439 / Loss: 9.567141896983486 / Norm of grad_w, grad_b: 0.323802749780911 , 0.035033248079358606 / Norm
of weights: 3.79742656107456
Iteration Number: 440 / Loss: 9.567191264736191 / Norm of grad_w, grad_b: 0.32263756037816577 , 0.03503212599831567 / Nor
m of weights: 3.7974381040617438
Iteration Number: 441 / Loss: 9.567240132030358 / Norm of grad_w, grad_b: 0.32147747250841086 , 0.03503100592752982 / Nor
m of weights: 3.7974495207787586
Iteration Number: 442 / Loss: 9.567288502548672 / Norm of grad_w, grad_b: 0.320322450328744 , 0.03502988785688643 / Norm
 of weights: 3.7974608121096076
Iteration Number: 443 / Loss: 9.567336379912305 / Norm of grad_w, grad_b: 0.31917245839955993 , 0.03502877177637329 / Nor
m of weights: 3.7974719789238094
Iteration Number: 444 / Loss: 9.567383767682148 / Norm of grad_w, grad_b: 0.3180274616775923 , 0.03502765767607907 / Norm
of weights: 3.797483022076685
Iteration Number: 445 / Loss: 9.567430669360027 / Norm of grad_w, grad_b: 0.3168874255091683 , 0.03502654554619183 / Norm
of weights: 3.797493942409647
Iteration Number: 446 / Loss: 9.56747708838989 / Norm of grad_w, grad_b: 0.31575231562373196 , 0.03502543537699725 / Norm
of weights: 3.7975047407504725
Iteration Number: 447 / Loss: 9.56752302815896 / Norm of grad_w, grad_b: 0.31462209812745845 , 0.0350243271588774 / Norm
 of weights: 3.7975154179135804
Iteration Number: 448 / Loss: 9.56756849199888 / Norm of grad_w, grad_b: 0.31349673949687706 , 0.03502322088230922 / Norm
of weights: 3.7975259747002963
Iteration Number: 449 / Loss: 9.56761348318683 / Norm of grad_w, grad_b: 0.3123762065729406 , 0.03502211653786284 / Norm
 of weights: 3.7975364118991144
Iteration Number: 450 / Loss: 9.567658004946612 / Norm of grad_w, grad_b: 0.3112604665548796 , 0.03502101411620045 / Norm
of weights: 3.797546730285954
Iteration Number: 451 / Loss: 9.567702060449715 / Norm of grad_w, grad_b: 0.3101494869945915 , 0.03501991360807478 / Norm
of weights: 3.797556930624411
Iteration Number: 452 / Loss: 9.56774565281638 / Norm of grad_w, grad_b: 0.3090432357907606 , 0.035018815004327734 / Norm
of weights: 3.797567013666004
Iteration Number: 453 / Loss: 9.567788785116605 / Norm of grad_w, grad_b: 0.3079416811834123 , 0.03501771829588904 / Norm
of weights: 3.797576980150413
Iteration Number: 454 / Loss: 9.567831460371163 / Norm of grad_w, grad_b: 0.30684479174845286 , 0.03501662347377492 / Nor
m of weights: 3.79758683080572
Iteration Number: 455 / Loss: 9.567873681552589 / Norm of grad_w, grad_b: 0.30575253639236083 , 0.03501553052908684 / Nor
m of weights: 3.797596566348636
Iteration Number: 456 / Loss: 9.567915451586135 / Norm of grad_w, grad_b: 0.30466488434699884 , 0.03501443945301022 / Nor
m of weights: 3.797606187484728
```

```
Iteration Number: 457 / Loss: 9.56795677335071 / Norm of grad_w, grad_b: 0.3035818051644968 , 0.03501335023681313 / Norm
 of weights: 3.7976156949086435
Iteration Number: 458 / Loss: 9.56799764967983 / Norm of grad_w, grad_b: 0.3025032687123698 , 0.035012262871845154 / Norm
of weights: 3.797625089304324
Iteration Number: 459 / Loss: 9.568038083362492 / Norm of grad_w, grad_b: 0.3014292451687021 , 0.0350111773495361 / Norm
 of weights: 3.797634371345222
Iteration Number: 460 / Loss: 9.56807807714409 / Norm of grad_w, grad_b: 0.3003597050172505 , 0.03501009366139483 / Norm
 of weights: 3.797643541694506
Iteration Number: 461 / Loss: 9.568117633727262 / Norm of grad_w, grad_b: 0.29929461904287374 , 0.03500901179900818 / Nor
m of weights: 3.797652601005266
Iteration Number: 462 / Loss: 9.568156755772758 / Norm of grad_w, grad_b: 0.29823395832705013 , 0.03500793175403971 / Nor
m of weights: 3.797661549920713
Iteration Number: 463 / Loss: 9.568195445900267 / Norm of grad_w, grad_b: 0.2971776942434047 , 0.035006853518228624 / Nor
m of weights: 3.7976703890743773
Iteration Number: 464 / Loss: 9.568233706689234 / Norm of grad_w, grad_b: 0.29612579845328896 , 0.03500577708338871 / Nor
m of weights: 3.7976791190902963
Iteration Number: 465 / Loss: 9.56827154067966 / Norm of grad_w, grad_b: 0.2950782429016939 , 0.03500470244140712 / Norm
 of weights: 3.797687740583207
Iteration Number: 466 / Loss: 9.568308950372899 / Norm of grad_w, grad_b: 0.2940349998128918 , 0.03500362958424348 / Norm
of weights: 3.797696254158726
Iteration Number: 467 / Loss: 9.568345938232397 / Norm of grad_w, grad_b: 0.29299604168647825 , 0.03500255850392878 / Nor
m of weights: 3.7977046604135314
Iteration Number: 468 / Loss: 9.568382506684475 / Norm of grad_w, grad_b: 0.2919613412933232 , 0.03500148919256427 / Norm
of weights: 3.7977129599355424
Iteration Number: 469 / Loss: 9.568418658119054 / Norm of grad_w, grad_b: 0.29093087167179243 , 0.03500042164232056 / Nor
m of weights: 3.7977211533040873
Iteration Number: 470 / Loss: 9.568454394890367 / Norm of grad_w, grad_b: 0.28990460612369034 , 0.03499935584543656 / Nor
m of weights: 3.7977292410900763
Iteration Number: 471 / Loss: 9.568489719317677 / Norm of grad_w, grad_b: 0.2888825182107333 , 0.034998291794218574 / Nor
m of weights: 3.797737223856167
Iteration Number: 472 / Loss: 9.568524633685975 / Norm of grad_w, grad_b: 0.28786458175072893 , 0.03499722948103926 / Nor
m of weights: 3.7977451021569264
Iteration Number: 473 / Loss: 9.568559140246645 / Norm of grad_w, grad_b: 0.28685077081410026 , 0.034996168898336726 / No
rm of weights: 3.79775287653899
Iteration Number: 474 / Loss: 9.56859324121813 / Norm of grad_w, grad_b: 0.28584105972019463 , 0.03499511003861367 / Norm
of weights: 3.797760547541219
Iteration Number: 475 / Loss: 9.568626938786595 / Norm of grad_w, grad_b: 0.28483542303404163 , 0.034994052894436384 / No
rm of weights: 3.797768115694853
Iteration Number: 476 / Loss: 9.568660235106552 / Norm of grad_w, grad_b: 0.2838338355627556 , 0.03499299745843388 / Norm
of weights: 3.797775581523658
Iteration Number: 477 / Loss: 9.568693132301492 / Norm of grad_w, grad_b: 0.2828362723523878 , 0.034991943723297045 / Nor
m of weights: 3.797782945544075
Iteration Number: 478 / Loss: 9.568725632464485 / Norm of grad_w, grad_b: 0.28184270868454736 , 0.03499089168177781 / Nor
m of weights: 3.797790208265363
Iteration Number: 479 / Loss: 9.568757737658796 / Norm of grad_w, grad_b: 0.2808531200732668 , 0.03498984132668825 / Norm
```

