

In [1]:

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

Data Preprocessing

1. Tried Normalization - Both Min-Max and Gaussian Normalization i tried but they gave worse results.

In [2]:

```
from mlxtend.data import loadlocal_mnist
```

In [3]:

```
test_X, test_Y=loadlocal_mnist(images_path='dataset/t10k-images.idx3-ubyte', labels_path='dataset/t10k-labels.idx1-ubyte')
train_X, train_Y=loadlocal_mnist(images_path='dataset/train-images.idx3-ubyte', labels_path='dataset/train-labels.idx1-ubyte')
```

Data dividing into test , train , validation

In [4]:

```
import pandas as pd
import numpy as np
total_data_X=[]
total_data_Y=[]
for i in range(len(train_Y)):
    mylist=list(train_X[i].ravel())
    total_data_X.append(mylist)
    total_data_Y.append(train_Y[i])
for i in range(len(test_Y)):
    mylist=list(test_X[i].ravel())
    total_data_X.append(mylist)
    total_data_Y.append(test_Y[i])

total_data_X=np.array(total_data_X)
#total_data_X = total_data_X/255
total_data_Y=np.array(total_data_Y)

def mytraintestvalsplit(total, valsize, testsize):
    notestrows=int(testsize*total.shape[0])
    novalrows=int(valsize*total.shape[0])
    notrainrows=total.shape[0]-notestrows-novalrows
    trainrows=total[:notrainrows]
    valrows=total[notrainrows:notrainrows+novalrows]
    testrows=total[notrainrows+novalrows:]
    return trainrows, valrows, testrows
```

```
X_train, X_val, X_test = mytraintestvalsplit(total_data_X, 0.2, 0.1)
Y_train, Y_val, Y_test = mytraintestvalsplit(total_data_Y, 0.2, 0.1)
```

In [5]:

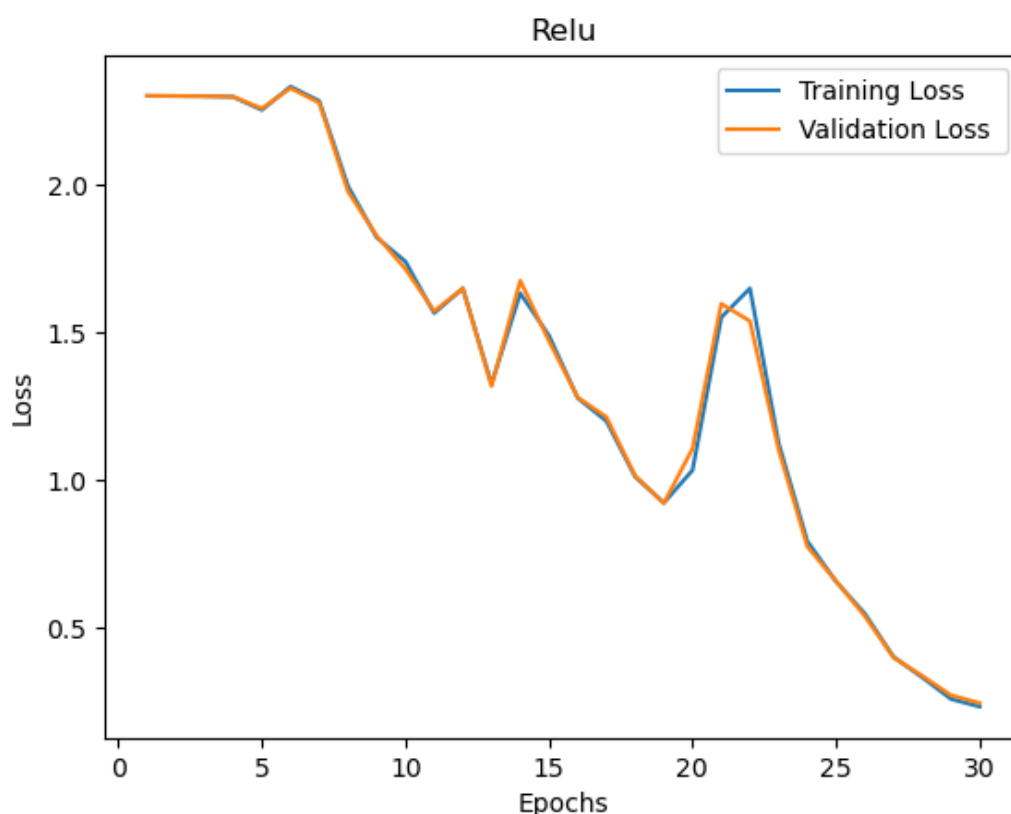
```
nn = neural_networks.NN(6, [784, 256, 128, 64, 32, 10], 'relu', 0.1, 'normal', len(X_train)//20, 30)
nn.fit(X_train, Y_train, X_val, Y_val)
print()
print("The Accuracy Score for Relu Activation is with normal Initialization is")
print(nn.score(X_test, Y_test))
```

Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch: 10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch: 19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch: 28 Epoch: 29 Epoch: 30

The Accuracy Score for Relu Activation is with normal Initialization is 0.9478571428571428

In [7]:

```
alltheweights={}
alltheweights['relu'] = nn.params
plt.plot(list(range(1,len(nn.train_losses) + 1)),nn.train_losses, label = "Training Loss " )
plt.plot(list(range(1,len(nn.val_losses) + 1)),nn.val_losses, label = "Validation Loss " )
plt.ylabel('Loss')
plt.legend()
plt.xlabel('Epochs')
plt.title("Relu")
plt.show()
```



In [10]:

```
nn = neural_networks.NN(6, [784, 256, 128, 64, 32, 10], 'relu', 0.08, 'normal', len(X_train)//20, 30)
nn.fit(X_train,Y_train,X_val,Y_val)
print()
print("The Accuracy Score for Relu Activation is with normal Initialization is")
print(nn.score(X_test,Y_test))
```

Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch: 10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch: 19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch: 28 Epoch: 29 Epoch: 30

The Accuracy Score for Relu Activation is with normal Initialization is 0.861

In [11]:

```
alltheweights={}
alltheweights['relu'] = nn.params
print(alltheweights['relu'])
```

