In [3]:

```
import numpy as np
import matplotlib.pyplot as plt
import utilis as u
import importlib
import model as m
import mlp as ml
```

```
Using TensorFlow backend.
```

load data

In [4]:

```
filename = '/data_batch_1'
X_train, y_train,Y_train = u.load_data(filename, reshape=False, clipping=True)
meanX = np.mean(X_train,axis=1)
stdX = np.std(X_train,axis=1)
X_train = (X_train-meanX.reshape((len(meanX),1)))/stdX.reshape((len(stdX),1))

filename = '/data_batch_2'
X_val, y_val,Y_val = u.load_data(filename, reshape=False, clipping=True)
X_val = (X_val-meanX.reshape((len(meanX),1)))/stdX.reshape((len(stdX),1))

filename = '/test_batch'
X_test, y_test,Y_test = u.load_data(filename, reshape=False, clipping=True)
X_test = (X_test-meanX.reshape((len(meanX),1)))/stdX.reshape((len(stdX),1))
```

In [32]:

```
data = {'X_train':X_train, 'Y_train':Y_train, 'y_train':y_train,'X_val':X_val, 'Y_val':
Y_val, 'y_val':y_val}
```

test gradients

In [51]:

```
importlib.reload(m)
mlp = m.MLP(dimensions=[100,50,10])
P = mlp.forward(X_train[:100,:1])
mlp.compute_gradients(X_train[:100,:1], Y_train[:,:1],P)
mlp.compareGradients(X_train[:100,:1],Y_train[:,:1])
for lmbda in [0,0.001,0.1]:
    mlp = m.MLP(dimensions=[100,50,10],lambda_=lmbda)
    P = mlp.forward(X_train[:100,:1])
    mlp.compute_gradients(X_train[:100,:1], Y_train[:,:1],P)
    rerr_w, rerr_b, aerr_w, aerr_b = mlp.compareGradients(X_train[:100,:1],Y_train[:,:1
])
    print(f"{rerr_w[0]} & {rerr_b[0]} & {rerr_w[1]} & {rerr_b[1]}")
```

```
2.175e-09 & 7.7e-11 & 1.77e-10 & 5.7e-11
5.7163e-08 & 7.7e-11 & 2.427e-08 & 5.7e-11
9.088e-09 & 2.38e-10 & 2.869e-09 & 1.66e-10
```

train Model 1

In [33]:

```
importlib.reload(m)
for seed in [1,2,3,4,5]:
    mlp = m.MLP(lambda_=0.01, seed=seed)
    GD_params = {"n_batch":100, "eta_min":1e-5, 'eta_max':1e-1, 'ns':500, 'n_cycles':1,
'freq':10}
    mlp.cyclicLearning(data, GD_params, 'cyclic_learning_ex1', True, True)
```

```
  0%|          | 0/10 [00:00<?, ?it/s]   Epoch 0: train_cost = 3.087885073
783367, val_cost = 3.0891523362602586,
         train_acc = 0.1187, val_acc = 0.1186

 10%|▉         | 1/10 [00:06<01:02,  6.97s/it]   Epoch 100: train_cost =
2.383763526802096, val_cost = 2.4574480863937884,
         train_acc = 0.3699, val_acc = 0.3394

 20%|██        | 2/10 [00:12<00:51,  6.45s/it]   Epoch 200: train_cost =
2.1382396086814266, val_cost = 2.28995103706939,
         train_acc = 0.4386, val_acc = 0.3782

 30%|███       | 3/10 [00:17<00:42,  6.06s/it]   Epoch 300: train_cost =
1.9892566154984441, val_cost = 2.1778864009526804,
         train_acc = 0.4609, val_acc = 0.3919

 40%|████      | 4/10 [00:22<00:34,  5.82s/it]   Epoch 400: train_cost =
1.8979753118928058, val_cost = 2.1135588609197216,
         train_acc = 0.4623, val_acc = 0.3933

 50%|█████     | 5/10 [00:27<00:28,  5.67s/it]   Epoch 500: train_cost =
1.9307837420529559, val_cost = 2.1664724795328145,
         train_acc = 0.4618, val_acc = 0.3887

 60%|██████    | 6/10 [00:32<00:21,  5.48s/it]   Epoch 600: train_cost =
1.6756557801973972, val_cost = 1.9411776004013221,