```
of weights: 3.7977973701897403
Iteration Number: 480 / Loss: 9.568789449918462 / Norm of grad_w, grad_b: 0.27986748226190206 , 0.03498879265089967 / Nor
m of weights: 3.7978044318125197
Iteration Number: 481 / Loss: 9.568820771248866 / Norm of grad_w, grad_b: 0.2788857712199979 , 0.03498774564734206 / Norm
of weights: 3.7978113936222466
Iteration Number: 482 / Loss: 9.568851703627299 / Norm of grad_w, grad_b: 0.2779079631403165 , 0.03498670030900303 / Norm
of weights: 3.7978182561008307
Iteration Number: 483 / Loss: 9.568882249003519 / Norm of grad_w, grad_b: 0.27693403443587866 , 0.03498565662892704 / Nor
m of weights: 3.797825019723674
Iteration Number: 484 / Loss: 9.568912409300285 / Norm of grad_w, grad_b: 0.2759639617370665 , 0.03498461460021483 / Norm
of weights: 3.7978316849597986
Iteration Number: 485 / Loss: 9.56894218641389 / Norm of grad_w, grad_b: 0.2749977218887839 , 0.034983574216022516 / Norm
of weights: 3.797838252271972
Iteration Number: 486 / Loss: 9.568971582214672 / Norm of grad_w, grad_b: 0.2740352919476518 , 0.03498253546956077 / Norm
of weights: 3.797844722116825
Iteration Number: 487 / Loss: 9.569000598547527 / Norm of grad_w, grad_b: 0.2730766491792962 , 0.03498149835409423 / Norm
of weights: 3.7978510949449777
Iteration Number: 488 / Loss: 9.569029237232412 / Norm of grad_w, grad_b: 0.2721217710556396 , 0.03498046286294073 / Norm
of weights: 3.79785737120115
Iteration Number: 489 / Loss: 9.56905750006482 / Norm of grad_w, grad_b: 0.27117063525227336 , 0.03497942898947053 / Norm
of weights: 3.797863551324278
Iteration Number: 490 / Loss: 9.569085388816266 / Norm of grad_w, grad_b: 0.2702232196458586 , 0.0349783967271058 / Norm
 of weights: 3.7978696357476274
Iteration Number: 491 / Loss: 9.569112905234743 / Norm of grad_w, grad_b: 0.2692795023115905 , 0.034977366069319595 / Nor
m of weights: 3.797875624898902
Iteration Number: 492 / Loss: 9.569140051045203 / Norm of grad_w, grad_b: 0.2683394615207112 , 0.03497633700963553 / Norm
of weights: 3.7978815192003506
Iteration Number: 493 / Loss: 9.56916682794997 / Norm of grad_w, grad_b: 0.2674030757380158 , 0.034975309541627024 / Norm
of weights: 3.797887319068875
Iteration Number: 494 / Loss: 9.569193237629221 / Norm of grad_w, grad_b: 0.2664703236195053 , 0.03497428365891641 / Norm
of weights: 3.7978930249161302
Iteration Number: 495 / Loss: 9.569219281741379 / Norm of grad_w, grad_b: 0.2655411840099941 , 0.03497325935517473 / Norm
of weights: 3.797898637148626
Iteration Number: 496 / Loss: 9.569244961923552 / Norm of grad_w, grad_b: 0.26461563594072124 , 0.03497223662412066 / Nor
m of weights: 3.7979041561678284
Iteration Number: 497 / Loss: 9.56927027979195 / Norm of grad_w, grad_b: 0.2636936586271908 , 0.0349712154595202 / Norm o
f weights: 3.797909582370253
Iteration Number: 498 / Loss: 9.56929523694228 / Norm of grad_w, grad_b: 0.26277523146680615 , 0.034970195855185925 / Nor
m of weights: 3.797914916147564
Iteration Number: 499 / Loss: 9.56931983495014 / Norm of grad_w, grad_b: 0.2618603340367093 , 0.034969177804976465 / Norm
of weights: 3.7979201578866615
Iteration Number: 500 / Loss: 9.569344075371413 / Norm of grad_w, grad_b: 0.2609489460917154 , 0.0349681613027958 / Norm
 of weights: 3.79792530796978
Iteration Number: 501 / Loss: 9.569367959742651 / Norm of grad_w, grad_b: 0.26004104756196034 , 0.034967146342592806 / No
rm of weights: 3.797930366774571
```

```
Iteration Number: 502 / Loss: 9.569391489581438 / Norm of grad_w, grad_b: 0.25913661855098347 , 0.03496613291836071 / Nor
m of weights: 3.7979353346741953
Iteration Number: 503 / Loss: 9.569414666386757 / Norm of grad_w, grad_b: 0.2582356393336342 , 0.034965121024136285 / Nor
m of weights: 3.7979402120374046
Iteration Number: 504 / Loss: 9.56943749163935 / Norm of grad_w, grad_b: 0.2573380903540029 , 0.03496411065399959 / Norm
 of weights: 3.797944999228629
Iteration Number: 505 / Loss: 9.569459966802066 / Norm of grad_w, grad_b: 0.2564439522233897 , 0.034963101802073335 / Nor
m of weights: 3.797949696608057
Iteration Number: 506 / Loss: 9.569482093320202 / Norm of grad_w, grad_b: 0.25555320571852363 , 0.034962094462522283 / No
rm of weights: 3.7979543045317166
Iteration Number: 507 / Loss: 9.569503872621844 / Norm of grad_w, grad_b: 0.2546658317794009 , 0.03496108862955273 / Norm
of weights: 3.797958823351554
Iteration Number: 508 / Loss: 9.569525306118184 / Norm of grad_w, grad_b: 0.2537818115075901 , 0.034960084297412125 / Nor
m of weights: 3.797963253415512
Iteration Number: 509 / Loss: 9.569546395203862 / Norm of grad_w, grad_b: 0.25290112616422816 , 0.03495908146038844 / Nor
m of weights: 3.797967595067604
Iteration Number: 510 / Loss: 9.56956714125726 / Norm of grad_w, grad_b: 0.2520237571682391 , 0.03495808011280967 / Norm
 of weights: 3.7979718486479888
Iteration Number: 511 / Loss: 9.569587545640825 / Norm of grad_w, grad_b: 0.25114968609458893 , 0.034957080249043405 / No
rm of weights: 3.7979760144930435
Iteration Number: 512 / Loss: 9.569607609701364 / Norm of grad_w, grad_b: 0.25027889467233433 , 0.03495608186349634 / Nor
m of weights: 3.797980092935434
Iteration Number: 513 / Loss: 9.569627334770342 / Norm of grad_w, grad_b: 0.24941136478309708 , 0.03495508495061376 / Nor
m of weights: 3.797984084304186
Iteration Number: 514 / Loss: 9.569646722164181 / Norm of grad_w, grad_b: 0.24854707845916907 , 0.03495408950487919 / Nor
m of weights: 3.797987988924752
Iteration Number: 515 / Loss: 9.569665773184536 / Norm of grad_w, grad_b: 0.24768601788189953 , 0.03495309552081366 / Nor
m of weights: 3.797991807119078
Iteration Number: 516 / Loss: 9.569684489118568 / Norm of grad_w, grad_b: 0.2468281653799969 , 0.03495210299297562 / Norm
of weights: 3.79799553920567
Iteration Number: 517 / Loss: 9.56970287123923 / Norm of grad_w, grad_b: 0.24597350342790697 , 0.03495111191596025 / Norm
of weights: 3.7979991854996578
Iteration Number: 518 / Loss: 9.569720920805526 / Norm of grad_w, grad_b: 0.24512201464418265 , 0.034950122284399116 / No
rm of weights: 3.798002746312858
Iteration Number: 519 / Loss: 9.569738639062779 / Norm of grad_w, grad_b: 0.24427368178992476 , 0.034949134092959654 / No
rm of weights: 3.798006221953834
Iteration Number: 520 / Loss: 9.569756027242876 / Norm of grad_w, grad_b: 0.24342848776714157 , 0.03494814733634489 / Nor
m of weights: 3.79800961272796
Iteration Number: 521 / Loss: 9.56977308656454 / Norm of grad_w, grad_b: 0.24258641561727712 , 0.03494716200929292 / Norm
of weights: 3.7980129189374754
Iteration Number: 522 / Loss: 9.569789818233552 / Norm of grad_w, grad_b: 0.24174744851964278 , 0.034946178106576575 / No
rm of weights: 3.798016140881546
Iteration Number: 523 / Loss: 9.569806223443006 / Norm of grad_w, grad_b: 0.24091156978996212 , 0.03494519562300283 / Nor
m of weights: 3.7980192788563207
Iteration Number: 524 / Loss: 9.569822303373547 / Norm of grad_w, grad_b: 0.24007876287887384 , 0.03494421455341274 / Nor
```