{'b1': array([[-1.56893561e-05, -1.15825096e-05, -2.41079062e-05,

-5.38508623e-05, -6.55510260e-05, -7.59930626e-05,
-2.50314360e-04, -1.19548163e-04, -3.41853136e-05,
-4.53058062e-05, 8.32991683e-06, -5.26804144e-05,
-4.44882097e-06, 1.05881104e-04, -1.23165361e-05,
5.50849820e-05, -1.82055182e-05, -2.81744252e-04,
4.71273482e-05, -1.29683658e-04, -7.48916218e-05,
-4.29451927e-05, -3.24913640e-05, 2.72072251e-05,
1.17727397e-06, -7.56800903e-05, -3.59553035e-05,
-7.12604812e-05, -1.01329678e-04, -1.45981161e-04,
-5.11241875e-05, -2.24709657e-05, -1.91640003e-04,
-4.28621583e-05, -2.67151076e-05, -2.22046735e-05,
-9.02026544e-05, -2.80241507e-04, -1.26862371e-04,
1.26032816e-05, -1.45255299e-05, -7.39325402e-06,
-2.37711693e-05, -1.08855271e-04, -2.03003393e-05,
-4.09165193e-05, -8.09600711e-05, -4.81825597e-05,
-1.29319411e-04, -5.96794602e-05, -1.40093906e-04,
-7.13815819e-06, -7.07279069e-05, -1.46439020e-05,
9.08656695e-05, -2.20740372e-04, -4.81572265e-05,
5.94142101e-05, -1.61388913e-05, 1.46960884e-04,
-1.39662985e-06, -3.97811877e-05, -2.17030301e-05,
-1.89554576e-04, -9.27833925e-06, -2.82188723e-05,
-5.79269018e-05, -5.92132932e-06, -2.00226423e-05,
-7.33331645e-06, -2.77076515e-05, -1.94771734e-05,
5.51145809e-05, -1.25403200e-05, 5.58187584e-05,
-4.06144367e-05, -6.91980620e-05, -2.39443231e-04,
1.06448746e-05, -2.65552966e-05, -1.51349160e-04,
-1.54161282e-04, -2.99765077e-05, 3.59906803e-06,
-3.32349744e-05, -2.91948845e-05, -6.62268743e-05,
-2.00623643e-05, -2.63067474e-05, -7.97286041e-05,
-1.26915496e-04, -2.73473156e-05, -5.84118618e-05,
4.44483640e-05, -2.29673238e-05, -1.37439464e-04,
-9.68321985e-05, -2.57001713e-05, -5.64954344e-05,
-7.70328432e-05, -7.67803695e-05, 4.55288172e-05,
-1.28459921e-06, -3.98712656e-06, 3.78547982e-05,
2.76473000e-05, -2.22665152e-05, -6.58138177e-06,
1.45597616e-05, -1.22268515e-05, -1.38827276e-05,
5.10883763e-06, -3.37581732e-05, -1.60902446e-04,
-1.74749267e-05, -1.13947324e-04, -2.67987982e-05,
-1.74853708e-04, -1.04200068e-04, -6.92347055e-05,
-5.73331557e-05, -1.74133593e-04, -1.91640193e-05,
2.47774783e-05, 8.33816867e-05, -9.31739532e-06,
3.31961803e-05, -7.68342219e-05, 8.31329087e-05,
2.61698543e-05, -4.41841443e-05, 1.21236345e-04,
-2.42352782e-05, -3.17710120e-05, -4.19532446e-06,
-4.80081167e-05, -5.04641041e-05, -4.36390910e-05,
-6.60673272e-06, -6.17007794e-05, -1.97820008e-04,
-3.04709354e-05, -6.13040121e-05, -5.97636770e-05,
-5.37413747e-05, -4.20595560e-05, 6.38977485e-05,
2.93165160e-05, 2.95373449e-05, -2.52865256e-05,
3.11412289e-05, 5.96696222e-05, -2.96463941e-05,
-6.92573063e-05, 2.65177855e-06, -3.66739941e-05,
-2.11655471e-05, -1.03831744e-05, -2.37959755e-05,
-4.94934894e-05, -5.79896304e-06, -4.41005259e-05,
-1.80239316e-05, -2.34326189e-05, -7.85330103e-05,
-6.32021992e-06, -7.63058954e-05, 9.00693743e-06,
-8.23958601e-06, 1.70840079e-05, -3.46450723e-04,
-5.52887692e-05, -1.75010213e-04, -1.26917078e-05,
-1.10639961e-04, -9.26123338e-05, 1.50967159e-05,
-9.54607686e-05, -6.49236235e-05, -8.69781428e-05,
-7.85544187e-05, 2.21059433e-06, -1.16480297e-06,
-4.41602977e-06, 2.40879520e-05, 1.26016654e-04,
-6.26061298e-05, -6.07311123e-06, -3.82809107e-05,
-9.80444813e-06, 4.68035987e-06, -2.90168329e-05,
3.16670212e-05, -1.35144013e-04, 7.67518762e-07,
-1.10122925e-04, -3.26076088e-05, 1.15543113e-05,
-6.64387234e-05, -1.89130295e-05, 3.04433659e-05,
-3.95012927e-05, -5.32919314e-05, -5.40845487e-05,
-5.50839904e-06, -4.22319430e-05, -4.81487105e-04,
-3.87742791e-05, -2.46678584e-05, 8.70072244e-06,
-6.62673428e-05, -7.80957676e-05, 1.01935371e-05,
-7.47391995e-05, 2.83714691e-05, -2.80044597e-05,
2.55458962e-06, -4.19029951e-05, -1.27346387e-04,

```
-6.93933311e-05, 1.82453792e-05, -2.81833358e-05,
-3.32677562e-05, -6.98782150e-05, -5.90864909e-05,
-2.24472422e-05, -7.49442736e-05, -3.24260685e-05,
-3.37444353e-06, -5.01835934e-05, 6.49134983e-06,
-1.36711471e-04, 3.83447022e-06, -6.37357439e-05,
-2.18119285e-05, -1.11354017e-04, -6.38918837e-05,
-2.48796244e-05, 7.50856377e-07, -3.13290380e-04,
-7.89119093e-06, -6.50053088e-05, -5.73647217e-05,
8.10722881e-05, 3.17377038e-05, -8.71135514e-05,
-6.65184897e-06, -7.59010354e-05, 3.70518267e-07,
-4.12159894e-05, -3.68099281e-05, -7.12006019e-06,
-7.89378461e-05, -6.92268159e-07, -8.46232424e-05,
-3.49453355e-05]]), 'W1': array([[ 1.79871217e-02, -2.02158083e-02, -1.09255743e-
02, ...,
8.77325459e-03, 4.55480214e-03, -2.19666327e-02],
[-1.26173855e-02, 1.05803844e-02, -1.11890006e-02, ...,
4.83699567e-03, -4.09598671e-03, 6.34268878e-03],
[-1.35816589e-03, 6.11750853e-05, -1.24064994e-02, ...,
2.34074613e-02, 2.32774258e-03, -8.30708319e-04],
...,
[-6.89518439e-03, -8.31815863e-03, -4.60122961e-03, ...,
-4.51767514e-03, 1.39846346e-02, -2.25045790e-03],
[-1.51260714e-02, -1.25835360e-03, -1.05992910e-03, ...,
4.34064240e-04, -1.68105760e-02, 6.49930013e-03],
[-1.82308194e-02, -3.70202106e-03, 1.69240987e-02, ...,
-9.33062949e-03, 1.12590515e-02, 1.27451648e-02]]), 'b2': array([[ 5.68870676e-
04, 3.51042877e-03, -3.23321582e-04,
-3.04370778e-04, 2.47809138e-04, -1.87266337e-03,
2.96237277e-04, 8.00040092e-04, -5.74079613e-05,
-4.61762900e-05, 7.41837457e-04, -3.22461867e-04,
-2.35188944e-04, 9.21397064e-04, -2.38448410e-04,
-1.07121584e-03, -1.73383073e-03, 4.71531315e-05,
-1.17043323e-04, -1.98971522e-04, 1.19201897e-04,
-3.78706939e-05, -1.05473755e-04, 2.17569691e-04,
2.96148061e-04, -3.64563809e-04, 4.38713092e-04,
1.12607849e-04, 3.53330162e-05, 4.69715089e-03,
4.68933537e-03, 3.83933841e-03, -2.99084909e-05,
1.13538781e-04, 2.32666731e-03, 1.80341841e-03,
2.99324565e-03, 4.05335701e-04, 9.99641775e-05,
-1.58153355e-04, 5.23408459e-04, 7.35589667e-04,
4.81547330e-04, 1.85537931e-03, -7.53017382e-05,
1.74070736e-03, -1.94373198e-04, -1.86343295e-04,
-9.13465266e-04, 2.78274356e-03, -1.64940933e-05,
5.42694385e-04, -4.00683303e-04, -1.08811472e-03,
2.37957835e-04, 2.91753395e-04, 3.06560658e-04,
-1.71078002e-04, 4.67188712e-03, 8.20977022e-04,
1.51160143e-04, -1.47821922e-05, -1.47183833e-04,
-4.39482337e-04, 6.25150100e-04, -3.96736052e-04,
-8.59316854e-04, 3.74720173e-04, -1.65802848e-03,
4.00950183e-03, 7.06159565e-04, 1.60466561e-03,
2.88391436e-04, 5.57033121e-04, 1.35317598e-03,
-3.66767218e-04, 4.97780445e-04, -3.31068481e-04,
1.11611695e-04, -7.44889163e-04, -3.04431651e-04,
-7.63916262e-04, -6.35834370e-05, -1.73440402e-03,
2.99353205e-04, 5.26338921e-05, 4.91903113e-05,
1.06649060e-03, -2.98818507e-04, 5.32963947e-04,
-5.57493013e-04, 3.30283817e-03, -3.03670712e-05,
-1.84810579e-04, -2.76279244e-05, -3.32560156e-04,
2.85509764e-03, 1.04765509e-04, 2.16337743e-04,
-1.69500420e-04, -4.21978722e-04, -4.32249656e-04,
8.99253696e-04, -2.58711404e-04, 2.01210832e-03,
9.61568715e-05, -4.43478037e-04, 6.47921172e-04,
1.30078236e-04, 8.82412727e-04, -9.35579433e-05,
-7.21158213e-04, 2.65747191e-05, 1.83731738e-03,
1.05879801e-04, 2.78902896e-03, -7.34605779e-04,
-6.32102878e-04, 1.20696997e-04, 2.75313458e-04,
4.28306976e-04, -3.34429640e-04, -4.26439756e-04,
1.13343585e-04, 4.04707983e-04, -2.14981368e-04,
2.31210300e-03, 9.66723749e-04]]), 'W2': array([[ 7.55531451e-03, 2.04818866e-
02, 5.76498569e-03, ...,
-8.91068684e-03, 1.29692794e-02, -8.80525126e-03],
[ 2.15592418e-02, 1.92209258e-03, -5.08897453e-03, ...,
```

```
8.23231201e-03, 2.82733674e-04, 5.96333560e-03],
[ 9.56522469e-03, -6.68269455e-03, -4.64271477e-03, ...,
 1.32290268e-02, 7.89802897e-04, 4.79259836e-03],
...,
[ 8.20507311e-03, -2.68594775e-03, 7.80483188e-03, ...,
 1.41629691e-02, 7.88267236e-03, -1.82412229e-02],
[ 5.39915512e-03, 3.43729926e-02, -4.39446608e-05, ...,
 5.28951219e-03, 3.15678489e-03, 1.30147577e-03],
[-1.19672990e-02, 4.24164259e-02, -1.28415044e-02, ...,
 -6.70120269e-03, 7.90548612e-03, 1.04127185e-02]], 'b3': array([[ 2.09798731e-
02, 8.68912624e-03, 1.96805042e-03,
 8.66077787e-04, 3.61083881e-03, 3.88917378e-03,
 3.20821287e-04, 3.93839279e-05, 7.41297241e-03,
 3.01735542e-03, 1.66292046e-03, 4.50570770e-03,
 -3.30922379e-04, 6.94693344e-03, 5.33510180e-04,
 -2.66416534e-04, 4.02203065e-03, 2.98592206e-04,
 -4.33459548e-05, 2.55226691e-03, 3.30277442e-03,
 1.45191646e-04, 1.64417069e-03, 1.33837040e-03,
 6.47439891e-06, 2.84252899e-03, -2.75849210e-04,
 1.88243567e-02, 2.34670832e-04, -5.44647458e-05,
 6.36222370e-03, -1.45740242e-03, 5.48417327e-03,
 2.20968710e-03, -2.13337941e-03, 2.01338835e-04,
 2.72093102e-03, 1.32432635e-03, 2.89296444e-03,
 -2.43353202e-05, 6.76961695e-04, -1.00824084e-03,
 -1.79243154e-04, -2.46716348e-04, 8.20509987e-04,
 9.93120278e-03, 6.06476427e-03, -2.62392130e-03,
 -1.50191490e-04, 2.55720176e-03, 3.41812207e-04,
 1.47992716e-04, 1.15744150e-03, 3.08787809e-02,
 2.20493545e-03, 2.81791081e-03, -9.67448836e-04,
 1.81872116e-03, 2.35665684e-03, 2.45486317e-03,
 -2.29210713e-04, 3.44589440e-03, -1.28373947e-04,
 2.85750116e-04])), 'W3': array([[ 3.79131188e-02, 1.95166056e-02, -1.63043256e-
02, ...,
 -3.03309507e-02, 4.20152521e-03, 4.64259407e-03],
[ 7.38739393e-02, 5.90232922e-02, -1.52766290e-02, ...,
 1.58663016e-02, -1.84412773e-03, -2.52940592e-02],
[-3.10801282e-03, -9.28271462e-03, -9.44903275e-03, ...,
 1.89548109e-02, 1.04395591e-02, -2.02213690e-02],
...,
[-2.94252312e-02, -9.22543069e-05, 7.23031357e-03, ...,
 -1.54152154e-02, 1.43298475e-03, 7.98928459e-03],
[ 1.09387604e-02, 2.28881609e-02, 4.94315878e-03, ...,
 -1.02069298e-02, -3.38003191e-03, 9.17740998e-03],
[ 7.50872195e-03, 7.46630195e-03, 1.80394264e-02, ...,
 1.87456248e-02, -1.63174893e-03, 1.24098117e-02])), 'b4': array([[ 0.00548796,
0.01477025, 0.02083374, 0.0139627, 0.00315365,
 0.02887193, -0.0045557, 0.02416423, 0.00748942, 0.02981625,
 0.01255559, -0.00966868, -0.00262511, 0.00665494, -0.00073229,
 0.02430002, 0.05986266, 0.00120553, -0.00100005, 0.01580725,
 0.00330991, 0.00924181, -0.00422713, 0.01810802, 0.00652873,
 -0.00114219, -0.00124239, -0.00705189, 0.03070448, -0.00341524,
 -0.00096566, 0.04914375])), 'W4': array([[ 0.01409198, -0.02812701, 0.03615941,
..., -0.0056535,
 -0.00693582, 0.13728497],
[-0.01402432, -0.01424383, -0.02158273, ..., 0.0009527,
 -0.00088423, 0.05324965],
[-0.01817492, -0.01918928, 0.01583974, ..., -0.00802284,
 -0.00412507, -0.01734023],
...,
[ 0.03461276, -0.01452504, 0.05288129, ..., -0.00195369,
 0.01533162, 0.03091855],
[ 0.01599604, 0.00845709, 0.00592998, ..., -0.0011993,
 0.01451499, -0.00893704],
[ 0.01833603, 0.00552378, 0.00129124, ..., 0.001298,
 0.00840168, 0.0193429 ]]), 'b5': array([[ -0.11126383, -0.5424993, -0.74163996,
-0.38414986, 0.92201955,
 -0.20393495, -0.53157191, 0.57984273, 0.09666943, 0.91652811]]), 'W5': array([
[-1.71175965e-01, 3.44421714e-01, 1.19893135e-01,
 -2.83806241e-01, 5.08359706e-02, -1.37021212e-01,
 2.85590508e-01, -1.55662227e-01, 1.33816220e-01,
 -2.17129388e-01],
[ 2.84333976e-01, -9.71553787e-02, -1.45158742e-01,
```