         train_acc = 0.5098, val_acc = 0.4225

 70%|███████   | 7/10 [00:37<00:16,  5.34s/it]   Epoch 700: train_cost =
1.5802125993671072, val_cost = 1.8866851594493628,
         train_acc = 0.5476, val_acc = 0.4396

 80%|████████  | 8/10 [00:42<00:10,  5.22s/it]   Epoch 800: train_cost =
1.5047996845596778, val_cost = 1.8411404556252313,
         train_acc = 0.5789, val_acc = 0.4486

 90%|█████████ | 9/10 [00:48<00:05,  5.22s/it]   Epoch 900: train_cost =
1.4689807206823635, val_cost = 1.832957227195731,
         train_acc = 0.5927, val_acc = 0.4493

100%|██████████| 10/10 [00:53<00:00,  5.33s/it]

  0%|          | 0/10 [00:00<?, ?it/s]   Epoch 0: train_cost = 3.054741771
0216336, val_cost = 3.0598608845680197,
         train_acc = 0.1009, val_acc = 0.0954

 10%|▉         | 1/10 [00:08<01:14,  8.24s/it]   Epoch 100: train_cost =
2.3914036071940683, val_cost = 2.463372966718284,
         train_acc = 0.3744, val_acc = 0.3315

 20%|██        | 2/10 [00:14<01:01,  7.69s/it]   Epoch 200: train_cost =
2.1459342167927655, val_cost = 2.2941478398107042,
         train_acc = 0.4302, val_acc = 0.3768

 30%|███       | 3/10 [00:20<00:50,  7.22s/it]   Epoch 300: train_cost =
2.0046148754780537, val_cost = 2.1913884531998495,
         train_acc = 0.4568, val_acc = 0.3945

 40%|████      | 4/10 [00:26<00:39,  6.66s/it]   Epoch 400: train_cost =
1.909489040175192, val_cost = 2.1215139075541147,
         train_acc = 0.4554, val_acc = 0.387
```

```
 50%|████      | 5/10 [00:31<00:31,  6.28s/it]   Epoch 500: train_cost =
1.8267246966221444, val_cost = 2.0571867007721436,
         train_acc = 0.4752, val_acc = 0.399

 60%|█████     | 6/10 [00:36<00:24,  6.03s/it]   Epoch 600: train_cost =
1.6707454704671798, val_cost = 1.9306660235507198,
         train_acc = 0.5115, val_acc = 0.4253

 70%|██████    | 7/10 [00:42<00:17,  5.84s/it]   Epoch 700: train_cost =
1.5969706048277676, val_cost = 1.895168157683294,
         train_acc = 0.5407, val_acc = 0.4313

 80%|███████   | 8/10 [00:47<00:11,  5.71s/it]   Epoch 800: train_cost =
1.5072137871746425, val_cost = 1.840937047669858,
         train_acc = 0.5795, val_acc = 0.4401

 90%|████████  | 9/10 [00:53<00:05,  5.60s/it]   Epoch 900: train_cost =
1.4705077015885661, val_cost = 1.828393412324239,
         train_acc = 0.5931, val_acc = 0.4501

100%|██████████| 10/10 [00:58<00:00,  5.84s/it]

  0%|          | 0/10 [00:00<?, ?it/s]   Epoch 0: train_cost = 3.246513211
2539825, val_cost = 3.2453023915208803,
         train_acc = 0.0842, val_acc = 0.0881

 10%|▉         | 1/10 [00:05<00:47,  5.28s/it]   Epoch 100: train_cost =
2.4186351481339132, val_cost = 2.4922370898756734,
         train_acc = 0.3504, val_acc = 0.3188

 20%|█▊        | 2/10 [00:10<00:42,  5.27s/it]   Epoch 200: train_cost =
2.1599770767435817, val_cost = 2.318017066329073,
         train_acc = 0.427, val_acc = 0.369

 30%|██▋       | 3/10 [00:15<00:37,  5.31s/it]   Epoch 300: train_cost =
2.014485921545223, val_cost = 2.2061107038869467,
         train_acc = 0.451, val_acc = 0.3832

 40%|███▌      | 4/10 [00:21<00:32,  5.45s/it]   Epoch 400: train_cost =