```
m of weights: 3.7980223331549854
Iteration Number: 525 / Loss: 9.5698380591936 / Norm of grad_w, grad_b: 0.23924901137046894 , 0.03494323489268074 / Norm
 of weights: 3.7980253040678176
Iteration Number: 526 / Loss: 9.569853492059586 / Norm of grad_w, grad_b: 0.23842229898084813 , 0.03494225663571438 / Nor
m of weights: 3.7980281918822407
Iteration Number: 527 / Loss: 9.569868603116161 / Norm of grad_w, grad_b: 0.2375986095567575 , 0.03494127977745406 / Norm
of weights: 3.7980309968828756
Iteration Number: 528 / Loss: 9.569883393496422 / Norm of grad_w, grad_b: 0.23677792707420506 , 0.034940304312872414 / No
rm of weights: 3.798033719351593
Iteration Number: 529 / Loss: 9.56989786432213 / Norm of grad_w, grad_b: 0.23596023563693952 , 0.03493933023697419 / Norm
of weights: 3.7980363595675617
Iteration Number: 530 / Loss: 9.569912016703901 / Norm of grad_w, grad_b: 0.23514551947535872 , 0.034938357544795726 / No
rm of weights: 3.7980389178072986
Iteration Number: 531 / Loss: 9.569925851741434 / Norm of grad_w, grad_b: 0.23433376294494643 , 0.03493738623140465 / Nor
m of weights: 3.7980413943447173
Iteration Number: 532 / Loss: 9.569939370523695 / Norm of grad_w, grad_b: 0.23352495052508643 , 0.034936416291899565 / No
rm of weights: 3.798043789451174
Iteration Number: 533 / Loss: 9.569952574129115 / Norm of grad_w, grad_b: 0.2327190668177017 , 0.0349354477214097 / Norm
 of weights: 3.7980461033955155
Iteration Number: 534 / Loss: 9.569965463625794 / Norm of grad_w, grad_b: 0.23191609654606865 , 0.03493448051509445 / Nor
m of weights: 3.798048336444122
Iteration Number: 535 / Loss: 9.569978040071668 / Norm of grad_w, grad_b: 0.23111602455345148 , 0.03493351466814323 / Nor
m of weights: 3.7980504888609543
Iteration Number: 536 / Loss: 9.569990304514716 / Norm of grad_w, grad_b: 0.23031883580195092 , 0.034932550175775 / Norm
 of weights: 3.798052560907594
Iteration Number: 537 / Loss: 9.570002257993126 / Norm of grad_w, grad_b: 0.2295245153712145 , 0.03493158703323807 / Norm
of weights: 3.79805455284329
Iteration Number: 538 / Loss: 9.570013901535484 / Norm of grad_w, grad_b: 0.22873304845729192 , 0.034930625235809656 / No
rm of weights: 3.7980564649249953
Iteration Number: 539 / Loss: 9.570025236160932 / Norm of grad_w, grad_b: 0.2279444203714338 , 0.03492966477879562 / Norm
of weights: 3.7980582974074117
Iteration Number: 540 / Loss: 9.570036262879347 / Norm of grad_w, grad_b: 0.22715861653886069 , 0.03492870565753023 / Nor
m of weights: 3.7980600505430293
Iteration Number: 541 / Loss: 9.570046982691515 / Norm of grad_w, grad_b: 0.2263756224976885 , 0.034927747867375736 / Nor
m of weights: 3.7980617245821633
Iteration Number: 542 / Loss: 9.570057396589277 / Norm of grad_w, grad_b: 0.22559542389776524 , 0.034926791403722186 / No
rm of weights: 3.7980633197729956
Iteration Number: 543 / Loss: 9.570067505555707 / Norm of grad_w, grad_b: 0.22481800649948125 , 0.03492583626198701 / Nor
m of weights: 3.7980648363616094
Iteration Number: 544 / Loss: 9.570077310565251 / Norm of grad_w, grad_b: 0.22404335617278248 , 0.03492488243761483 / Nor
m of weights: 3.798066274592028
Iteration Number: 545 / Loss: 9.570086812583893 / Norm of grad_w, grad_b: 0.2232714588959306 , 0.034923929926077193 / Nor
m of weights: 3.798067634706251
Iteration Number: 546 / Loss: 9.570096012569305 / Norm of grad_w, grad_b: 0.22250230075460034 , 0.03492297872287215 / Nor
m of weights: 3.7980689169442887
```

```
Iteration Number: 547 / Loss: 9.570104911470983 / Norm of grad_w, grad_b: 0.22173586794062433 , 0.03492202882352411 / Nor
m of weights: 3.7980701215441965
Iteration Number: 548 / Loss: 9.570113510230405 / Norm of grad_w, grad_b: 0.2209721467510917 , 0.03492108022358358 / Norm
of weights: 3.79807124874211
Iteration Number: 549 / Loss: 9.570121809781162 / Norm of grad_w, grad_b: 0.22021112358724187 , 0.03492013291862675 / Nor
m of weights: 3.7980722987722784
Iteration Number: 550 / Loss: 9.570129811049101 / Norm of grad_w, grad_b: 0.21945278495343662 , 0.03491918690425541 / Nor
m of weights: 3.798073271867096
Iteration Number: 551 / Loss: 9.570137514952469 / Norm of grad_w, grad_b: 0.21869711745617554 , 0.0349182421760965 / Norm
of weights: 3.7980741682571333
Iteration Number: 552 / Loss: 9.570144922402022 / Norm of grad_w, grad_b: 0.21794410780311865 , 0.03491729872980208 / Nor
m of weights: 3.7980749881711717
Iteration Number: 553 / Loss: 9.570152034301183 / Norm of grad_w, grad_b: 0.21719374280205203 , 0.03491635656104887 / Nor
m of weights: 3.7980757318362306
Iteration Number: 554 / Loss: 9.570158851546148 / Norm of grad_w, grad_b: 0.21644600935995326 , 0.03491541566553808 / Nor
m of weights: 3.7980763994776003
Iteration Number: 555 / Loss: 9.57016537502603 / Norm of grad_w, grad_b: 0.2157008944820097 , 0.034914476038995244 / Norm
of weights: 3.7980769913188697
Iteration Number: 556 / Loss: 9.570171605622965 / Norm of grad_w, grad_b: 0.2149583852706574 , 0.03491353767716984 / Norm
of weights: 3.7980775075819535
Iteration Number: 557 / Loss: 9.570177544212235 / Norm of grad_w, grad_b: 0.2142184689247741 , 0.0349126005758351 / Norm
 of weights: 3.798077948487126
Iteration Number: 558 / Loss: 9.5701831916624 / Norm of grad_w, grad_b: 0.21348113273858188 , 0.034911664730787835 / Norm
of weights: 3.7980783142530417
Iteration Number: 559 / Loss: 9.57018854883539 / Norm of grad_w, grad_b: 0.21274636410080872 , 0.034910730137848164 / Nor
m of weights: 3.798078605096769
Iteration Number: 560 / Loss: 9.57019361658664 / Norm of grad_w, grad_b: 0.21201415049385863 , 0.03490979679285922 / Norm
of weights: 3.7980788212338124
Iteration Number: 561 / Loss: 9.570198395765185 / Norm of grad_w, grad_b: 0.21128447949278867 , 0.034908864691687067 / No
rm of weights: 3.79807896287814
Iteration Number: 562 / Loss: 9.570202887213782 / Norm of grad_w, grad_b: 0.2105573387645837 , 0.034907933830220346 / Nor
m of weights: 3.79807903024221
Iteration Number: 563 / Loss: 9.570207091768996 / Norm of grad_w, grad_b: 0.20983271606717693 , 0.03490700420437006 / Nor
m of weights: 3.7980790235369932
Iteration Number: 564 / Loss: 9.570211010261325 / Norm of grad_w, grad_b: 0.20911059924861683 , 0.03490607581006954 / Nor
m of weights: 3.7980789429720003
Iteration Number: 565 / Loss: 9.570214643515289 / Norm of grad_w, grad_b: 0.20839097624629285 , 0.03490514864327399 / Nor
m of weights: 3.7980787887553045
Iteration Number: 566 / Loss: 9.570217992349544 / Norm of grad_w, grad_b: 0.20767383508601522 , 0.03490422269996043 / Nor
m of weights: 3.7980785610935652
Iteration Number: 567 / Loss: 9.570221057576955 / Norm of grad_w, grad_b: 0.20695916388126467 , 0.0349032979761274 / Norm
of weights: 3.798078260192051
Iteration Number: 568 / Loss: 9.57022384000472 / Norm of grad_w, grad_b: 0.20624695083233435 , 0.03490237446779487 / Norm
of weights: 3.7980778862546622
Iteration Number: 569 / Loss: 9.570226340434445 / Norm of grad_w, grad_b: 0.20553718422551995 , 0.03490145217100395 / Nor
```