-1.42792032e-01, -7.04403861e-02, 1.73692884e-01,
1.51711365e-01, -8.22194805e-02, -1.99049054e-02,
-4.24887441e-02],
[-1.75962975e-01, 1.05666779e-01, 2.81282487e-01,
-4.95907939e-02, 9.29067770e-02, -5.36364705e-02,
2.93041905e-01, -2.94781928e-01, 3.58567486e-02,
-2.23536223e-01],
[-5.91023477e-02, -2.40142336e-02, 1.24816992e-02,
3.81554947e-02, -2.45498468e-02, 1.25738320e-02,
-4.15159186e-02, -9.18414263e-03, 5.92056509e-02,
4.18932813e-02],
[4.73247306e-02, -3.63014178e-02, 1.90727589e-03,
3.61501783e-02, 5.60809554e-02, 3.27441591e-02,
1.76468046e-03, -4.80939636e-02, 1.99172727e-02,
-7.22984742e-02],
[-1.75846826e-03, 3.29433851e-01, -1.40319754e-01,
-1.22754944e-01, -5.86494837e-02, -8.67387773e-02,
-2.27444272e-01, 4.68239600e-02, 2.29932907e-01,
-1.62085363e-02],
[-3.98181241e-03, -9.67466766e-03, -3.01916500e-02,
-9.39584057e-04, -3.84928055e-03, 3.59155241e-04,
-2.01719532e-02, 1.58856223e-02, 9.73766357e-04,
-1.26640988e-03],
[-5.63795396e-02, -1.11988789e-01, -8.30562547e-02,
-1.74580834e-01, 1.54670116e-01, -1.53811971e-01,
-1.17954653e-01, 3.78983894e-01, -1.49280061e-01,
2.57698572e-01],
[4.51664563e-02, -3.23812401e-02, 1.13457388e-02,
2.56619031e-02, 7.00721589e-02, 3.91037737e-02,
-6.36613323e-04, -5.38512439e-02, 4.28316747e-02,
-9.14341656e-02],
[-5.58199036e-02, -1.33534726e-01, 8.23530519e-02,
-1.61185014e-02, 2.79260183e-01, -7.65834806e-03,
2.14637219e-02, -1.43904865e-01, -8.80656033e-02,
1.93630842e-02],
[-7.73657252e-02, -9.90546775e-03, -6.88023761e-02,
7.30961288e-02, 1.00118015e-02, 6.72600598e-02,
-6.05193712e-03, -9.00438075e-03, 2.91882868e-02,
4.07414497e-02],
[-1.04868309e-03, -3.87452409e-03, 6.76071887e-03,
-9.45890435e-03, -9.23557352e-03, 4.90569626e-03,
3.21725319e-03, -8.39099232e-03, 7.65898045e-03,
-1.08663074e-02],
[5.16269088e-02, -2.07525991e-02, 3.47874609e-03,
-3.69296388e-02, -1.54678787e-03, -2.22030376e-02,
4.68087987e-02, 8.59891516e-05, -1.25582776e-02,
-1.58354932e-02],
[1.77853030e-01, -5.87391599e-02, -8.27745650e-02,
-8.01899628e-02, -4.83919578e-02, 1.14806317e-01,
5.69230594e-02, -4.63587719e-02, -7.27353445e-03,
-4.66363075e-02],
[2.85806091e-02, -3.59330490e-02, 2.01906321e-03,
2.23451802e-02, 4.26773600e-02, 2.51228610e-02,
-1.71066029e-02, -6.92070694e-02, 3.14099271e-02,
-6.58211115e-02],
[-1.46846246e-01, -7.46934959e-02, -4.68594939e-02,
-7.48835887e-02, 2.07160685e-01, -1.59634355e-01,
-1.53371676e-01, 3.13648351e-01, -1.16503715e-01,
2.34169481e-01],
[-2.15392552e-01, -2.66311059e-01, 1.74923626e-01,
3.76478576e-01, -1.34970371e-01, 2.12147344e-01,
-2.74018833e-01, -1.62340451e-01, 1.82559334e-01,
1.58828081e-01],
[8.82448733e-02, -5.97818298e-02, -2.79570904e-02,
3.80553437e-02, 9.31379615e-02, 5.90431880e-02,
-1.11363513e-02, -1.58222174e-01, 1.04431259e-01,
-1.37163669e-01],
[-5.82197974e-03, 6.91591789e-03, -2.06493740e-03,
1.75488103e-02, 3.60146313e-03, 1.42731702e-02,
6.33031903e-03, 2.74878937e-03, -2.09758072e-02,
9.21475366e-03],
[2.99408501e-01, -1.29014199e-01, -1.29558409e-01,

```

-8.20585701e-02, -1.10238812e-01, 1.80495557e-01,
7.24516483e-02, -9.88038509e-02, 5.62188147e-03,
-2.86526912e-02],
[-1.06752848e-02, -8.62038297e-03, -3.86176409e-03,
1.35646649e-02, 2.53432206e-02, -1.26539793e-02,
5.44706898e-03, -1.72820630e-02, -9.70815634e-03,
7.97053325e-03],
[-4.65263540e-02, -5.34553513e-03, -4.74180421e-02,
3.30568253e-02, -4.25525542e-04, 6.52228234e-03,
-9.84033767e-03, 7.71237494e-03, 4.70403755e-02,
1.92046968e-02],
[ 1.14629875e-01, -2.23332471e-02, -2.61691222e-02,
-3.75737798e-02, -1.80764595e-02, 3.39239926e-02,
5.49214847e-02, -2.12692330e-03, 1.58940446e-02,
-2.98590912e-02],
[-6.97430444e-02, -1.87584559e-02, 7.16018096e-02,
8.94905366e-02, -3.22181466e-02, 1.48623606e-04,
-1.94744321e-02, -5.98537627e-02, 7.38995575e-02,
-9.34850871e-03],
[ 6.97656156e-02, -4.07965348e-02, -2.27650299e-02,
3.60535977e-02, 5.90046014e-02, 3.68699626e-02,
-2.75352642e-02, -1.13279486e-01, 7.62366593e-02,
-1.10627640e-01],
[ 1.30942451e-01, -5.57738922e-02, -2.90269961e-02,
2.85175202e-02, -6.41159865e-02, 1.11356763e-01,
-9.07462843e-02, -2.17497789e-02, 3.22838117e-02,
-4.26609274e-03],
[-2.53527523e-01, -1.75525512e-01, 2.81308138e-01,
1.76775108e-01, -8.78590313e-02, 6.42559168e-02,
1.00242118e-01, -7.55856371e-02, 4.01876550e-02,
-1.05311825e-01],
[-7.43872204e-02, -3.23953300e-02, 1.85856930e-01,
4.99661970e-02, -6.29346570e-03, -7.15626559e-02,
8.45457324e-02, -6.18686997e-02, -3.84964805e-02,
-9.24084299e-02],
[-7.35924539e-02, -4.54414452e-02, -3.46318087e-03,
9.58531126e-02, -1.56587071e-02, 5.75107809e-02,
-1.05680764e-01, -3.47310461e-02, 8.39961783e-02,
5.82180335e-02],
[-4.66406766e-03, 1.99657690e-02, 6.56735727e-03,
-1.32200215e-02, -6.94371503e-04, 7.96111202e-03,
8.52651207e-03, -3.57402915e-03, -6.46029210e-03,
-1.77755340e-02],
[-2.50551322e-02, -1.67524208e-02, -3.25948981e-05,
1.69864377e-02, 1.78632067e-02, -3.32345528e-04,
-1.41316034e-03, -1.90838504e-02, 2.42332531e-02,
-9.29983541e-03],
[-2.26565378e-01, -1.48914766e-01, 1.49222961e-01,
2.48685491e-01, -1.08223684e-01, 8.69246094e-02,
-1.52387443e-01, -1.15538534e-01, 1.94590660e-01,
8.70598260e-02]]})

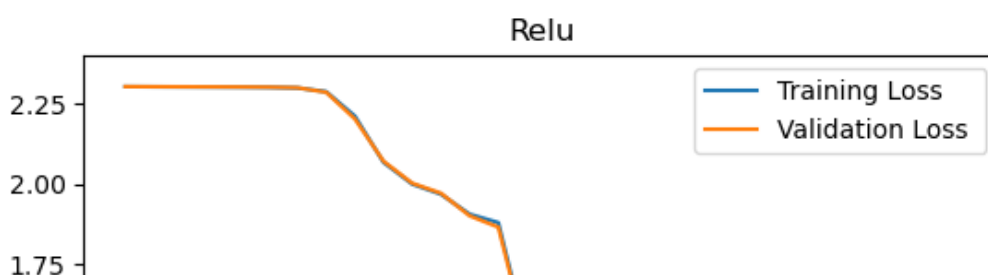
```

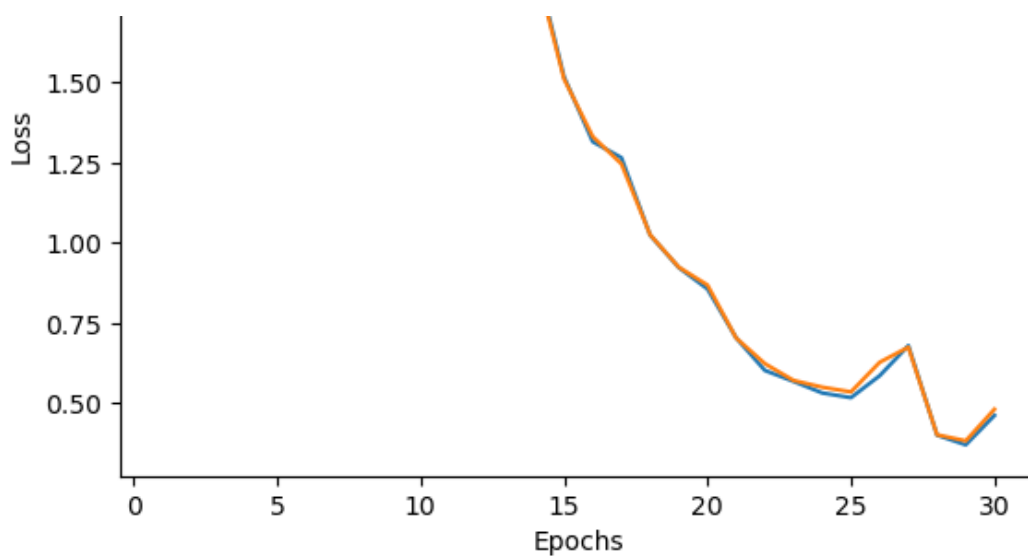
In [19]:

```

#plot with relu learning rate 0.08
plt.plot(nn.train_losses, label = "Training Loss ")
plt.plot(nn.val_losses, label = "Validation Loss ")
plt.ylabel('Loss')
plt.legend()
plt.xlabel('Epochs')
plt.title("Relu")
plt.show()

```





In [12]:

```
import pickle
```

In []:

```
# save Relu the trained models
```

In [18]:

```
f = open("demofile", "wb")
pickle.dump(nn,f)
f.close()
```

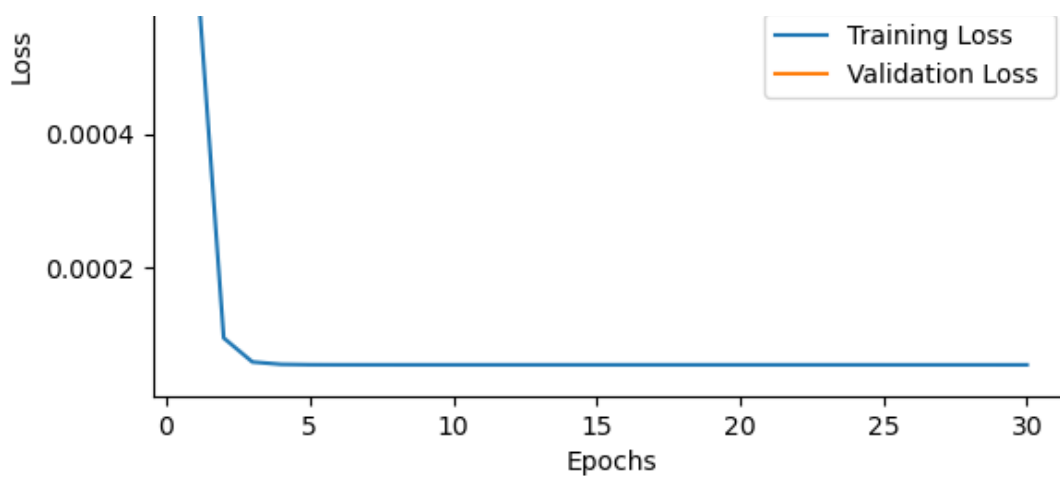
In [8]:

```
alltheactivations=['sigmoid','tanh','linear','leakyrelu', 'softmax']

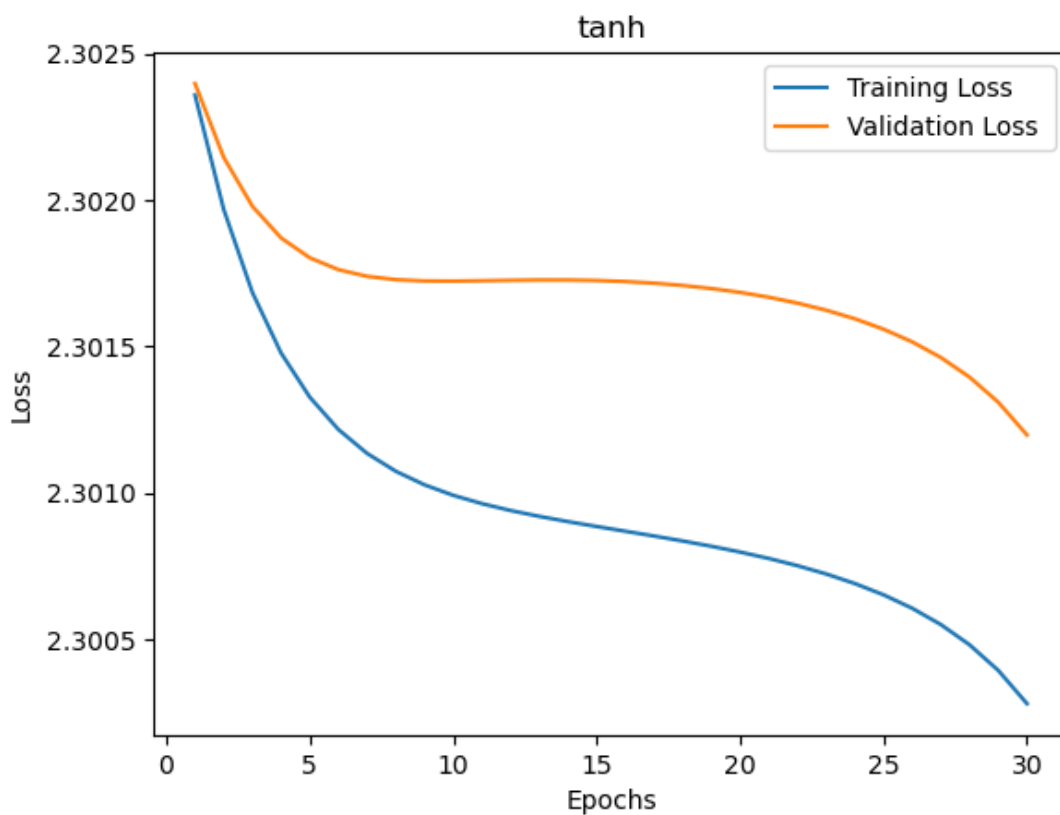
for x in alltheactivations:
    nn = neural_networks.NN(6, [784,256, 128, 64, 32,10], x, 0.08, 'normal', len(X_train)/
/20, 30)
    nn.fit(X_train,Y_train,X_val,Y_val)
    print(f"The Accuracy Score for {x} activation fn is :")
    print(nn.score(X_test,Y_test))
    alltheweights[x] = nn.params
    plt.plot(list(range(1,len(nn.train_losses) + 1)),nn.train_losses, label = "Training Lo
ss ")
    plt.plot(list(range(1,len(nn.val_losses) + 1)),nn.val_losses, label = "Validation Loss
")
    plt.ylabel('Loss')
    plt.legend()
    plt.xlabel('Epochs')
    plt.title(x)
    plt.show()
```

Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch:
10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch:
19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch:
28 Epoch: 29 Epoch: 30 The Accuracy Score for sigmoid activation fn is :
0.11357142857142857





Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch: 10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch: 19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch: 28 Epoch: 29 Epoch: 30 The Accuracy Score for tanh activation fn is : 0.11357142857142857