1.8912897605824386, val_cost = 2.1115581420596077,
         train_acc = 0.4692, val_acc = 0.399

 50%|████▌     | 5/10 [00:27<00:27,  5.43s/it]   Epoch 500: train_cost =
1.8054848627852746, val_cost = 2.0418651800053516,
         train_acc = 0.4856, val_acc = 0.3982

 60%|█████▍    | 6/10 [00:32<00:21,  5.41s/it]   Epoch 600: train_cost =
1.6692151872287755, val_cost = 1.9359684169556564,
         train_acc = 0.5193, val_acc = 0.423

 70%|██████▎   | 7/10 [00:37<00:16,  5.41s/it]   Epoch 700: train_cost =
1.5916592522636122, val_cost = 1.898506276439245,
         train_acc = 0.5412, val_acc = 0.4353

 80%|███████▏  | 8/10 [00:43<00:10,  5.48s/it]   Epoch 800: train_cost =
1.5020801776589952, val_cost = 1.8401703742748756,
         train_acc = 0.5824, val_acc = 0.4465

 90%|████████▏ | 9/10 [00:49<00:05,  5.66s/it]   Epoch 900: train_cost =
1.4684386457567897, val_cost = 1.8295860324004025,
```

```
        train_acc = 0.5944, val_acc = 0.456

100%|██████████| 10/10 [00:55<00:00,  5.51s/it]

  0%|          | 0/10 [00:00<?, ?it/s]   Epoch 0: train_cost = 3.017894170
2125304, val_cost = 3.016581439574926,
        train_acc = 0.1017, val_acc = 0.1004

 10%|█         | 1/10 [00:05<00:50,  5.61s/it]   Epoch 100: train_cost =
2.3965656396277257, val_cost = 2.469760730965862,
        train_acc = 0.362, val_acc = 0.3354

 20%|██        | 2/10 [00:11<00:44,  5.61s/it]   Epoch 200: train_cost =
2.1539772507791684, val_cost = 2.299818490663023,
        train_acc = 0.4317, val_acc = 0.3715

 30%|███       | 3/10 [00:16<00:38,  5.57s/it]   Epoch 300: train_cost =
2.0480297323152, val_cost = 2.231785520133223,
        train_acc = 0.439, val_acc = 0.381

 40%|████      | 4/10 [00:22<00:33,  5.56s/it]   Epoch 400: train_cost =
1.9105013895328953, val_cost = 2.126365943741584,
        train_acc = 0.4568, val_acc = 0.3862

 50%|█████     | 5/10 [00:27<00:27,  5.60s/it]   Epoch 500: train_cost =
1.793871079020222, val_cost = 2.0268634343541088,
        train_acc = 0.4901, val_acc = 0.407

 60%|██████    | 6/10 [00:33<00:22,  5.66s/it]   Epoch 600: train_cost =
1.6865303604066684, val_cost = 1.9422627058930144,
        train_acc = 0.5072, val_acc = 0.4223

 70%|███████   | 7/10 [00:39<00:16,  5.60s/it]   Epoch 700: train_cost =
1.5901824077348614, val_cost = 1.8960703546012105,
        train_acc = 0.5416, val_acc = 0.4295

 80%|████████  | 8/10 [00:44<00:11,  5.59s/it]   Epoch 800: train_cost =
1.5022431722959888, val_cost = 1.841127182096286,
        train_acc = 0.5815, val_acc = 0.445

 90%|█████████ | 9/10 [00:50<00:05,  5.61s/it]   Epoch 900: train_cost =
1.4674810887344365, val_cost = 1.828514219068934,
        train_acc = 0.5972, val_acc = 0.4502

100%|██████████| 10/10 [00:56<00:00,  5.61s/it]

  0%|          | 0/10 [00:00<?, ?it/s]   Epoch 0: train_cost = 3.135510407
719734, val_cost = 3.141959603178897,
        train_acc = 0.1181, val_acc = 0.11

 10%|█         | 1/10 [00:05<00:51,  5.67s/it]   Epoch 100: train_cost =
2.4134933137737886, val_cost = 2.4820877756785573,
        train_acc = 0.3606, val_acc = 0.3246

 20%|██        | 2/10 [00:11<00:45,  5.75s/it]   Epoch 200: train_cost =