```
m of weights: 3.7980774394839543
Iteration Number: 570 / Loss: 9.57022855966224 / Norm of grad_w, grad_b: 0.20482985243236163 , 0.0349005310818167 / Norm
 of weights: 3.798076920081157
Iteration Number: 571 / Loss: 9.570230498478818 / Norm of grad_w, grad_b: 0.20412494390887312 , 0.03489961119631591 / Nor
m of weights: 3.7980763282461987
Iteration Number: 572 / Loss: 9.570232157669572 / Norm of grad_w, grad_b: 0.20342244719472408 , 0.03489869251060504 / Nor
m of weights: 3.7980756641777247
Iteration Number: 573 / Loss: 9.57023353801467 / Norm of grad_w, grad_b: 0.20272235091248295 , 0.03489777502080791 / Norm
of weights: 3.798074928073119
Iteration Number: 574 / Loss: 9.570234640289128 / Norm of grad_w, grad_b: 0.20202464376691445 , 0.03489685872306845 / Nor
m of weights: 3.7980741201285246
Iteration Number: 575 / Loss: 9.570235465262915 / Norm of grad_w, grad_b: 0.201329314544138 , 0.03489594361355073 / Norm
 of weights: 3.798073240538862
Iteration Number: 576 / Loss: 9.570236013701011 / Norm of grad_w, grad_b: 0.20063635211101624 , 0.03489502968843861 / Nor
m of weights: 3.7980722894978496
Iteration Number: 577 / Loss: 9.570236286363501 / Norm of grad_w, grad_b: 0.19994574541434107 , 0.03489411694393555 / Nor
m of weights: 3.798071267198021
Iteration Number: 578 / Loss: 9.57023628400565 / Norm of grad_w, grad_b: 0.1992574834801049 , 0.03489320537626458 / Norm
 of weights: 3.7980701738307445
Iteration Number: 579 / Loss: 9.570236007377975 / Norm of grad_w, grad_b: 0.19857155541287164 , 0.034892294981667946 / No
rm of weights: 3.7980690095862433
Iteration Number: 580 / Loss: 9.570235457226332 / Norm of grad_w, grad_b: 0.19788795039497337 , 0.03489138575640712 / Nor
m of weights: 3.7980677746536076
Iteration Number: 581 / Loss: 9.570234634291973 / Norm of grad_w, grad_b: 0.19720665768590437 , 0.03489047769676242 / Nor
m of weights: 3.798066469220819
Iteration Number: 582 / Loss: 9.570233539311634 / Norm of grad_w, grad_b: 0.19652766662161236 , 0.03488957079903307 / Nor
m of weights: 3.7980650934747615
Iteration Number: 583 / Loss: 9.570232173017596 / Norm of grad_w, grad_b: 0.19585096661378726 , 0.03488866505953684 / Nor
m of weights: 3.798063647601243
Iteration Number: 584 / Loss: 9.57023053613776 / Norm of grad_w, grad_b: 0.19517654714919613 , 0.03488776047461002 / Norm
of weights: 3.7980621317850076
Iteration Number: 585 / Loss: 9.570228629395716 / Norm of grad_w, grad_b: 0.19450439778911613 , 0.03488685704060715 / Nor
m of weights: 3.7980605462097548
Iteration Number: 586 / Loss: 9.570226453510795 / Norm of grad_w, grad_b: 0.19383450816852807 , 0.034885954753900945 / No
rm of weights: 3.7980588910581523
Iteration Number: 587 / Loss: 9.570224009198153 / Norm of grad_w, grad_b: 0.1931668679955778 , 0.03488505361088214 / Norm
of weights: 3.7980571665118563
Iteration Number: 588 / Loss: 9.570221297168832 / Norm of grad_w, grad_b: 0.19250146705093835 , 0.03488415360795923 / Nor
m of weights: 3.79805537275152
Iteration Number: 589 / Loss: 9.57021831812981 / Norm of grad_w, grad_b: 0.1918382951871272 , 0.03488325474155844 / Norm
 of weights: 3.7980535099568136
Iteration Number: 590 / Loss: 9.57021507278407 / Norm of grad_w, grad_b: 0.19117734232785633 , 0.03488235700812357 / Norm
of weights: 3.7980515783064384
Iteration Number: 591 / Loss: 9.570211561830678 / Norm of grad_w, grad_b: 0.19051859846759134 , 0.03488146040411562 / Nor
m of weights: 3.798049577978136
```

```
Iteration Number: 592 / Loss: 9.5702077859648 / Norm of grad_w, grad_b: 0.18986205367069425 , 0.034880564926013075 / Norm
of weights: 3.7980475091487103
Iteration Number: 593 / Loss: 9.570203745877809 / Norm of grad_w, grad_b: 0.1892076980710341 , 0.03487967057031131 / Norm
of weights: 3.798045371994033
Iteration Number: 594 / Loss: 9.570199442257298 / Norm of grad_w, grad_b: 0.18855552187124885 , 0.03487877733352272 / Nor
m of weights: 3.798043166689065
Iteration Number: 595 / Loss: 9.570194875787172 / Norm of grad_w, grad_b: 0.1879055153422172 , 0.034877885212176545 / Nor
m of weights: 3.7980408934078618
Iteration Number: 596 / Loss: 9.570190047147673 / Norm of grad_w, grad_b: 0.18725766882250147 , 0.03487699420281862 / Nor
m of weights: 3.7980385523235927
Iteration Number: 597 / Loss: 9.570184957015456 / Norm of grad_w, grad_b: 0.18661197271774116 , 0.034876104302011324 / No
rm of weights: 3.7980361436085492
Iteration Number: 598 / Loss: 9.57017960606363 / Norm of grad_w, grad_b: 0.18596841750004525 , 0.034875215506333455 / Nor
m of weights: 3.7980336674341597
Iteration Number: 599 / Loss: 9.570173994961808 / Norm of grad_w, grad_b: 0.1853269937075072 , 0.03487432781238004 / Norm
of weights: 3.798031123971002
Iteration Number: 600 / Loss: 9.570168124376167 / Norm of grad_w, grad_b: 0.18468769194361992 , 0.034873441216762245 / No
rm of weights: 3.798028513388812
Iteration Number: 601 / Loss: 9.570161994969485 / Norm of grad_w, grad_b: 0.18405050287667318 , 0.034872555716107236 / No
rm of weights: 3.798025835856499
Iteration Number: 602 / Loss: 9.570155607401198 / Norm of grad_w, grad_b: 0.183415417239343 , 0.034871671307058 / Norm of
weights: 3.7980230915421562
Iteration Number: 603 / Loss: 9.570148962327446 / Norm of grad_w, grad_b: 0.18278242582798654 , 0.03487078798627333 / Nor
m of weights: 3.798020280613071
Iteration Number: 604 / Loss: 9.57014206040112 / Norm of grad_w, grad_b: 0.18215151950221223 , 0.03486990575042755 / Norm
of weights: 3.798017403235736
Iteration Number: 605 / Loss: 9.570134902271903 / Norm of grad_w, grad_b: 0.1815226891843904 , 0.03486902459621056 / Norm
of weights: 3.798014459575861
Iteration Number: 606 / Loss: 9.570127488586321 / Norm of grad_w, grad_b: 0.18089592585903108 , 0.03486814452032754 / Nor
m of weights: 3.798011449798383
Iteration Number: 607 / Loss: 9.570119819987783 / Norm of grad_w, grad_b: 0.18027122057231976 , 0.034867265519498945 / No
rm of weights: 3.7980083740674746
Iteration Number: 608 / Loss: 9.570111897116622 / Norm of grad_w, grad_b: 0.17964856443166993 , 0.03486638759046038 / Nor
m of weights: 3.798005232546559
Iteration Number: 609 / Loss: 9.570103720610142 / Norm of grad_w, grad_b: 0.17902794860508278 , 0.03486551072996242 / Nor
m of weights: 3.798002025398316
Iteration Number: 610 / Loss: 9.570095291102662 / Norm of grad_w, grad_b: 0.1784093643207813 , 0.03486463493477049 / Norm
of weights: 3.7979987527846912
Iteration Number: 611 / Loss: 9.570086609225545 / Norm of grad_w, grad_b: 0.17779280286669946 , 0.034863760201664845 / No
rm of weights: 3.7979954148669095
Iteration Number: 612 / Loss: 9.570077675607246 / Norm of grad_w, grad_b: 0.1771782555898713 , 0.03486288652744039 / Norm
of weights: 3.797992011805481
Iteration Number: 613 / Loss: 9.570068490873354 / Norm of grad_w, grad_b: 0.17656571389612963 , 0.034862013908906565 / No
rm of weights: 3.797988543760212
Iteration Number: 614 / Loss: 9.57005905564662 / Norm of grad_w, grad_b: 0.17595516924952412 , 0.03486114234288721 / Norm
```