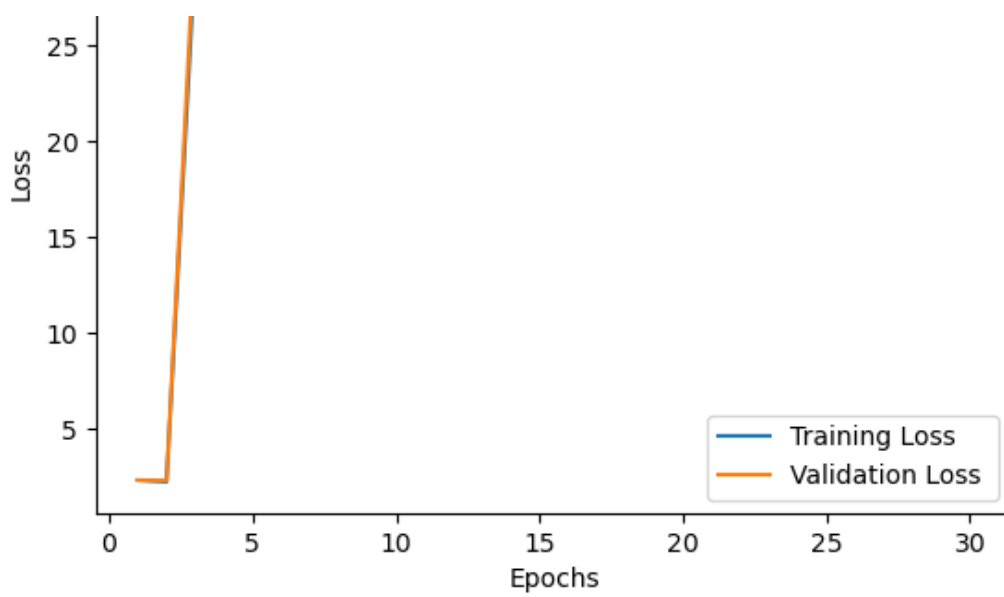


Epoch: 1 Epoch: 2 Epoch: 3

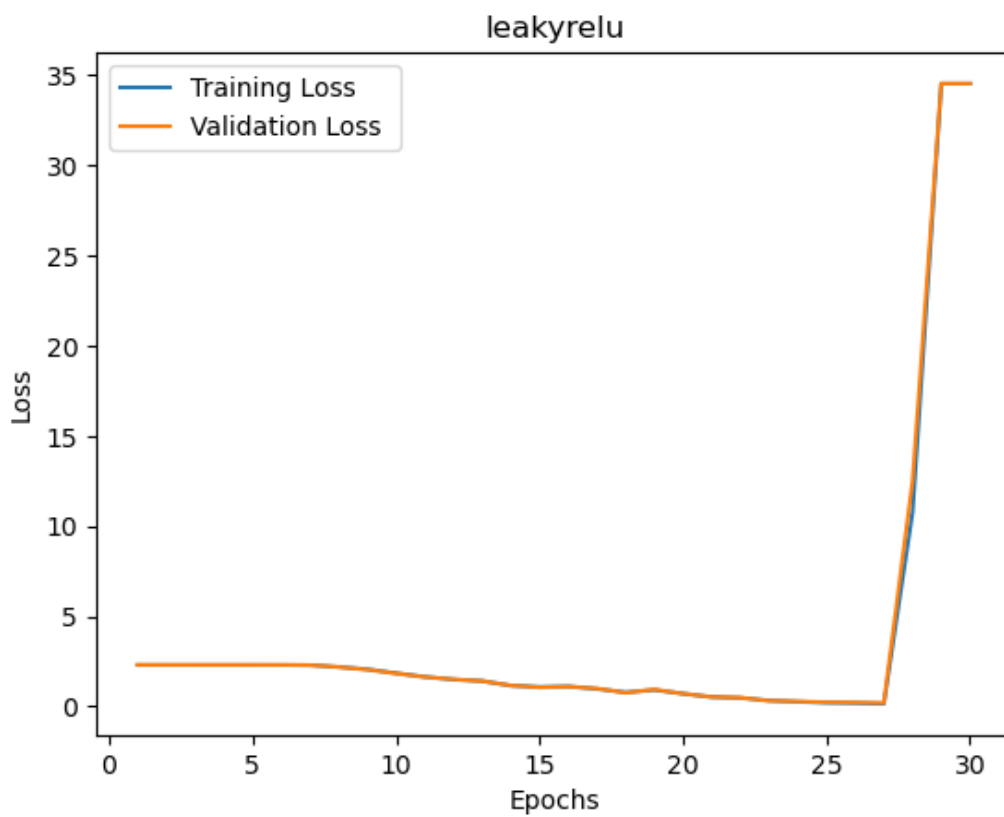
```
d:\Desktop\Ml\Ml_assignment\assignment_3\new_Q2\neural_networks.py:77: RuntimeWarning: overflow encountered in exp
  AL = (np.exp(ZL)/(np.sum(np.exp(ZL),axis = 1, keepdims = True))) # SoftMax
d:\Desktop\Ml\Ml_assignment\assignment_3\new_Q2\neural_networks.py:77: RuntimeWarning: invalid value encountered in true_divide
  AL = (np.exp(ZL)/(np.sum(np.exp(ZL),axis = 1, keepdims = True))) # SoftMax
```

Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch: 10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch: 19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch: 28 Epoch: 29 Epoch: 30 The Accuracy Score for linear activation fn is : 0.10128571428571428

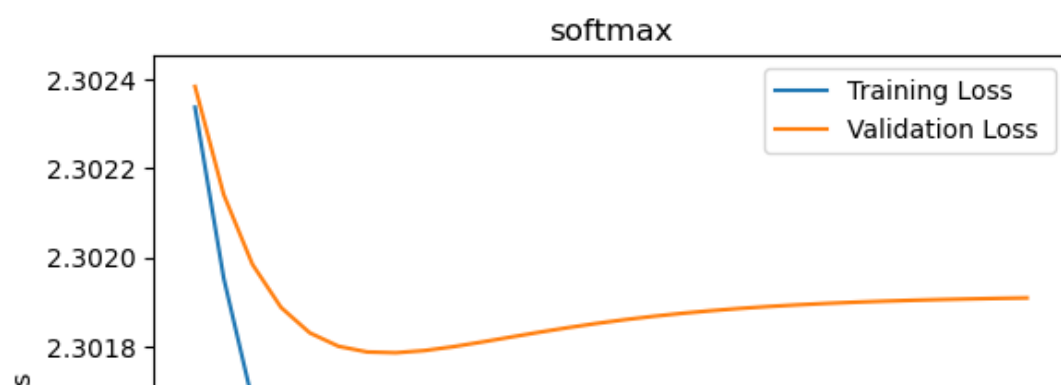


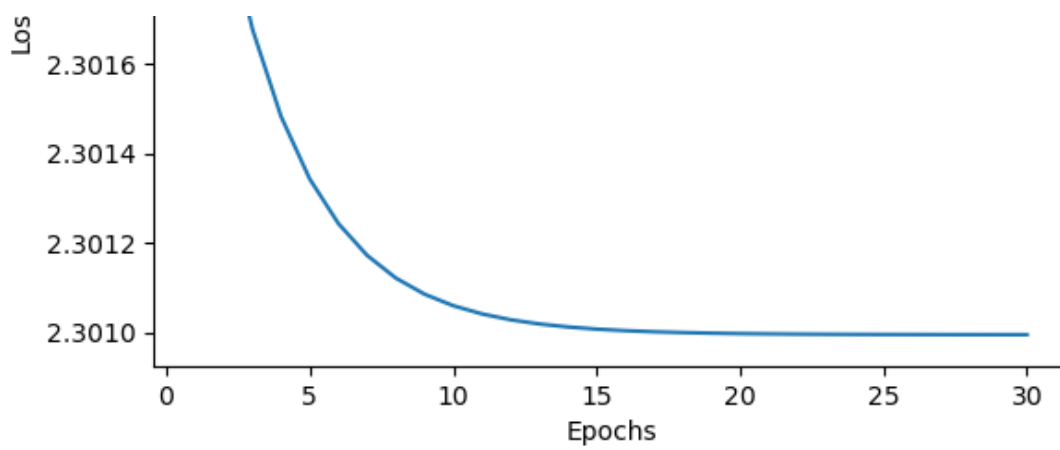


Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch: 10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch: 19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch: 28 Epoch: 29 Epoch: 30 The Accuracy Score for leakyrelu activation fn is : 0.10128571428571428



Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch: 10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch: 19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch: 28 Epoch: 29 Epoch: 30 The Accuracy Score for softmax activation fn is : 0.11357142857142857





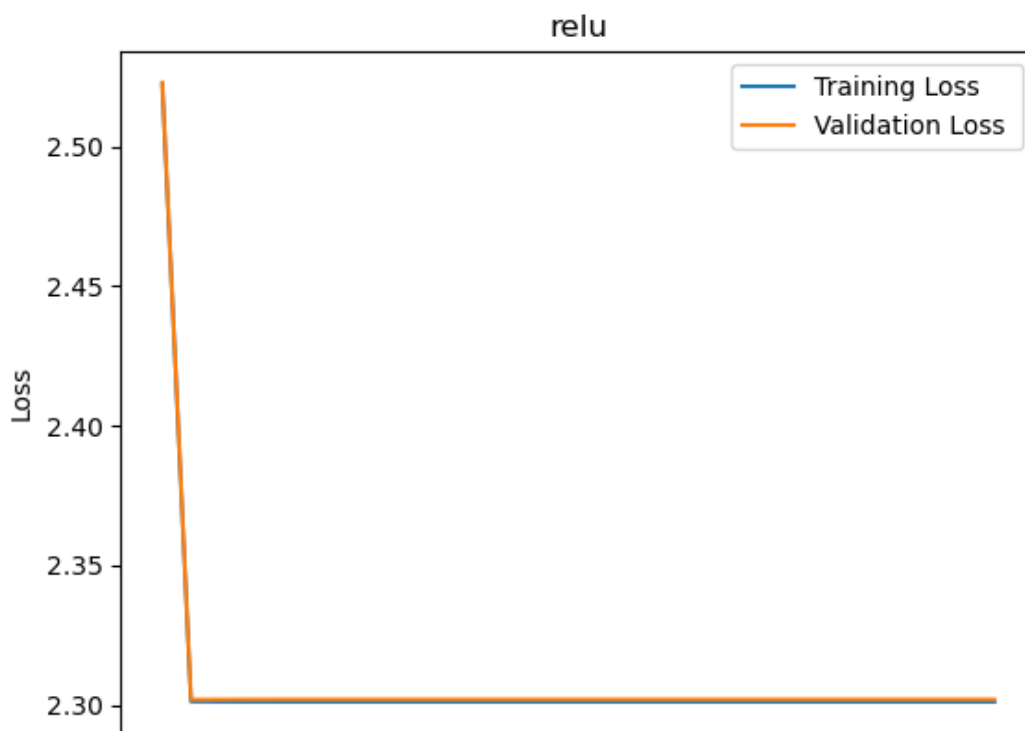
with random weight and lr = 0.08

In [20]:

```
alltheactivations=['relu','sigmoid','tanh','linear','leakyrelu','softmax']
alltheweights={}

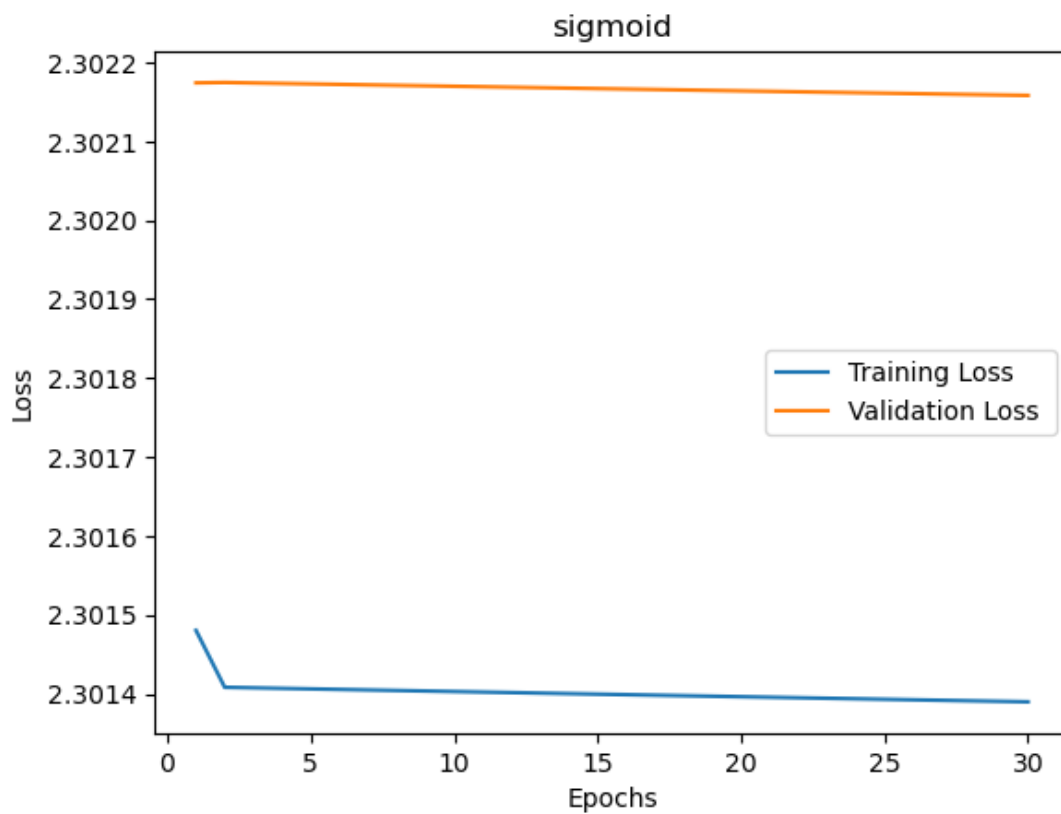
for x in alltheactivations:
    nn = neural_networks.NN(6, [784,256, 128, 64, 32,10], x, 0.08, 'random', len(X_train)/
/128, 30)
    nn.fit(X_train,Y_train,X_val,Y_val)
    print(f"The Accuracy Score for {x} activation fn is :")
    print(nn.score(X_test,Y_test))
    alltheweights[x] = nn.params
    plt.plot(list(range(1,len(nn.train_losses) + 1)),nn.train_losses, label = "Training Lo
ss ")
    plt.plot(list(range(1,len(nn.val_losses) + 1)),nn.val_losses, label = "Validation Loss
")
    plt.ylabel('Loss')
    plt.legend()
    plt.xlabel('Epochs')
    plt.title(x)
    plt.show()
```

Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch:
10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch:
19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch:
28 Epoch: 29 Epoch: 30 The Accuracy Score for relu activation fn is :
0.11357142857142857

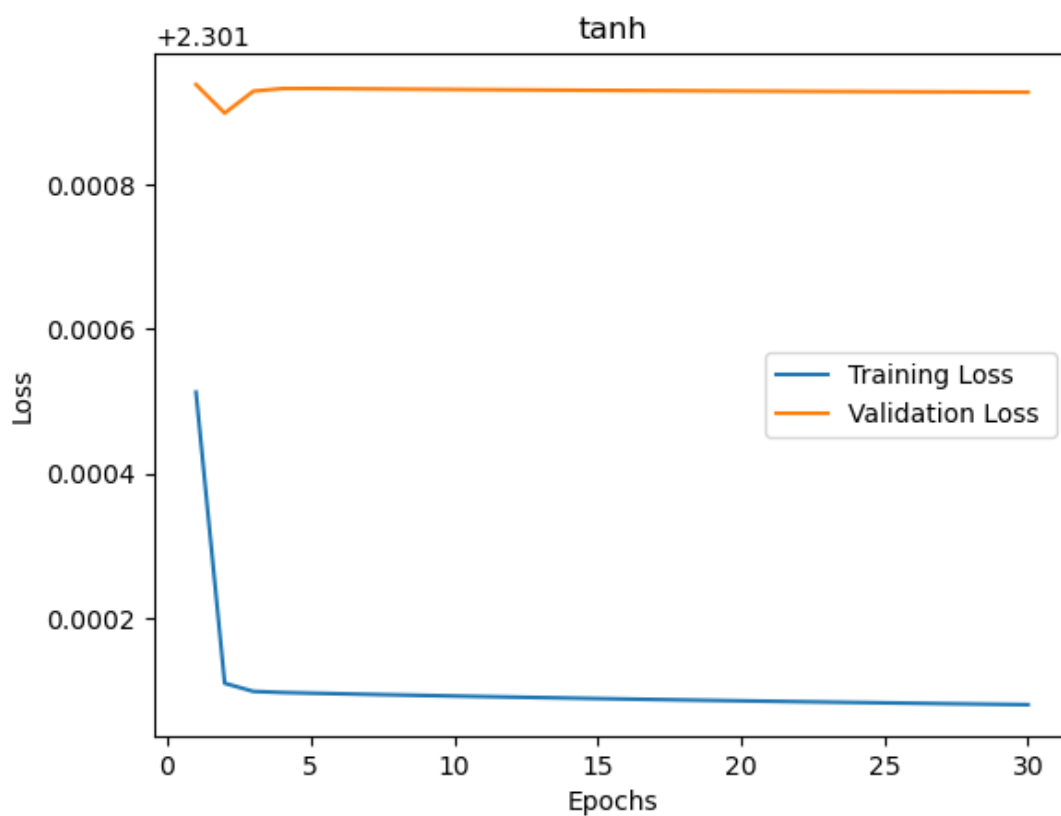


0 5 10 15 20 25 30
Epochs

Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch:
10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch:
19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch:
28 Epoch: 29 Epoch: 30 The Accuracy Score for sigmoid activation fn is :
0.11357142857142857



Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch:
10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch:
19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch:
28 Epoch: 29 Epoch: 30 The Accuracy Score for tanh activation fn is :
0.11357142857142857



Epoch: 1

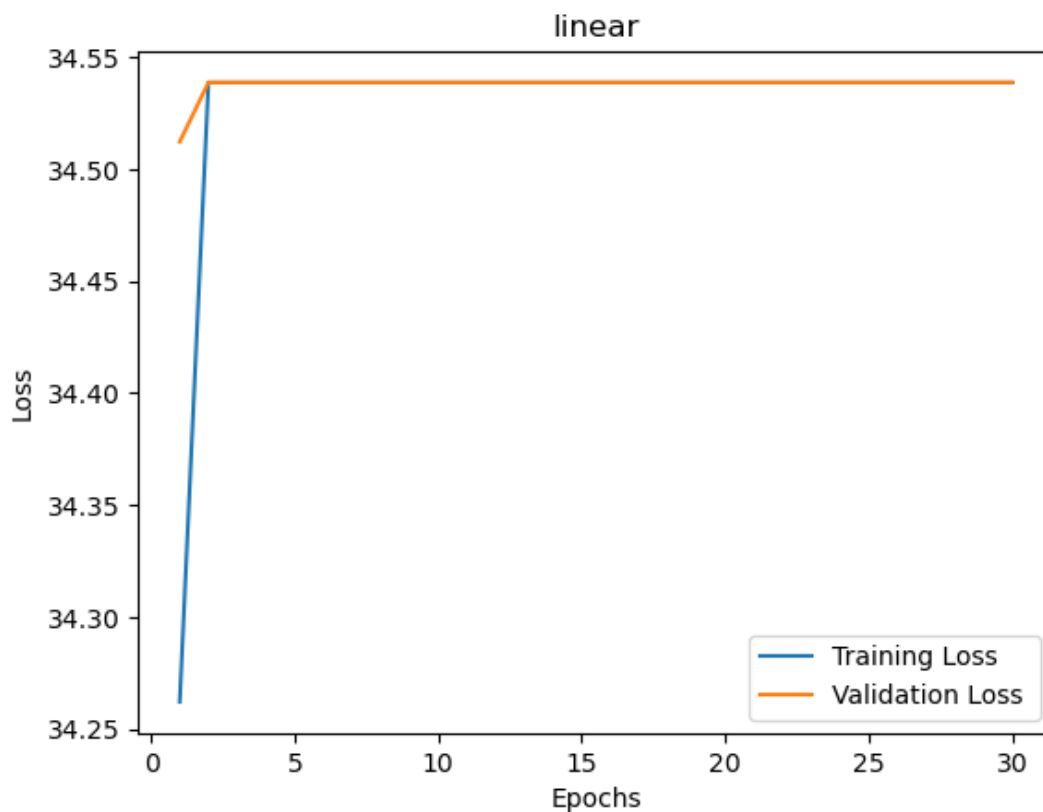
d:\Desktop\Ml\Ml_assignment\assignment_3\new_Q2\neural_networks.py:77: RuntimeWarning: overflow encountered in exp

```
AL = (np.exp(ZL)/(np.sum(np.exp(ZL),axis = 1, keepdims = True))) # SoftMax
```

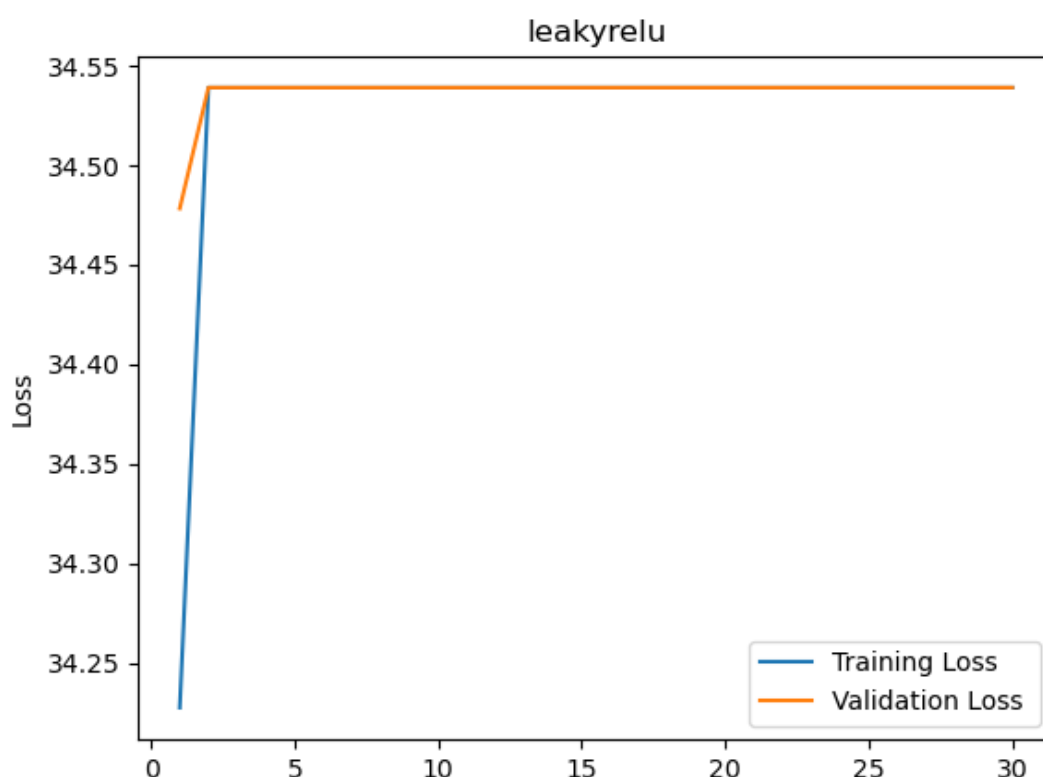
d:\Desktop\Ml\Ml_assignment\assignment_3\new_Q2\neural_networks.py:77: RuntimeWarning: in valid value encountered in true_divide

```
AL = (np.exp(ZL)/(np.sum(np.exp(ZL),axis = 1, keepdims = True))) # SoftMax
```

Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch: 10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch: 19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch: 28 Epoch: 29 Epoch: 30 The Accuracy Score for linear activation fn is : 0.10128571428571428

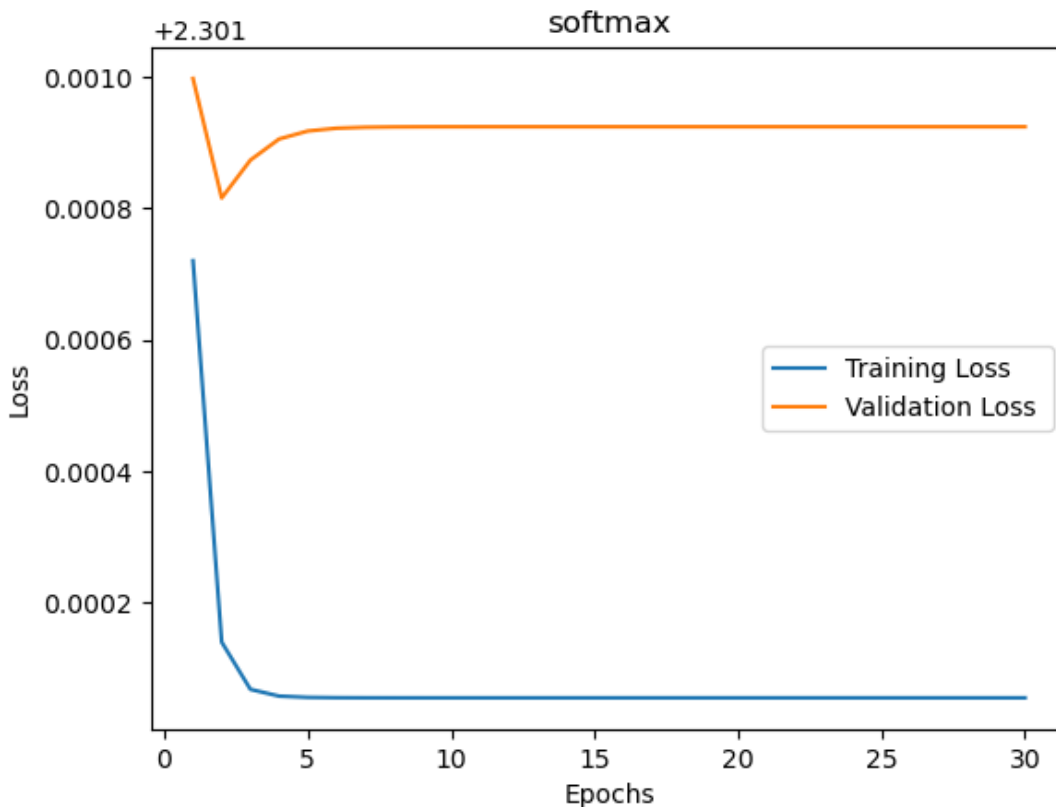


Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch: 10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch: 19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch: 28 Epoch: 29 Epoch: 30 The Accuracy Score for leakyrelu activation fn is : 0.10128571428571428



Epochs

Epoch: 1 Epoch: 2 Epoch: 3 Epoch: 4 Epoch: 5 Epoch: 6 Epoch: 7 Epoch: 8 Epoch: 9 Epoch: 10 Epoch: 11 Epoch: 12 Epoch: 13 Epoch: 14 Epoch: 15 Epoch: 16 Epoch: 17 Epoch: 18 Epoch: 19 Epoch: 20 Epoch: 21 Epoch: 22 Epoch: 23 Epoch: 24 Epoch: 25 Epoch: 26 Epoch: 27 Epoch: 28 Epoch: 29 Epoch: 30 The Accuracy Score for softmax activation fn is : 0.11357142857142857



(4th) Explaniation

- From the above test accuracies, it is quite evident that “ReLU” works best. It has the best test accuracy.
- The worst performance is shown in the case of the “linear” activation function as the accuracy is least in that case. It does not perform any activation and is suitable for single layers only, in our multi-layer case it does not perform well.
- For “tanh”, also we weren't able to make a good model. Maybe with more epochs we could get a desired result, but since i was low on computation power. I used only 30 epochs for these experiments.
- For “relu”, it shows the best behaviour it reacts very fast to the changes and to the data. The only noticeable problem is that after certain number of epochs we tend to go away from the global
- For “linear”, the training loss keeps in decreasing while validation loss fluctuates a bit and then stablizes.
- For “sigmoid”, it was stuck at a local minima for a long time and the moment it got out, the training loss started decreasing while validation loss shot and then decreased to a nominal value.