2.1548169491126568, val_cost = 2.3021406042014525,
        train_acc = 0.4295, val_acc = 0.3794

 30%|███       | 3/10 [00:17<00:40,  5.82s/it]   Epoch 300: train_cost =
1.9936667484787467, val_cost = 2.1852738992723713,
        train_acc = 0.4593, val_acc = 0.3856
```

```
 40%|████      | 4/10 [00:23<00:35,  5.96s/it]   Epoch 400: train_cost =
1.9250588543216638, val_cost = 2.146053345439446,
         train_acc = 0.4544, val_acc = 0.3869

 50%|█████     | 5/10 [00:31<00:32,  6.53s/it]   Epoch 500: train_cost =
1.8190897249814182, val_cost = 2.066496323232418,
         train_acc = 0.4787, val_acc = 0.3935

 60%|██████    | 6/10 [00:37<00:25,  6.27s/it]   Epoch 600: train_cost =
1.6772588524073824, val_cost = 1.9508074461042848,
         train_acc = 0.5129, val_acc = 0.4133

 70%|███████   | 7/10 [00:45<00:20,  6.75s/it]   Epoch 700: train_cost =
1.5722621558214351, val_cost = 1.8868143106236337,
         train_acc = 0.5418, val_acc = 0.4379

 80%|████████  | 8/10 [00:52<00:13,  6.97s/it]   Epoch 800: train_cost =
1.493706854033864, val_cost = 1.8398162885953901,
         train_acc = 0.5863, val_acc = 0.4468

 90%|█████████ | 9/10 [00:59<00:06,  6.77s/it]   Epoch 900: train_cost =
1.4548620151918146, val_cost = 1.8274104497666832,
         train_acc = 0.5997, val_acc = 0.4506

100%|██████████| 10/10 [01:05<00:00,  6.50s/it]
```

train Model 2

In [38]:

```
importlib.reload(m)
for seed in [1,2,3,4,5]:
    mlp = m.MLP(lambda_=0.01,seed=seed)
    GD_params = {"n_batch":100, "eta_min":1e-5, 'eta_max':1e-1, 'ns':800, 'n_cycles':3,
'freq':10}
    mlp.cyclicLearning(data, GD_params, 'cyclic_learning_ex2', True, True)
```

```
c = 0.6911, val_acc = 0.4649

 69%|██████▉   | 33/48 [02:42<01:11,  4.80s/it]         Epoch 160: train_
cost = 1.2877624732692872, val_cost = 1.8678564400960758,
         train_acc = 0.6837, val_acc = 0.4585

 71%|███████   | 34/48 [02:48<01:12,  5.15s/it]
 73%|███████▎  | 35/48 [02:50<00:56,  4.38s/it]         Epoch 320: train_
cost = 1.3332691706232866, val_cost = 1.9230267730103368,
         train_acc = 0.6614, val_acc = 0.4443

 75%|███████▌  | 36/48 [02:56<00:58,  4.86s/it]  Epoch 480: train_cost =
1.3668607985804977, val_cost = 1.9457343458661116,
         train_acc = 0.6453, val_acc = 0.4353

 77%|███████▋  | 37/48 [03:02<00:56,  5.17s/it]
 79%|███████▉  | 38/48 [03:05<00:43,  4.38s/it]         Epoch 640: train_
cost = 1.4960233182345868, val_cost = 2.0400693913635912,
         train_acc = 0.5765, val_acc = 0.407

 81%|████████  | 39/48 [03:11<00:43,  4.84s/it]
 83%|████████▎ | 40/48 [03:13<00:33,  4.20s/it]         Epoch 800: train_
cost = 1.6465450833838364, val_cost = 2.1053631717751062,
         train_acc = 0.5338, val_acc = 0.394

 85%|████████▌ | 41/48 [03:19<00:33,  4.72s/it]  Epoch 960: train_cost =
1.5911762078782936, val_cost = 2.065298053122249,
         train_acc = 0.5621, val_acc = 0.4025

 88%|████████▊ | 42/48 [03:25<00:30,  5.09s/it]
 90%|████████▉ | 43/48 [03:28<00:21,  4.32s/it]         Epoch 1120: train
_cost = 1.3688182650685603, val_cost = 1.8845888785222475,