```
of weights: 3.797985010890213
Iteration Number: 615 / Loss: 9.570049370547004 / Norm of grad_w, grad_b: 0.1753466131718703 , 0.03486027182622053 / Norm
of weights: 3.797981413353909
Iteration Number: 616 / Loss: 9.57003943619171 / Norm of grad_w, grad_b: 0.17474003724229947 , 0.03485940235575891 / Norm
of weights: 3.797977751309046
Iteration Number: 617 / Loss: 9.570029253195209 / Norm of grad_w, grad_b: 0.1741354330968212 , 0.034858533928368816 / Nor
m of weights: 3.7979740249127025
Iteration Number: 618 / Loss: 9.5700188221693 / Norm of grad_w, grad_b: 0.17353279242779518 , 0.03485766654093084 / Norm
 of weights: 3.797970234321295
Iteration Number: 619 / Loss: 9.57000814372312 / Norm of grad_w, grad_b: 0.17293210698353037 , 0.034856800190339286 / Nor
m of weights: 3.797966379690589
Iteration Number: 620 / Loss: 9.569997218463193 / Norm of grad_w, grad_b: 0.17233336856789272 , 0.03485593487350236 / Nor
m of weights: 3.7979624611757035
Iteration Number: 621 / Loss: 9.569986046993455 / Norm of grad_w, grad_b: 0.17173656903976012 , 0.034855070587341906 / No
rm of weights: 3.797958478931122
Iteration Number: 622 / Loss: 9.569974629915288 / Norm of grad_w, grad_b: 0.17114170031263093 , 0.034854207328793435 / No
rm of weights: 3.7979544331107005
Iteration Number: 623 / Loss: 9.569962967827568 / Norm of grad_w, grad_b: 0.17054875435426245 , 0.03485334509480582 / Nor
m of weights: 3.7979503238676706
Iteration Number: 624 / Loss: 9.569951061326655 / Norm of grad_w, grad_b: 0.16995772318614769 , 0.03485248388234139 / Nor
m of weights: 3.7979461513546533
Iteration Number: 625 / Loss: 9.56993891100648 / Norm of grad_w, grad_b: 0.1693685988831298 , 0.03485162368837577 / Norm
 of weights: 3.7979419157236625
Iteration Number: 626 / Loss: 9.56992651745853 / Norm of grad_w, grad_b: 0.1687813735730115 , 0.034850764509897734 / Norm
of weights: 3.797937617126111
Iteration Number: 627 / Loss: 9.569913881271898 / Norm of grad_w, grad_b: 0.16819603943611144 , 0.034849906343909186 / No
rm of weights: 3.797933255712822
Iteration Number: 628 / Loss: 9.569901003033308 / Norm of grad_w, grad_b: 0.16761258870487258 , 0.03484904918742496 / Nor
m of weights: 3.797928831634033
Iteration Number: 629 / Loss: 9.569887883327151 / Norm of grad_w, grad_b: 0.16703101366344916 , 0.0348481930374729 / Norm
of weights: 3.7979243450394002
Iteration Number: 630 / Loss: 9.569874522735487 / Norm of grad_w, grad_b: 0.16645130664733326 , 0.034847337891093594 / No
rm of weights: 3.797919796078012
Iteration Number: 631 / Loss: 9.569860921838112 / Norm of grad_w, grad_b: 0.16587346004296627 , 0.03484648374534035 / Nor
m of weights: 3.797915184898389
Iteration Number: 632 / Loss: 9.569847081212552 / Norm of grad_w, grad_b: 0.1652974662872293 , 0.034845630597279124 / Nor
m of weights: 3.7979105116484937
Iteration Number: 633 / Loss: 9.569833001434105 / Norm of grad_w, grad_b: 0.16472331786727937 , 0.034844778443988436 / No
rm of weights: 3.797905776475736
Iteration Number: 634 / Loss: 9.569818683075862 / Norm of grad_w, grad_b: 0.16415100731999807 , 0.034843927282559146 / No
rm of weights: 3.7979009795269794
Iteration Number: 635 / Loss: 9.569804126708737 / Norm of grad_w, grad_b: 0.16358052723158192 , 0.03484307711009468 / Nor
m of weights: 3.797896120948547
Iteration Number: 636 / Loss: 9.569789332901486 / Norm of grad_w, grad_b: 0.1630118702374017 , 0.034842227923710535 / Nor
m of weights: 3.797891200886227
```

```
Iteration Number: 637 / Loss: 9.569774302220736 / Norm of grad_w, grad_b: 0.1624450290213726 , 0.03484137972053454 / Norm
of weights: 3.7978862194852803
Iteration Number: 638 / Loss: 9.569759035231005 / Norm of grad_w, grad_b: 0.16187999631572506 , 0.03484053249770652 / Nor
m of weights: 3.797881176890443
Iteration Number: 639 / Loss: 9.56974353249473 / Norm of grad_w, grad_b: 0.16131676490062602 , 0.03483968625237843 / Norm
of weights: 3.7978760732459365
Iteration Number: 640 / Loss: 9.56972779457229 / Norm of grad_w, grad_b: 0.160755327603809 , 0.034838840981714116 / Norm
 of weights: 3.797870908695468
Iteration Number: 641 / Loss: 9.56971182202202 / Norm of grad_w, grad_b: 0.16019567730023582 , 0.03483799668288923 / Norm
of weights: 3.7978656833822395
Iteration Number: 642 / Loss: 9.56969561540025 / Norm of grad_w, grad_b: 0.1596378069116915 , 0.0348371533530913 / Norm o
f weights: 3.797860397448952
Iteration Number: 643 / Loss: 9.569679175261303 / Norm of grad_w, grad_b: 0.1590817094065507 , 0.03483631098951945 / Norm
of weights: 3.797855051037811
Iteration Number: 644 / Loss: 9.569662502157549 / Norm of grad_w, grad_b: 0.15852737779931145 , 0.034835469589384535 / No
rm of weights: 3.7978496442905296
Iteration Number: 645 / Loss: 9.569645596639386 / Norm of grad_w, grad_b: 0.15797480515031334 , 0.03483462914990885 / Nor
m of weights: 3.797844177348338
Iteration Number: 646 / Loss: 9.5696284592553 / Norm of grad_w, grad_b: 0.15742398456546916 , 0.0348337896683262 / Norm o
f weights: 3.797838650351983
Iteration Number: 647 / Loss: 9.56961109055186 / Norm of grad_w, grad_b: 0.15687490919581412 , 0.03483295114188171 / Norm
of weights: 3.797833063441737
Iteration Number: 648 / Loss: 9.569593491073741 / Norm of grad_w, grad_b: 0.15632757223725444 , 0.034832113567831856 / No
rm of weights: 3.797827416757401
Iteration Number: 649 / Loss: 9.569575661363753 / Norm of grad_w, grad_b: 0.15578196693015964 , 0.034831276943444435 / No
rm of weights: 3.7978217104383085
Iteration Number: 650 / Loss: 9.569557601962858 / Norm of grad_w, grad_b: 0.1552380865591879 , 0.03483044126599822 / Norm
of weights: 3.7978159446233337
Iteration Number: 651 / Loss: 9.56953931341018 / Norm of grad_w, grad_b: 0.1546959244528258 , 0.034829606532783226 / Norm
of weights: 3.7978101194508898
Iteration Number: 652 / Loss: 9.569520796243026 / Norm of grad_w, grad_b: 0.15415547398320387 , 0.03482877274110033 / Nor
m of weights: 3.79780423505894
Iteration Number: 653 / Loss: 9.569502050996919 / Norm of grad_w, grad_b: 0.1536167285655988 , 0.03482793988826149 / Norm
of weights: 3.7977982915849977
Iteration Number: 654 / Loss: 9.569483078205593 / Norm of grad_w, grad_b: 0.15307968165836752 , 0.03482710797158943 / Nor
m of weights: 3.7977922891661313
Iteration Number: 655 / Loss: 9.569463878401027 / Norm of grad_w, grad_b: 0.15254432676243654 , 0.034826276988417726 / No
rm of weights: 3.7977862279389694
Iteration Number: 656 / Loss: 9.569444452113455 / Norm of grad_w, grad_b: 0.15201065742108769 , 0.03482544693609064 / Nor
m of weights: 3.7977801080397056
Iteration Number: 657 / Loss: 9.569424799871388 / Norm of grad_w, grad_b: 0.15147866721973077 , 0.034824617811963096 / No
rm of weights: 3.7977739296040993
Iteration Number: 658 / Loss: 9.56940492220162 / Norm of grad_w, grad_b: 0.1509483497854481 , 0.03482378961340061 / Norm
 of weights: 3.7977676927674824
Iteration Number: 659 / Loss: 9.569384819629253 / Norm of grad_w, grad_b: 0.15041969878679698 , 0.03482296233777926 / Nor
```