Testing own sklearn models

In []:

```
for activation in ["logistic", "tanh", "identity", "relu"]:  
    nn = MLPClassifier(activation=activation, hidden_layer_sizes=[256, 128, 64, 32], learning_rate_init=0.08, max_iter=30, solver="sgd", alpha = 0)  
    nn.fit(X_train, Y_train)  
    print(f'Test accuracy for {activation} = {nn.score(X_test, Y_test)}')
```

Test accuracy for logistic = 0.814

Test accuracy for tanh = 0.103

Test accuracy for identity = 0

Test accuracy for relu = 0.861

In case of sigmoid (logistic) sklearn implementation is far superior. In the case of “linear”, the accuracy on test in custom implementation comes out to be better than the one in sklearn’s as it did not converge with the given parameters. In case of tanh we get similar results as 30 epochs are not sufficient which is what we have taken for our experiments. Sklearn also performs best for relu just like ours and the accuracies are also closeby.

In [1]:

```
!git clone https://github.com/zalandoresearch/fashion-mnist.git
```

fatal: destination path 'fashion-mnist' already exists and is not an empty directory.

In [1]:

```
import numpy as np
```

In [2]:

```
import sys
sys.path.insert(0, 'fashion-mnist/utils')
```

In [3]:

```
import mnist_reader
```

In [5]:

```
X, Y = mnist_reader.load_mnist('fashion-mnist/data/fashion', kind='train')
X_test, y_test = mnist_reader.load_mnist('fashion-mnist/data/fashion', kind='t10k')
```

In [6]:

```
from sklearn.model_selection import train_test_split
from sklearn.neural_network import MLPClassifier as MLPRegressor
import matplotlib.pyplot as plt
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import train_test_split
# from keras.utils import to_categorical
from sklearn.metrics import log_loss
```

**According to question we have to
(divide training set into 85:15 train to validation set).**

In [7]:

```
X_train, X_val, y_train, y_val = train_test_split(X,Y,test_size=0.15,random_state=10)
```

Part A

In [8]:

```
clf = MLPRegressor(max_iter=450,batch_size=35, hidden_layer_sizes=(256, 32), activation=
'logistic',random_state=10,verbose= False)

train_loss = []
valid_loss = []

for j in range(0,100):

    clf.partial_fit(X_train,y_train,[0,1,2,3,4,5,6,7,8,9])

    # y_cal_t = clf.predict_proba(X_train)

    y_cal_v = clf.predict_proba(X_val)
    valid_loss.append(log_loss(y_val, y_cal_v))
train_loss = clf.loss_curve_
print(clf.score(X_test,y_test))
```

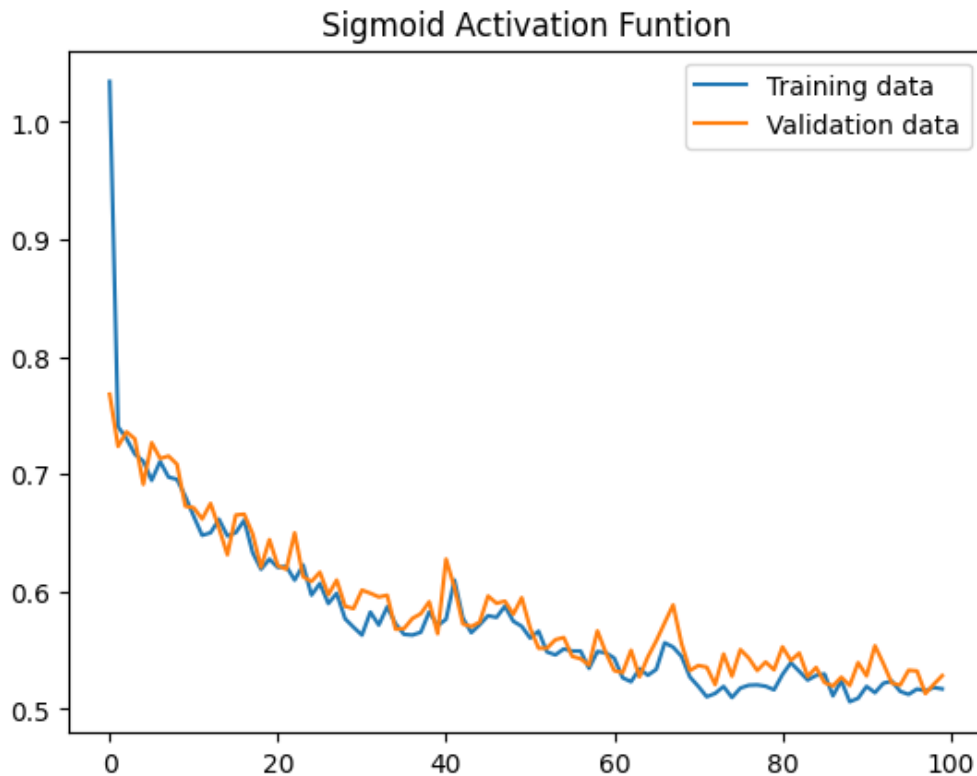

0.8037

In [9]:

```
plt.plot(train_loss)
plt.plot(valid_loss)
plt.legend(['Training data', 'Validation data'])
plt.title('Sigmoid Activation Funtion')
```

Out[9]:

Text(0.5, 1.0, 'Sigmoid Activation Funtion')



In [9]:

```
Relu = MLPRegressor(max_iter=450,batch_size=35, hidden_layer_sizes=(256, 32), activation
='relu',random_state=10,verbose= False)

train_loss = []
valid_loss = []

for j in range(0,100):

    Relu.partial_fit(X_train,y_train,[0,1,2,3,4,5,6,7,8,9])

    # y_cal_t = clf.predict_proba(X_train)

    y_cal_v = Relu.predict_proba(X_val)
    valid_loss.append(log_loss(y_val, y_cal_v))
train_loss = Relu.loss_curve_
print(Relu.score(X_test,y_test))
```

0.8512

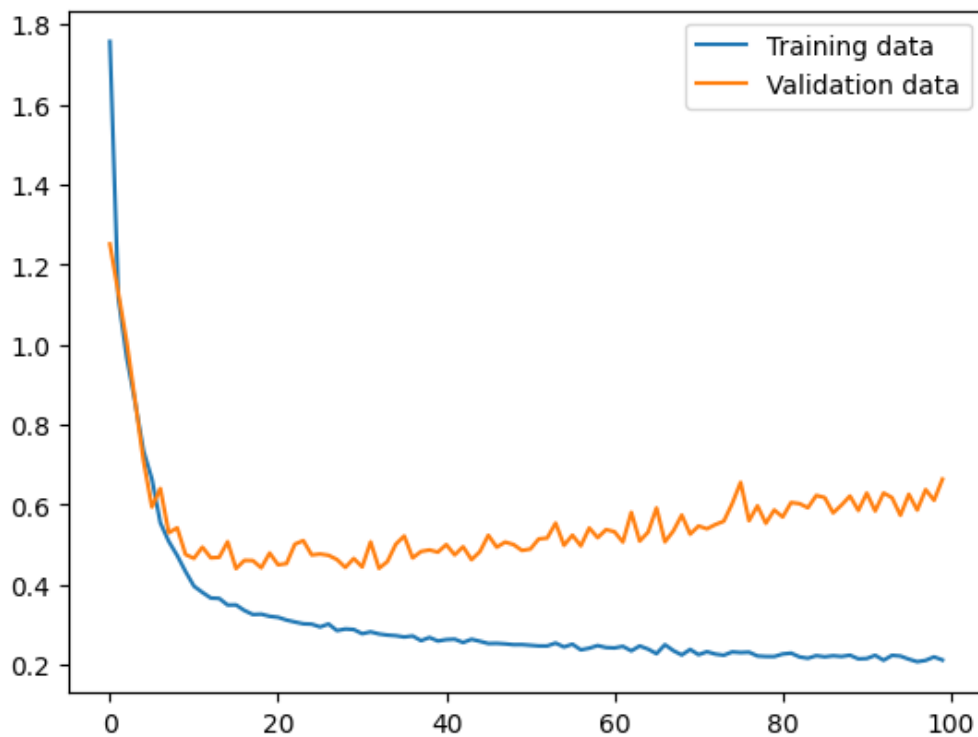
In [11]:

```
plt.plot(train_loss)
plt.plot(valid_loss)
plt.legend(['Training data', 'Validation data'])
plt.title('Relu Activation Funtion')
```

Out[11]:

Text(0.5, 1.0, 'Relu Activation Funtion')

Relu Activation Funtion



In [10]:

```
# tanh
clf = MLPRegressor(max_iter=450, batch_size=35, hidden_layer_sizes=(256, 32), activation=
'tanh', random_state=10, verbose= False)

train_loss = []
valid_loss = []

for j in range(0,100):

    clf.partial_fit(X_train,y_train,[0,1,2,3,4,5,6,7,8,9])

    y_cal_v = clf.predict_proba(X_val)
    valid_loss.append(log_loss(y_val, y_cal_v))
train_loss = clf.loss_curve_
print(clf.score(X_test,y_test))
```

0.7704

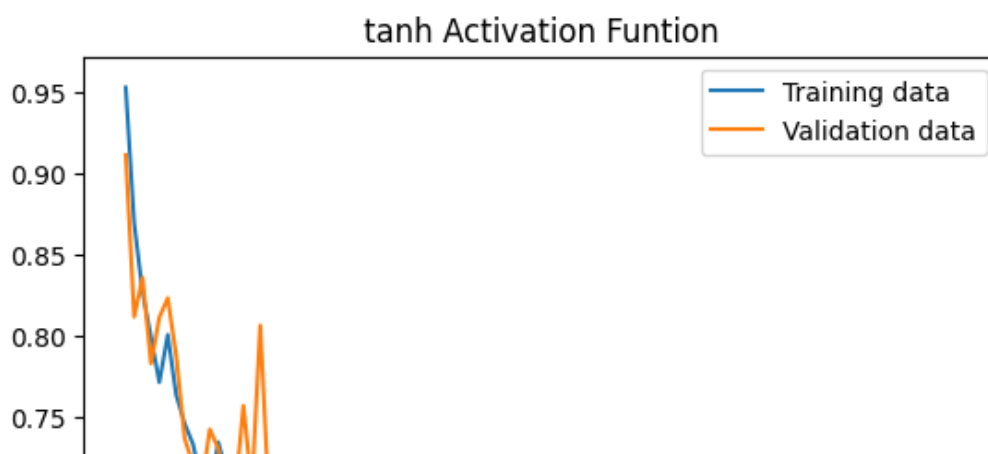
In [10]:

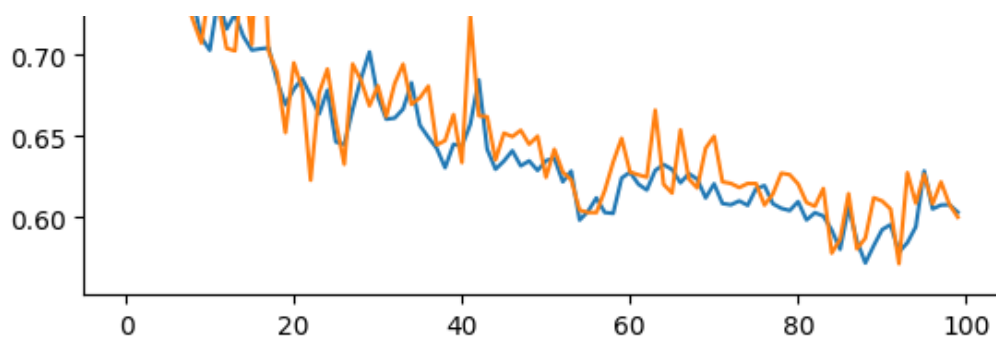
```
plt.plot(train_loss)
plt.plot(valid_loss)
plt.legend(['Training data', 'Validation data'])

plt.title('tanh Activation Funtion')
```