         train_acc = 0.6449, val_acc = 0.4538

 92%|█████████▏| 44/48 [03:34<00:19,  4.80s/it]         Epoch 1280: train
_cost = 1.3254824012546043, val_cost = 1.882475922870935,
         train_acc = 0.6716, val_acc = 0.4554

 94%|█████████▍| 45/48 [03:40<00:15,  5.17s/it]
 96%|█████████▌| 46/48 [03:42<00:08,  4.40s/it]  Epoch 1440: train_cost =
1.2776714541605192, val_cost = 1.8773834831103762,
         train_acc = 0.6938, val_acc = 0.459

 98%|█████████▊| 47/48 [03:48<00:04,  4.85s/it]
100%|██████████| 48/48 [03:51<00:00,  4.82s/it]

  0%|          | 0/48 [00:00<?, ?it/s]   Epoch 0: train_cost = 3.246513211
2539825, val_cost = 3.2453023915208803,
         train_acc = 0.0842, val_acc = 0.0881

  2%|▏         | 1/48 [00:06<04:59,  6.38s/it]  Epoch 160: train_cost =
2.3082779769254014, val_cost = 2.39799075896617,
         train_acc = 0.3904, val_acc = 0.3646

  4%|▍         | 2/48 [00:12<04:52,  6.35s/it]
  6%|▋         | 3/48 [00:15<03:54,  5.21s/it]  Epoch 320: train_cost =
2.0495804703645297, val_cost = 2.225165986857247,
         train_acc = 0.4583, val_acc = 0.3927

  8%|▊         | 4/48 [00:21<03:59,  5.44s/it]  Epoch 480: train_cost =
1.8822855127480929, val_cost = 2.113433341026148,
```

```
        train_acc = 0.4815, val_acc = 0.3948

 10%|▉         | 5/48 [00:27<03:59,  5.58s/it]
 12%|▉         | 6/48 [00:29<03:16,  4.69s/it]  Epoch 640: train_cost =
1.7432406709072465, val_cost = 2.0143102275518903,
        train_acc = 0.5049, val_acc = 0.4034

 15%|▉         | 7/48 [00:35<03:27,  5.07s/it]
 17%|▉         | 8/48 [00:38<02:53,  4.33s/it]  Epoch 800: train_cost =
1.6505050421917193, val_cost = 1.9416086319267507,
        train_acc = 0.5225, val_acc = 0.4171

 19%|▉         | 9/48 [00:44<03:07,  4.80s/it]  Epoch 960: train_cost =
1.5482452298180154, val_cost = 1.884877126306992,
        train_acc = 0.5498, val_acc = 0.4259

 21%|▉         | 10/48 [00:50<03:15,  5.15s/it]
 23%|▉         | 11/48 [00:52<02:41,  4.38s/it]         Epoch 1120: train
_cost = 1.481636260686799, val_cost = 1.8599693403189828,
        train_acc = 0.5797, val_acc = 0.4351

 25%|▉         | 12/48 [00:58<02:54,  4.84s/it]  Epoch 1280: train_cost =
1.4244175284106932, val_cost = 1.8389025918024493,
        train_acc = 0.6058, val_acc = 0.4482

 27%|▉         | 13/48 [01:04<03:00,  5.17s/it]
 29%|▉         | 14/48 [01:07<02:28,  4.38s/it]         Epoch 1440: train
_cost = 1.3782974060559574, val_cost = 1.8252382793192585,
        train_acc = 0.6249, val_acc = 0.4544

 31%|▉         | 15/48 [01:13<02:39,  4.84s/it]
 33%|▉         | 16/48 [01:15<02:13,  4.17s/it]         Epoch 0: train_co
st = 1.337804224637523, val_cost = 1.804986020466942,
        train_acc = 0.647, val_acc = 0.4654

 35%|▉         | 17/48 [01:21<02:26,  4.72s/it]  Epoch 160: train_cost =
1.3526493830932114, val_cost = 1.8289253446958638,
        train_acc = 0.6437, val_acc = 0.4592

 38%|▉         | 18/48 [01:27<02:32,  5.08s/it]
 40%|▉         | 19/48 [01:30<02:05,  4.32s/it]         Epoch 320: train_
cost = 1.370699369854632, val_cost = 1.8663568075803898,
        train_acc = 0.6375, val_acc = 0.4449