```
m of weights: 3.7977613976647646
Iteration Number: 660 / Loss: 9.569364492677723 / Norm of grad_w, grad_b: 0.14989270793359336 , 0.03482213598248548 / Nor
m of weights: 3.7977550444304327
Iteration Number: 661 / Loss: 9.569343941868787 / Norm of grad_w, grad_b: 0.1493673709764811 , 0.03482131054491614 / Norm
of weights: 3.7977486331985575
Iteration Number: 662 / Loss: 9.569323167722565 / Norm of grad_w, grad_b: 0.14884368170672832 , 0.03482048602247846 / Nor
m of weights: 3.7977421641027975
Iteration Number: 663 / Loss: 9.569302170757545 / Norm of grad_w, grad_b: 0.1483216339559364 , 0.03481966241258988 / Norm
of weights: 3.7977356372764013
Iteration Number: 664 / Loss: 9.569280951490597 / Norm of grad_w, grad_b: 0.14780122159582412 , 0.03481883971267794 / Nor
m of weights: 3.7977290528522114
Iteration Number: 665 / Loss: 9.569259510436988 / Norm of grad_w, grad_b: 0.1472824385377845 , 0.03481801792018049 / Norm
of weights: 3.7977224109626686
Iteration Number: 666 / Loss: 9.569237848110397 / Norm of grad_w, grad_b: 0.146765278732841 , 0.03481719703254527 / Norm
 of weights: 3.7977157117398144
Iteration Number: 667 / Loss: 9.56921596502293 / Norm of grad_w, grad_b: 0.14624973617118814 , 0.03481637704723015 / Norm
of weights: 3.797708955315295
Iteration Number: 668 / Loss: 9.569193861685136 / Norm of grad_w, grad_b: 0.14573580488204824 , 0.03481555796170279 / Nor
m of weights: 3.797702141820364
Iteration Number: 669 / Loss: 9.56917153860601 / Norm of grad_w, grad_b: 0.14522347893334933 , 0.03481473977344087 / Norm
of weights: 3.7976952713858867
Iteration Number: 670 / Loss: 9.56914899629302 / Norm of grad_w, grad_b: 0.14471275243149265 , 0.03481392247993175 / Norm
of weights: 3.7976883441423426
Iteration Number: 671 / Loss: 9.569126235252117 / Norm of grad_w, grad_b: 0.1442036195211211 , 0.034813106078672634 / Nor
m of weights: 3.7976813602198276
Iteration Number: 672 / Loss: 9.569103255987732 / Norm of grad_w, grad_b: 0.14369607438475035 , 0.034812290567170376 / No
rm of weights: 3.79767431974806
Iteration Number: 673 / Loss: 9.569080059002818 / Norm of grad_w, grad_b: 0.1431901112426782 , 0.034811475942941504 / Nor
m of weights: 3.7976672228563793
Iteration Number: 674 / Loss: 9.569056644798835 / Norm of grad_w, grad_b: 0.14268572435269128 , 0.034810662203511984 / No
rm of weights: 3.797660069673753
Iteration Number: 675 / Loss: 9.569033013875769 / Norm of grad_w, grad_b: 0.1421829080096536 , 0.03480984934641752 / Norm
of weights: 3.7976528603287787
Iteration Number: 676 / Loss: 9.569009166732165 / Norm of grad_w, grad_b: 0.1416816565455431 , 0.034809037369203065 / Nor
m of weights: 3.7976455949496843
Iteration Number: 677 / Loss: 9.568985103865106 / Norm of grad_w, grad_b: 0.1411819643289982 , 0.03480822626942308 / Norm
of weights: 3.7976382736643344
Iteration Number: 678 / Loss: 9.568960825770246 / Norm of grad_w, grad_b: 0.14068382576519065 , 0.03480741604464134 / Nor
m of weights: 3.7976308966002317
Iteration Number: 679 / Loss: 9.568936332941819 / Norm of grad_w, grad_b: 0.14018723529550567 , 0.03480660669243094 / Nor
m of weights: 3.7976234638845185
Iteration Number: 680 / Loss: 9.56891162587264 / Norm of grad_w, grad_b: 0.13969218739742081 , 0.03480579821037413 / Norm
of weights: 3.797615975643982
Iteration Number: 681 / Loss: 9.568886705054133 / Norm of grad_w, grad_b: 0.13919867658416563 , 0.034804990596062374 / No
rm of weights: 3.7976084320050534
```

```
Iteration Number: 682 / Loss: 9.568861570976319 / Norm of grad_w, grad_b: 0.1387066974045665 , 0.03480418384709631 / Norm
of weights: 3.7976008330938145
Iteration Number: 683 / Loss: 9.56883622412785 / Norm of grad_w, grad_b: 0.1382162444427834 , 0.03480337796108555 / Norm
 of weights: 3.7975931790359976
Iteration Number: 684 / Loss: 9.568810664996004 / Norm of grad_w, grad_b: 0.13772731231816296 , 0.03480257293564877 / Nor
m of weights: 3.797585469956988
Iteration Number: 685 / Loss: 9.568784894066697 / Norm of grad_w, grad_b: 0.13723989568482162 , 0.034801768768413635 / No
rm of weights: 3.7975777059818294
Iteration Number: 686 / Loss: 9.568758911824506 / Norm of grad_w, grad_b: 0.1367539892317056 , 0.034800965457016664 / Nor
m of weights: 3.7975698872352224
Iteration Number: 687 / Loss: 9.568732718752663 / Norm of grad_w, grad_b: 0.13626958768223388 , 0.03480016299910323 / Nor
m of weights: 3.797562013841528
Iteration Number: 688 / Loss: 9.568706315333069 / Norm of grad_w, grad_b: 0.13578668579397798 , 0.034799361392327546 / No
rm of weights: 3.7975540859247725
Iteration Number: 689 / Loss: 9.568679702046307 / Norm of grad_w, grad_b: 0.13530527835863998 , 0.034798560634352584 / No
rm of weights: 3.7975461036086466
Iteration Number: 690 / Loss: 9.56865287937165 / Norm of grad_w, grad_b: 0.13482536020178412 , 0.03479776072284996 / Norm
of weights: 3.7975380670165113
Iteration Number: 691 / Loss: 9.568625847787077 / Norm of grad_w, grad_b: 0.13434692618259195 , 0.03479696165549995 / Nor
m of weights: 3.797529976271394
Iteration Number: 692 / Loss: 9.56859860776926 / Norm of grad_w, grad_b: 0.1338699711936588 , 0.03479616342999153 / Norm
 of weights: 3.797521831495998
Iteration Number: 693 / Loss: 9.568571159793605 / Norm of grad_w, grad_b: 0.13339449016088203 , 0.0347953660440221 / Norm
of weights: 3.7975136328127
Iteration Number: 694 / Loss: 9.568543504334235 / Norm of grad_w, grad_b: 0.13292047804311122 , 0.03479456949529759 / Nor
m of weights: 3.797505380343553
Iteration Number: 695 / Loss: 9.568515641864005 / Norm of grad_w, grad_b: 0.13244792983209686 , 0.034793773781532485 / No
rm of weights: 3.7974970742102894
Iteration Number: 696 / Loss: 9.568487572854522 / Norm of grad_w, grad_b: 0.13197684055225334 , 0.03479297890044954 / Nor
m of weights: 3.7974887145343237
Iteration Number: 697 / Loss: 9.56845929777614 / Norm of grad_w, grad_b: 0.13150720526039117 , 0.034792184849779945 / Nor
m of weights: 3.7974803014367517
Iteration Number: 698 / Loss: 9.568430817097974 / Norm of grad_w, grad_b: 0.1310390190456487 , 0.03479139162726321 / Norm
of weights: 3.7974718350383556
Iteration Number: 699 / Loss: 9.568402131287908 / Norm of grad_w, grad_b: 0.1305722770291807 , 0.034790599230647035 / Nor
m of weights: 3.7974633154596034
Iteration Number: 700 / Loss: 9.568373240812596 / Norm of grad_w, grad_b: 0.13010697436409904 , 0.034789807657687404 / No
rm of weights: 3.7974547428206527
Iteration Number: 701 / Loss: 9.568344146137486 / Norm of grad_w, grad_b: 0.1296431062351718 , 0.03478901690614845 / Norm
of weights: 3.797446117241352
Iteration Number: 702 / Loss: 9.56831484772681 / Norm of grad_w, grad_b: 0.12918066785874044 , 0.03478822697380245 / Norm
of weights: 3.7974374388412424
Iteration Number: 703 / Loss: 9.568285346043606 / Norm of grad_w, grad_b: 0.12871965448248487 , 0.03478743785842969 / Nor
m of weights: 3.7974287077395603
Iteration Number: 704 / Loss: 9.568255641549712 / Norm of grad_w, grad_b: 0.12826006138522603 , 0.0347866495578186 / Norm
```