Out[10]:

Text(0.5, 1.0, 'tanh Activation Funtion')





In [11]:

```
# Linear
clf = MLPRegressor(max_iter=450,batch_size=35, hidden_layer_sizes=(256, 32), activation=
'identity',random_state=10,verbose= False)

train_loss = []
valid_loss = []

for j in range(0,100):

    clf.partial_fit(X_train,y_train,[0,1,2,3,4,5,6,7,8,9])

    # y_cal_t = clf.predict_proba(X_train)

    y_cal_v = clf.predict_proba(X_val)
    valid_loss.append(log_loss(y_val, y_cal_v))
    train_loss = clf.loss_curve_
    print(clf.score(X_test,y_test))
```

0.7986

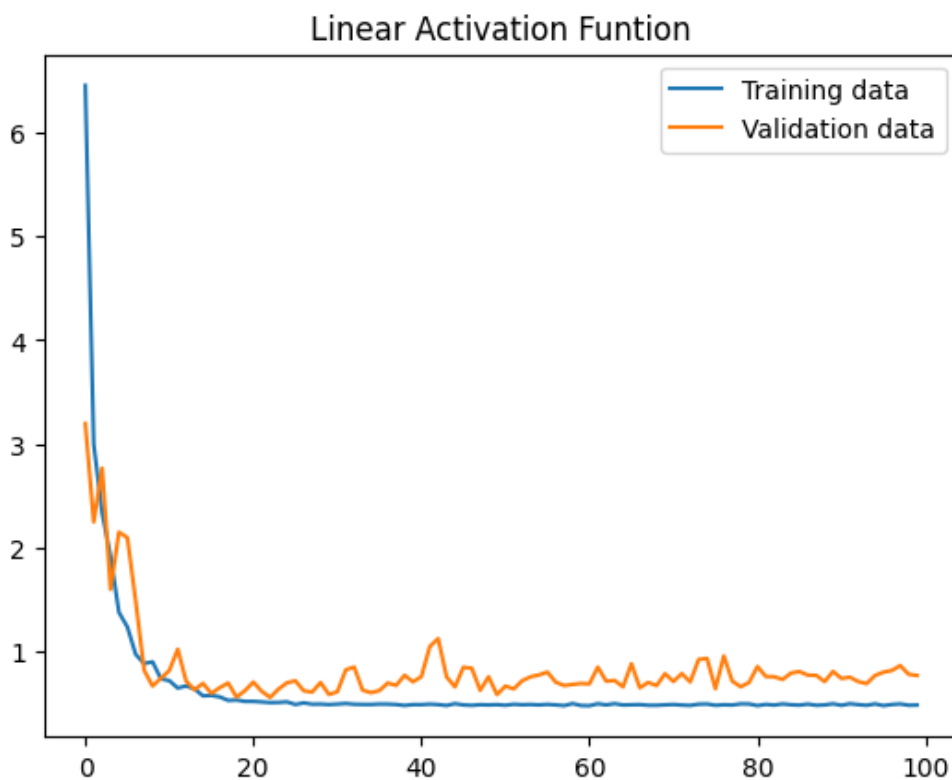
In [12]:

```
plt.plot(train_loss)
plt.plot(valid_loss)
plt.legend(['Training data', 'Validation data'])

plt.title('Linear Activation Funtion')
```

Out[12]:

Text(0.5, 1.0, 'Linear Activation Funtion')



Choosing the right activation function is important and it can be consider as a type of hyperparameter tuning in which , I manually choosen the activation function on the bases of Accuracy and plot(loss v/s epochs and validation loss v/s epochs)

The accuracy of diffent activation are:

Simgoid : 0.8037

Relu : 0.8512

tanh: 0.7704

linear:0.7986

As we can see from the above plots that the differnce of training and validation loss for relu is beacuse Relu counters the problem of vanishing gradients that decreases the learnability of a neural network. Due to this reason, it performs better than others.

we can also see that accuracy of Relu best on testing set also

clearly the Relu activation function performance is best

Part B

In [12]:

```
clf = MLPRegressor(learning_rate_init=0.1,batch_size=35, hidden_layer_sizes=(256, 32), a
ctivation='relu',random_state=10,verbose= False)

train_loss = []
valid_loss = []

for j in range(0,35):
    # print(j)
    clf.partial_fit(X_train,y_train,[0,1,2,3,4,5,6,7,8,9])

    # y_cal_t = clf.predict_proba(X_train)

    y_cal_v = clf.predict_proba(X_val)
    valid_loss.append(log_loss(y_val, y_cal_v))

train_loss = clf.loss_curve_
print(clf.score(X_test,y_test))
```

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

25
26
27
28
29
30
31
32
33
34
0.1

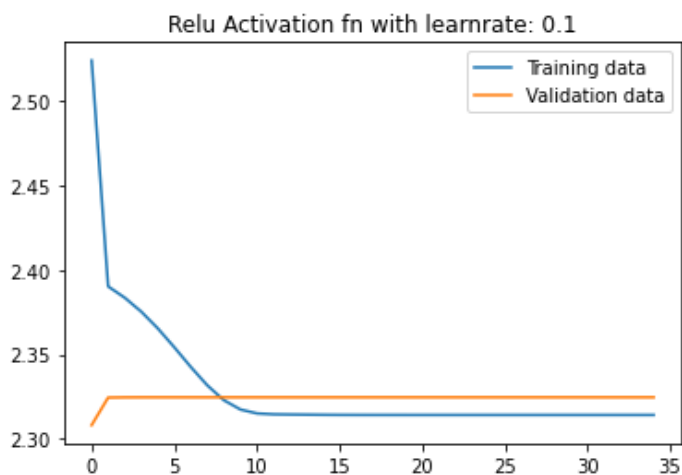
In [13]:

```
plt.plot(train_loss)
plt.plot(valid_loss)
plt.legend(['Training data', 'Validation data'])

plt.title("Relu Activation fn with learnrate: 0.1")
```

Out[13]:

Text(0.5, 1.0, 'Relu Activation fn with learnrate: 0.1')



In [13]:

```
clf = MLPRegressor(learning_rate_init=0.01,batch_size=35, hidden_layer_sizes=(256, 32),
activation='relu',random_state=10,verbose= False)

train_loss = []
valid_loss = []

for j in range(0,35):

    clf.partial_fit(X_train,y_train,[0,1,2,3,4,5,6,7,8,9])

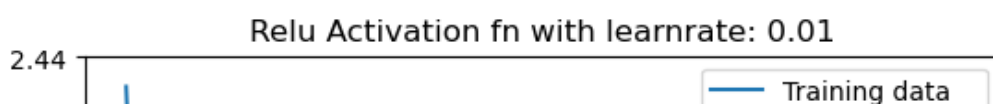
    # y_cal_t = clf.predict_proba(X_train)

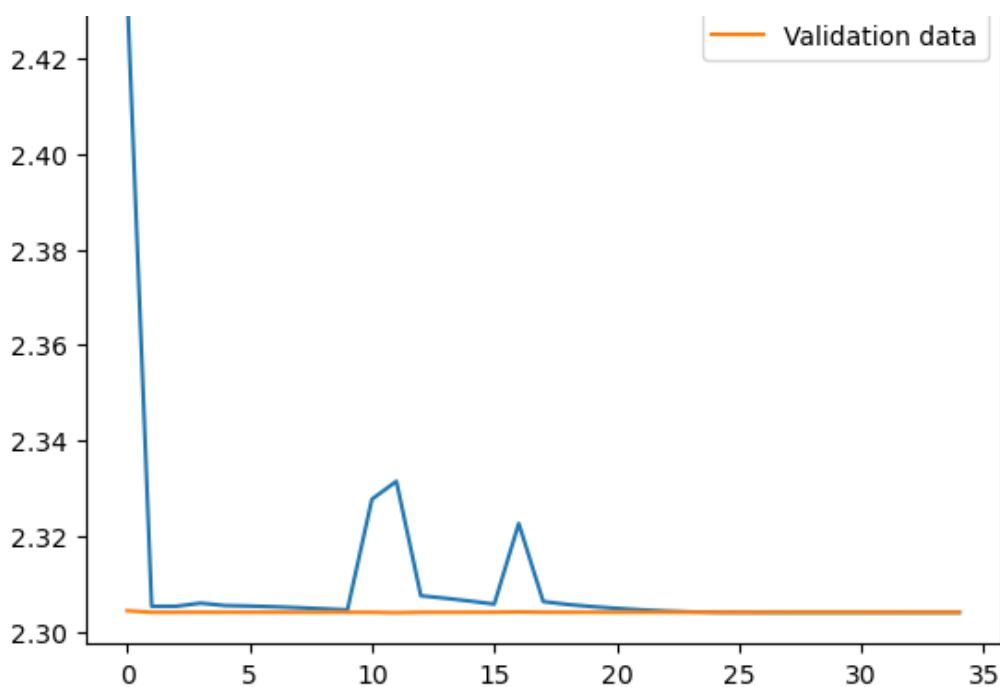
    y_cal_v = clf.predict_proba(X_val)
    valid_loss.append(log_loss(y_val, y_cal_v))

train_loss = clf.loss_curve_
print(clf.score(X_test,y_test))
plt.plot(train_loss)
plt.plot(valid_loss)
plt.legend(['Training data', 'Validation data'])

plt.title("Relu Activation fn with learnrate: 0.01")
train_loss.clear()
valid_loss.clear()
```

0.1





In [14]:

```
clf = MLPRegressor(learning_rate_init=0.001,batch_size=35, hidden_layer_sizes=(256, 32),
activation='relu',random_state=10,verbose= False)

train_loss = []
valid_loss = []

for j in range(0,35):

    clf.partial_fit(X_train,y_train,[0,1,2,3,4,5,6,7,8,9])

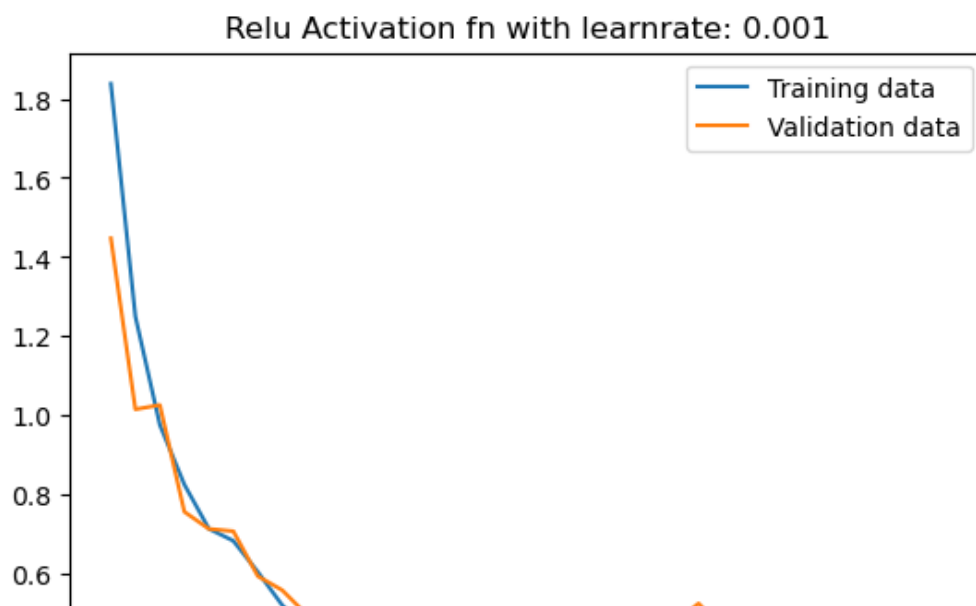
    # y_cal_t = clf.predict_proba(X_train)

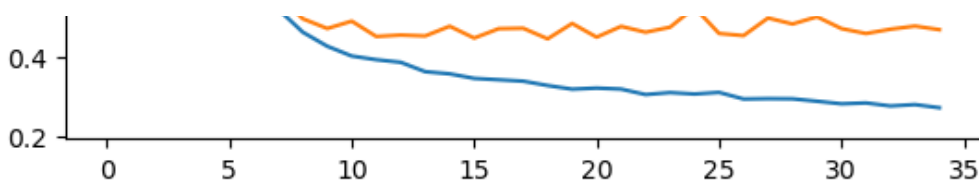
    y_cal_v = clf.predict_proba(X_val)
    valid_loss.append(log_loss(y_val, y_cal_v))

train_loss = clf.loss_curve_
print(clf.score(X_test,y_test))
plt.plot(train_loss)
plt.plot(valid_loss)
plt.legend(['Training data', 'Validation data'])

plt.title("Relu Activation fn with learnrate: 0.001")
train_loss.clear()
valid_loss.clear()
```

0.8645





In [2]:

```
rate_arr = [0.1,0.01,0.001]
for i in rate_arr:
    print("Relu Activation fn with learnrate:", i)
```

```
Relu Activation fn with learnrate: 0.1
Relu Activation fn with learnrate: 0.01
Relu Activation fn with learnrate: 0.001
```