 42%|▉         | 20/48 [01:36<02:14,  4.81s/it]         Epoch 480: train_
cost = 1.471748526761118, val_cost = 1.945563208825552,
        train_acc = 0.5846, val_acc = 0.4217

 44%|▉         | 21/48 [01:41<02:18,  5.13s/it]
 46%|▉         | 22/48 [01:44<01:53,  4.37s/it]  Epoch 640: train_cost =
1.5286565611401526, val_cost = 1.9947839215441634,
        train_acc = 0.5577, val_acc = 0.4037

 48%|▉         | 23/48 [01:50<02:00,  4.82s/it]
 50%|▉         | 24/48 [01:52<01:39,  4.14s/it]  Epoch 800: train_cost =
1.5316667187554573, val_cost = 1.9578822955075887,
        train_acc = 0.56, val_acc = 0.4167

 52%|▉         | 25/48 [01:58<01:47,  4.69s/it]         Epoch 960: train_
cost = 1.4785765513407023, val_cost = 1.9175046383886845,
        train_acc = 0.5866, val_acc = 0.4259
```

```
 54%|█████     | 26/48 [02:04<01:51,  5.05s/it]
 56%|█████▌    | 27/48 [02:07<01:30,  4.31s/it]         Epoch 1120: train
_cost = 1.3980188069051724, val_cost = 1.8783175354389263,
         train_acc = 0.62, val_acc = 0.4485

 58%|█████▊    | 28/48 [02:13<01:36,  4.81s/it]         Epoch 1280: train
_cost = 1.3544773191995316, val_cost = 1.8785193624481793,
         train_acc = 0.6414, val_acc = 0.4514

 60%|██████    | 29/48 [02:19<01:37,  5.12s/it]
 62%|██████▎   | 30/48 [02:21<01:18,  4.35s/it]         Epoch 1440: train
_cost = 1.300192381650958, val_cost = 1.8538501359356008,
         train_acc = 0.6766, val_acc = 0.4593

 65%|██████▍   | 31/48 [02:27<01:22,  4.82s/it]
 67%|██████▋   | 32/48 [02:30<01:06,  4.16s/it]         Epoch 0: train_co
st = 1.2700014922197203, val_cost = 1.8368110153667196,
         train_acc = 0.6967, val_acc = 0.4694

 69%|██████▉   | 33/48 [02:36<01:10,  4.69s/it]         Epoch 160: train_
cost = 1.2831174985924256, val_cost = 1.8543484749398893,
         train_acc = 0.6888, val_acc = 0.4664

 71%|███████   | 34/48 [02:42<01:10,  5.06s/it]
 73%|███████▎  | 35/48 [02:44<00:56,  4.32s/it]         Epoch 320: train_
cost = 1.329946770210396, val_cost = 1.9073203539611943,
         train_acc = 0.6604, val_acc = 0.4461

 75%|███████▌  | 36/48 [02:50<00:57,  4.80s/it]  Epoch 480: train_cost =
1.3716169143938186, val_cost = 1.9417851255846466,
         train_acc = 0.642, val_acc = 0.4436

 77%|███████▋  | 37/48 [02:57<00:57,  5.26s/it]
 79%|███████▉  | 38/48 [02:59<00:44,  4.45s/it]         Epoch 640: train_
cost = 1.467668596791103, val_cost = 2.0011095141251967,
         train_acc = 0.5957, val_acc = 0.4262

 81%|████████  | 39/48 [03:05<00:44,  4.90s/it]
 83%|████████▎ | 40/48 [03:08<00:33,  4.20s/it]         Epoch 800: train_
cost = 1.6568959216331316, val_cost = 2.087704465296719,
         train_acc = 0.522, val_acc = 0.3921

 85%|████████▌ | 41/48 [03:14<00:33,  4.71s/it]  Epoch 960: train_cost =
1.51631356593705, val_cost = 1.9923119738819604,
         train_acc = 0.5814, val_acc = 0.4185

 88%|████████▊ | 42/48 [03:19<00:30,  5.09s/it]
 90%|████████▉ | 43/48 [03:22<00:21,  4.34s/it]         Epoch 1120: train
_cost = 1.3778195600035488, val_cost = 1.8921703995599684,
         train_acc = 0.6379, val_acc = 0.451

 92%|█████████▏| 44/48 [03:28<00:19,  4.83s/it]         Epoch 1280: train