```
of weights: 3.797419924055238
Iteration Number: 705 / Loss: 9.568225734705788 / Norm of grad_w, grad_b: 0.1278018838768163 , 0.03478586206976552 / Norm
of weights: 3.7974110879069056
Iteration Number: 706 / Loss: 9.568195625971306 / Norm of grad_w, grad_b: 0.12734511729789263 , 0.034785075392074795 / No
rm of weights: 3.797402199412894
Iteration Number: 707 / Loss: 9.568165315804578 / Norm of grad_w, grad_b: 0.12688975701981337 , 0.034784289522558534 / No
rm of weights: 3.797393258691236
Iteration Number: 708 / Loss: 9.568134804662742 / Norm of grad_w, grad_b: 0.12643579844434616 , 0.03478350445903692 / Nor
m of weights: 3.797384265859667
Iteration Number: 709 / Loss: 9.568104093001786 / Norm of grad_w, grad_b: 0.12598323700361203 , 0.034782720199337765 / No
rm of weights: 3.7973752210356286
Iteration Number: 710 / Loss: 9.568073181276548 / Norm of grad_w, grad_b: 0.12553206815987863 , 0.03478193674129677 / Nor
m of weights: 3.797366124336268
Iteration Number: 711 / Loss: 9.568042069940711 / Norm of grad_w, grad_b: 0.12508228740540617 , 0.03478115408275732 / Nor
m of weights: 3.797356975878443
Iteration Number: 712 / Loss: 9.568010759446839 / Norm of grad_w, grad_b: 0.12463389026230157 , 0.034780372221570446 / No
rm of weights: 3.79734777577872
Iteration Number: 713 / Loss: 9.567979250246351 / Norm of grad_w, grad_b: 0.12418687228226881 , 0.03477959115559489 / Nor
m of weights: 3.797338524153377
Iteration Number: 714 / Loss: 9.567947542789547 / Norm of grad_w, grad_b: 0.12374122904656809 , 0.034778810882697 / Norm
 of weights: 3.7973292211184058
Iteration Number: 715 / Loss: 9.567915637525612 / Norm of grad_w, grad_b: 0.12329695616582458 , 0.034778031400750656 / No
rm of weights: 3.797319866789513
Iteration Number: 716 / Loss: 9.567883534902617 / Norm of grad_w, grad_b: 0.12285404927985172 , 0.03477725270763725 / Nor
m of weights: 3.7973104612821222
Iteration Number: 717 / Loss: 9.56785123536753 / Norm of grad_w, grad_b: 0.1224125040574848 , 0.03477647480124568 / Norm
 of weights: 3.797301004711374
Iteration Number: 718 / Loss: 9.567818739366214 / Norm of grad_w, grad_b: 0.12197231619649145 , 0.034775697679472285 / No
rm of weights: 3.7972914971921288
Iteration Number: 719 / Loss: 9.567786047343445 / Norm of grad_w, grad_b: 0.12153348142336244 , 0.034774921340220866 / No
rm of weights: 3.797281938838968
Iteration Number: 720 / Loss: 9.567753159742912 / Norm of grad_w, grad_b: 0.12109599549323673 , 0.03477414578140242 / Nor
m of weights: 3.7972723297661948
Iteration Number: 721 / Loss: 9.567720077007218 / Norm of grad_w, grad_b: 0.120659854189613 , 0.03477337100093546 / Norm
 of weights: 3.7972626700878367
Iteration Number: 722 / Loss: 9.567686799577894 / Norm of grad_w, grad_b: 0.12022505332443902 , 0.03477259699674564 / Nor
m of weights: 3.7972529599176466
Iteration Number: 723 / Loss: 9.567653327895403 / Norm of grad_w, grad_b: 0.11979158873771303 , 0.03477182376676595 / Nor
m of weights: 3.7972431993691025
Iteration Number: 724 / Loss: 9.567619662399133 / Norm of grad_w, grad_b: 0.11935945629751352 , 0.034771051308936575 / No
rm of weights: 3.7972333885554117
Iteration Number: 725 / Loss: 9.567585803527427 / Norm of grad_w, grad_b: 0.11892865189985419 , 0.034770279621204804 / No
rm of weights: 3.7972235275895105
Iteration Number: 726 / Loss: 9.567551751717566 / Norm of grad_w, grad_b: 0.11849917146847204 , 0.03476950870152512 / Nor
m of weights: 3.797213616584065
```

```
Iteration Number: 727 / Loss: 9.56751750740578 / Norm of grad_w, grad_b: 0.11807101095469733 , 0.03476873854785913 / Norm
of weights: 3.797203655651474
Iteration Number: 728 / Loss: 9.567483071027269 / Norm of grad_w, grad_b: 0.11764416633742962 , 0.034767969158175406 / No
rm of weights: 3.797193644903869
Iteration Number: 729 / Loss: 9.567448443016179 / Norm of grad_w, grad_b: 0.1172186336228404 , 0.03476720053044966 / Norm
of weights: 3.797183584453117
Iteration Number: 730 / Loss: 9.567413623805635 / Norm of grad_w, grad_b: 0.11679440884439245 , 0.03476643266266446 / Nor
m of weights: 3.797173474410819
Iteration Number: 731 / Loss: 9.56737861382773 / Norm of grad_w, grad_b: 0.11637148806270652 , 0.034765665552809404 / Nor
m of weights: 3.797163314888314
Iteration Number: 732 / Loss: 9.567343413513534 / Norm of grad_w, grad_b: 0.11594986736523623 , 0.03476489919888104 / Nor
m of weights: 3.7971531059966788
Iteration Number: 733 / Loss: 9.567308023293098 / Norm of grad_w, grad_b: 0.11552954286639831 , 0.03476413359888267 / Nor
m of weights: 3.7971428478467293
Iteration Number: 734 / Loss: 9.567272443595463 / Norm of grad_w, grad_b: 0.1151105107073535 , 0.03476336875082454 / Norm
of weights: 3.7971325405490237
Iteration Number: 735 / Loss: 9.567236674848663 / Norm of grad_w, grad_b: 0.11469276705577286 , 0.034762604652723676 / No
rm of weights: 3.7971221842138596
Iteration Number: 736 / Loss: 9.567200717479722 / Norm of grad_w, grad_b: 0.11427630810593124 , 0.03476184130260384 / Nor
m of weights: 3.797111778951279
Iteration Number: 737 / Loss: 9.567164571914674 / Norm of grad_w, grad_b: 0.11386113007837192 , 0.03476107869849565 / Nor
m of weights: 3.797101324871068
Iteration Number: 738 / Loss: 9.567128238578551 / Norm of grad_w, grad_b: 0.11344722922004927 , 0.034760316838436243 / No
rm of weights: 3.797090822082756
Iteration Number: 739 / Loss: 9.567091717895394 / Norm of grad_w, grad_b: 0.11303460180386078 , 0.03475955572046961 / Nor
m of weights: 3.7970802706956217
Iteration Number: 740 / Loss: 9.567055010288271 / Norm of grad_w, grad_b: 0.11262324412889309 , 0.03475879534264621 / Nor
m of weights: 3.7970696708186895
Iteration Number: 741 / Loss: 9.567018116179261 / Norm of grad_w, grad_b: 0.11221315252005876 , 0.03475803570302327 / Nor
m of weights: 3.797059022560732
Iteration Number: 742 / Loss: 9.56698103598946 / Norm of grad_w, grad_b: 0.11180432332814635 , 0.03475727679966444 / Norm
of weights: 3.7970483260302723
Iteration Number: 743 / Loss: 9.566943770139003 / Norm of grad_w, grad_b: 0.11139675292961343 , 0.034756518630640006 / No
rm of weights: 3.7970375813355837
Iteration Number: 744 / Loss: 9.566906319047051 / Norm of grad_w, grad_b: 0.11099043772647167 , 0.03475576119402671 / Nor
m of weights: 3.7970267885846907
Iteration Number: 745 / Loss: 9.566868683131801 / Norm of grad_w, grad_b: 0.11058537414625777 , 0.03475500448790778 / Nor
m of weights: 3.797015947885373
Iteration Number: 746 / Loss: 9.5668308628105 / Norm of grad_w, grad_b: 0.11018155864193993 , 0.03475424851037285 / Norm
 of weights: 3.7970050593451607
Iteration Number: 747 / Loss: 9.566792858499424 / Norm of grad_w, grad_b: 0.10977898769166307 , 0.03475349325951803 / Nor
m of weights: 3.7969941230713404
Iteration Number: 748 / Loss: 9.566754670613905 / Norm of grad_w, grad_b: 0.10937765779882973 , 0.03475273873344572 / Nor
m of weights: 3.796983139170955
Iteration Number: 749 / Loss: 9.56671629956833 / Norm of grad_w, grad_b: 0.10897756549188904 , 0.03475198493026475 / Norm
```