Result

Learning rate refers to how frequently the weights are updated during the training process

As seen in the graphs, lr = 0.001 works the best, and the most likely reason is that the learning rate is excellent for finding a minimum, as opposed to the other two examples. The weights appear to shift violently in the other two situations, thus no minima could be established

- we can also see that accuracy of 0.001 best on testing set which is 0.8645

Part-c

In [15]:

```
# for hidden layer (130,66)
Relu = MLPRegressor(batch_size=35, hidden_layer_sizes=(130, 66), activation='relu',random_state=10,verbose=False)

train_loss = []
valid_loss = []

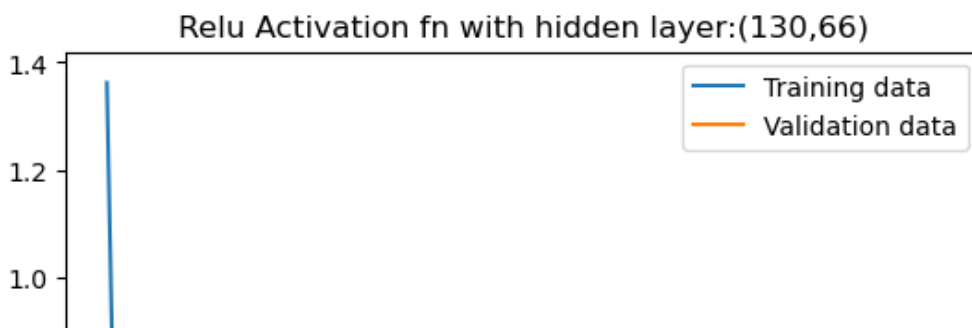
for j in range(0,100):

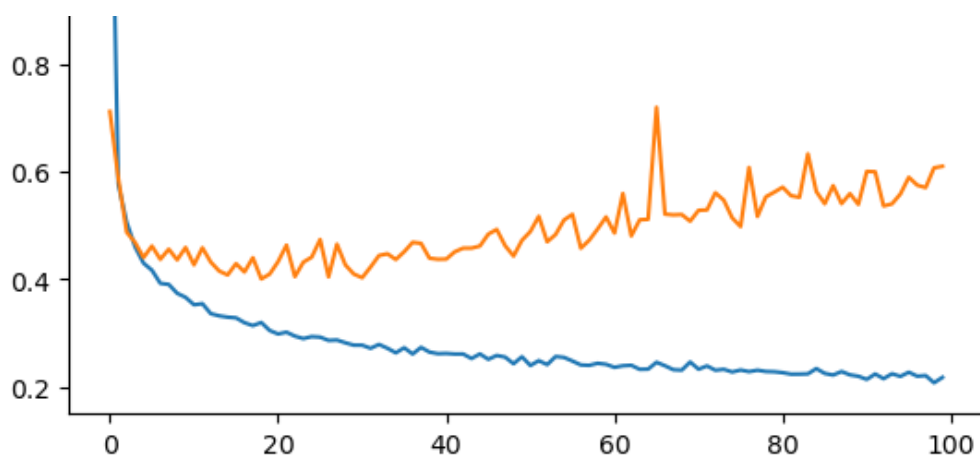
    Relu.partial_fit(X_train,y_train,[0,1,2,3,4,5,6,7,8,9])

    y_cal_v = Relu.predict_proba(X_val)
    valid_loss.append(log_loss(y_val, y_cal_v))
train_loss = Relu.loss_curve_
print(Relu.score(X_test,y_test))
plt.plot(train_loss)
plt.plot(valid_loss)
plt.legend(['Training data', 'Validation data'])

plt.title("Relu Activation fn with hidden layer:(130,66)")
train_loss.clear()
valid_loss.clear()
```

0.8561





In [16]:

```
# for hidden layer (64,32)
Relu = MLPRegressor(batch_size=35, hidden_layer_sizes=(64, 32), activation='relu', random
_state=10,verbose= False)

train_loss = []
valid_loss = []

for j in range(0,100):

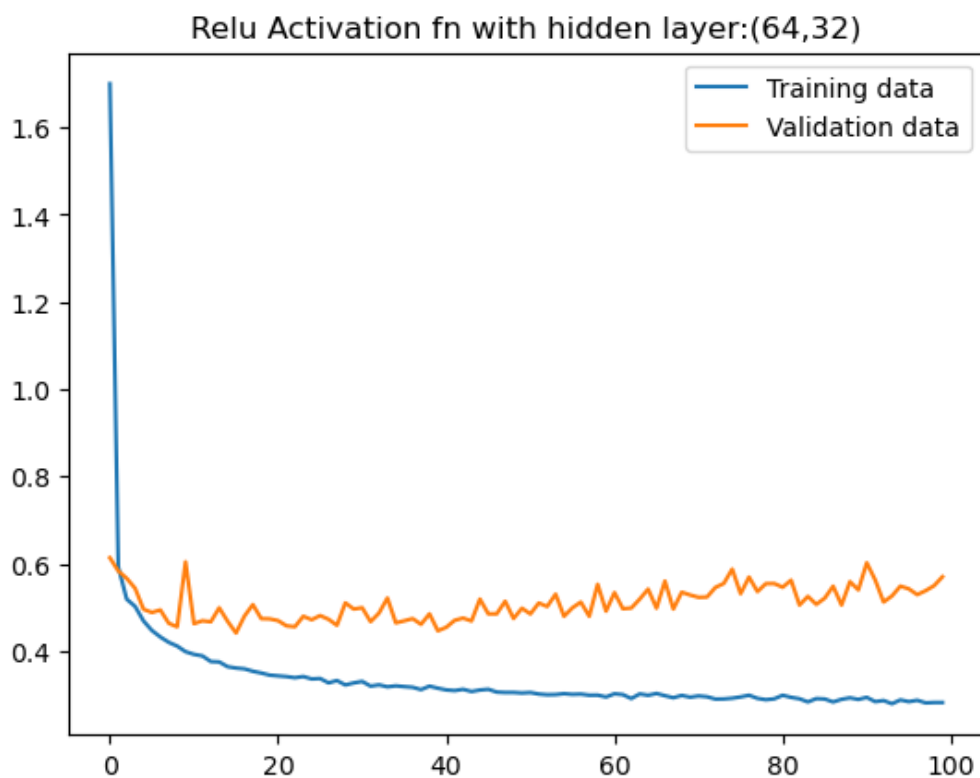
    Relu.partial_fit(X_train,y_train,[0,1,2,3,4,5,6,7,8,9])

    # y_cal_t = clf.predict_proba(X_train)

    y_cal_v = Relu.predict_proba(X_val)
    valid_loss.append(log_loss(y_val, y_cal_v))
train_loss = Relu.loss_curve_
print(Relu.score(X_test,y_test))
plt.plot(train_loss)
plt.plot(valid_loss)
plt.legend(['Training data', 'Validation data'])

plt.title("Relu Activation fn with hidden layer:(64,32)")
train_loss.clear()
valid_loss.clear()
```

0.8553



In [17]:

```
# for hidden layer (32,16)
Relu = MLPRegressor(batch_size=35, hidden_layer_sizes=(32,16), activation='relu', random_
state=10,verbose= False)

train_loss = []
valid_loss = []

for j in range(0,100):

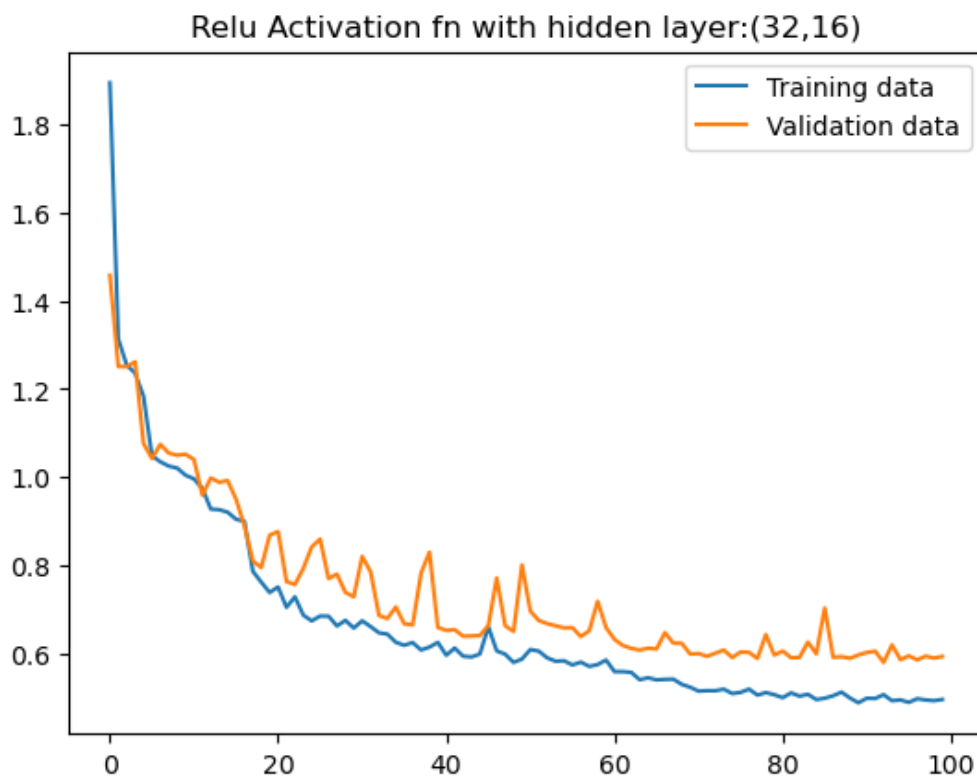
    Relu.partial_fit(X_train,y_train,[0,1,2,3,4,5,6,7,8,9])

    # y_cal_t = clf.predict_proba(X_train)

    y_cal_v = Relu.predict_proba(X_val)
    valid_loss.append(log_loss(y_val, y_cal_v))
train_loss = Relu.loss_curve_
print(Relu.score(X_test,y_test))
plt.plot(train_loss)
plt.plot(valid_loss)
plt.legend(['Training data', 'Validation data'])

plt.title("Relu Activation fn with hidden layer:(32,16)")
train_loss.clear()
valid_loss.clear()
```

0.8012



Result

We can see that as the number of neurons in the hidden layers decreases, our training loss and validation loss increase, affecting the model's performance because when a neural network has too few hidden neurons, it lacks the capacity to learn enough. As the number of neurons increases, the NN begins to perform better.

- Accuracy when hidden layer size: (130,66) : 0.8561
- Accuracy when hidden layer size: (32,16): 0.8012
- Accuracy when hidden layer size: (65,32): 0.8553

The [130,36] layers perform the best. The most likely explanation is that the number of neurons required for learning is ideal. More neurons may result in the learning of irrelevant features, while fewer neurons may result in the learning of less relevant ones.

In [7]:

```
clf = MLPRegressor(max_iter=50)
```

In [9]:

```
parameter_space = {
    'hidden_layer_sizes': [(256, 32), (128, 16), (64, 28), (32, 4)],
    'activation': ['tanh', 'relu', 'logistic', 'identity'],
    'batch_size': [20, 35, 50, 150],
    'max_iter': [25, 50]

}
```

In [12]:

```
clf_new = GridSearchCV(clf, parameter_space, n_jobs=-1, cv=3).fit(X_train, y_train)
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (50) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

```
/home/dheeraj20194/.conda/envs/dheeraj_new_env/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:702: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (25) reached and the optimization hasn't converged yet.
```

```
warnings.warn(
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```
'max_iter': [100,120]

}
```

In []:

```
clf_new = GridSearchCV(clf, parameter_space, n_jobs=-1, cv=3).fit(X_train, y_train)
```

In [17]:

```
print('Best parameters found:\n', clf_new.best_params_)
```

Best parameters found:

```
{'activation': 'relu', 'batch_size': 150, 'hidden_layer_sizes': (256, 32), 'max_iter': 120}
```

In []:

```
clf = MLPRegressor()
parameter_space = {
    'hidden_layer_sizes': [(256, 32), (128, 16)],
    'activation': ['tanh', 'relu', 'logistic', 'identity'],
    'batch_size': [20, 35, 50, 150],
    'solver': ['sgd', 'adam']

}
```

```
clf_new = GridSearchCV(clf, parameter_space, n_jobs=-1, cv=3).fit(X_train, y_train)
```

In [19]:

```
print('Best parameters found:\n', clf_new.best_params_)
```

Best parameters found:

```
{'activation': 'relu', 'batch_size': 150, 'hidden_layer_sizes': (256, 32), 'solver': 'adam'}
```

Result

Grid Search uses a different combination of all the specified hyperparameters and their values and calculates the performance for each combination and selects the best value for the hyperparameters

- **Maybe with more epochs we could get a desired result, but since i was low on computation power. I used only 50 to 120 epochs for these experiments which also take 8 hour minimum to run and max 16 hour**

So we can say that:

- **max_iter : 120 , Learning rate : 0.001 , Activation Function : Relu, Batch size : 150 , Hidden layer size : (256,32) , solver : adam**

All this are best parameter for our case which is matching the above part A ,B,C