```
of weights: 3.7969721077508036
Iteration Number: 750 / Loss: 9.566677745776143 / Norm of grad_w, grad_b: 0.10857870732425726 , 0.03475123184809019 / Nor
m of weights: 3.7969610289174427
Iteration Number: 751 / Loss: 9.56663900964984 / Norm of grad_w, grad_b: 0.10818107987433531 , 0.0347504794850434 / Norm
 of weights: 3.796949902777188
Iteration Number: 752 / Loss: 9.56660009160099 / Norm of grad_w, grad_b: 0.10778467974516732 , 0.03474972783925204 / Norm
of weights: 3.7969387294361145
Iteration Number: 753 / Loss: 9.566560992040223 / Norm of grad_w, grad_b: 0.10738950356456786 , 0.03474897690884995 / Nor
m of weights: 3.796927509000059
Iteration Number: 754 / Loss: 9.566521711377248 / Norm of grad_w, grad_b: 0.10699554798498713 , 0.034748226691977135 / No
rm of weights: 3.7969162415746185
Iteration Number: 755 / Loss: 9.566482250020844 / Norm of grad_w, grad_b: 0.10660280968333165 , 0.03474747718677982 / Nor
m of weights: 3.7969049272651536
Iteration Number: 756 / Loss: 9.566442608378871 / Norm of grad_w, grad_b: 0.10621128536096752 , 0.034746728391410284 / No
rm of weights: 3.7968935661767866
Iteration Number: 757 / Loss: 9.56640278685827 / Norm of grad_w, grad_b: 0.10582097174355198 , 0.034745980304026994 / Nor
m of weights: 3.7968821584144066
Iteration Number: 758 / Loss: 9.566362785865072 / Norm of grad_w, grad_b: 0.10543186558107316 , 0.03474523292279437 / Nor
m of weights: 3.7968707040826657
Iteration Number: 759 / Loss: 9.566322605804391 / Norm of grad_w, grad_b: 0.10504396364757505 , 0.034744486245883006 / No
rm of weights: 3.796859203285984
Iteration Number: 760 / Loss: 9.566282247080446 / Norm of grad_w, grad_b: 0.10465726274128967 , 0.03474374027146937 / Nor
m of weights: 3.7968476561285467
Iteration Number: 761 / Loss: 9.566241710096538 / Norm of grad_w, grad_b: 0.10427175968439883 , 0.034742994997736046 / No
rm of weights: 3.7968360627143083
Iteration Number: 762 / Loss: 9.566200995255079 / Norm of grad_w, grad_b: 0.10388745132290174 , 0.03474225042287143 / Nor
m of weights: 3.796824423146991
Iteration Number: 763 / Loss: 9.56616010295758 / Norm of grad_w, grad_b: 0.10350433452682276 , 0.03474150654506996 / Norm
of weights: 3.796812737530088
Iteration Number: 764 / Loss: 9.566119033604664 / Norm of grad_w, grad_b: 0.1031224061898266 , 0.03474076336253191 / Norm
of weights: 3.7968010059668615
Iteration Number: 765 / Loss: 9.56607778759606 / Norm of grad_w, grad_b: 0.10274166322926662 , 0.034740020873463416 / Nor
m of weights: 3.7967892285603453
Iteration Number: 766 / Loss: 9.566036365330607 / Norm of grad_w, grad_b: 0.10236210258615103 , 0.03473927907607645 / Nor
m of weights: 3.7967774054133447
Iteration Number: 767 / Loss: 9.565994767206268 / Norm of grad_w, grad_b: 0.10198372122492896 , 0.034738537968588855 / No
rm of weights: 3.79676553662844
Iteration Number: 768 / Loss: 9.565952993620124 / Norm of grad_w, grad_b: 0.10160651613357725 , 0.03473779754922423 / Nor
m of weights: 3.796753622307982
Iteration Number: 769 / Loss: 9.565911044968374 / Norm of grad_w, grad_b: 0.10123048432348716 , 0.03473705781621182 / Nor
m of weights: 3.7967416625540986
Iteration Number: 770 / Loss: 9.565868921646352 / Norm of grad_w, grad_b: 0.10085562282924494 , 0.03473631876778681 / Nor
m of weights: 3.796729657468692
Iteration Number: 771 / Loss: 9.565826624048514 / Norm of grad_w, grad_b: 0.10048192870880544 , 0.03473558040218991 / Nor
m of weights: 3.79671760715344
```

```
Iteration Number: 772 / Loss: 9.565784152568453 / Norm of grad_w, grad_b: 0.10010939904324243 , 0.034734842717667525 / No
rm of weights: 3.796705511709798
Iteration Number: 773 / Loss: 9.565741507598895 / Norm of grad_w, grad_b: 0.09973803093670104 , 0.03473410571247181 / Nor
m of weights: 3.7966933712389976
```

In [32]:

```
print("Train accuracy: ", accuracy(xtrain_normal, ytrain, model))
```

```
Train accuracy:  0.9253846153846154
```

In [33]:

```
#grid search for finding the best hyperparams and model

best_model = None
best_val = -1
for lr in [0.01, 0.001, 0.0001, 0.00001]:
    for la in [5, 2, 1, 0.1, 0.01]:
        model = fit(xtrain_normal, ytrain, lr, la, 10000, verbose=0)
        val_acc = accuracy(xval_normal, yval, model)
        print(lr, la, val_acc)
        if val_acc > best_val:
            best_val = val_acc
            best_model = model
```

```
0.01 5 0.9310689310689311
0.01 2 0.8771228771228772
0.01 1 0.9170829170829171
0.01 0.1 0.8941058941058941
0.01 0.01 0.9020979020979021
0.001 5 0.926073926073926
0.001 2 0.9330669330669331
0.001 1 0.9270729270729271
0.001 0.1 0.9250749250749251
0.001 0.01 0.919080919080919
0.0001 5 0.9230769230769231
0.0001 2 0.9340659340659341
0.0001 1 0.929070929070929
0.0001 0.1 0.9110889110889111
0.0001 0.01 0.9300699300699301
1e-05 5 0.9140859140859141
1e-05 2 0.9090909090909091
1e-05 1 0.9100899100899101
```

```
1e-05 0.1 0.926073926073926
1e-05 0.01 0.9230769230769231
```

In [34]:

```
print("Test accuracy: ", accuracy(xtest_normal, ytest, best_model))
```

```
Test accuracy:  0.943
```