_cost = 1.317731293270429, val_cost = 1.868440705041928,
         train_acc = 0.6808, val_acc = 0.4625

 94%|█████████▍| 45/48 [03:34<00:15,  5.19s/it]
 96%|█████████▌| 46/48 [03:38<00:09,  4.78s/it]  Epoch 1440: train_cost =
1.2767497319773151, val_cost = 1.8592001301385448,
         train_acc = 0.7006, val_acc = 0.4685
```

```
 98%|▉▉▉▉▉▉▉▉▉ | 47/48 [03:44<00:05,  5.16s/it]
100%|▉▉▉▉▉▉▉▉▉| 48/48 [03:46<00:00,  4.73s/it]

  0%|          | 0/48 [00:00<?, ?it/s]   Epoch 0: train_cost = 3.017894170
2125304, val_cost = 3.016581439574926,
         train_acc = 0.1017, val_acc = 0.1004

  2%|▏          | 1/48 [00:06<04:46,  6.09s/it]  Epoch 160: train_cost =
2.322414136863542, val_cost = 2.411747315497604,
         train_acc = 0.3799, val_acc = 0.346

  4%|▍          | 2/48 [00:12<04:38,  6.06s/it]
  6%|▋          | 3/48 [00:14<03:46,  5.04s/it]  Epoch 320: train_cost =
2.050870146413944, val_cost = 2.233458299403188,
         train_acc = 0.4556, val_acc = 0.3936

  8%|▉          | 4/48 [00:20<03:54,  5.33s/it]  Epoch 480: train_cost =
1.868982827822487, val_cost = 2.095758431302961,
         train_acc = 0.4844, val_acc = 0.4022

 10%|█          | 5/48 [00:26<03:58,  5.54s/it]
 12%|█▎         | 6/48 [00:29<03:15,  4.66s/it]  Epoch 640: train_cost =
1.7326954803859895, val_cost = 2.0063749496319585,
         train_acc = 0.5099, val_acc = 0.4088

 15%|█▌         | 7/48 [00:35<03:27,  5.07s/it]
 17%|█▊         | 8/48 [00:37<02:52,  4.32s/it]  Epoch 800: train_cost =
1.6759019327909617, val_cost = 1.9697627087408993,
         train_acc = 0.512, val_acc = 0.4071

 19%|██         | 9/48 [00:43<03:07,  4.82s/it]  Epoch 960: train_cost =
1.549459694470854, val_cost = 1.8804946625847663,
         train_acc = 0.5522, val_acc = 0.4273

 21%|██▏        | 10/48 [00:49<03:15,  5.15s/it]
 23%|██▍        | 11/48 [00:52<02:42,  4.39s/it]         Epoch 1120: train
_cost = 1.4797402929457522, val_cost = 1.8586190459431429,
         train_acc = 0.5775, val_acc = 0.4381

 25%|██▋        | 12/48 [00:58<02:55,  4.87s/it]  Epoch 1280: train_cost =
1.4224616893175013, val_cost = 1.8408280845262204,
         train_acc = 0.6062, val_acc = 0.4447

 27%|██▉        | 13/48 [01:05<03:08,  5.38s/it]
 29%|███        | 14/48 [01:07<02:36,  4.61s/it]         Epoch 1440: train
_cost = 1.3735128792709363, val_cost = 1.8201640764559326,
         train_acc = 0.6286, val_acc = 0.4511

 31%|███▎       | 15/48 [01:14<02:49,  5.12s/it]
 33%|███▌       | 16/48 [01:16<02:20,  4.39s/it]         Epoch 0: train_co
st = 1.3390812120403393, val_cost = 1.8042562221663656,
         train_acc = 0.6528, val_acc = 0.4583

 35%|███▊       | 17/48 [01:22<02:31,  4.87s/it]  Epoch 160: train_cost =
1.351335050715528, val_cost = 1.8273894127454449,
         train_acc = 0.6441, val_acc = 0.4513

 38%|████       | 18/48 [01:28<02:30,  5.02s/it]
 40%|████▏      | 19/48 [01:30<02:02,  4.21s/it]         Epoch 320: train_
cost = 1.3741987522120203, val_cost = 1.8647090948468317,
         train_acc = 0.6419, val_acc = 0.4475
```

```
 42%|████      | 20/48 [01:35<02:07,  4.54s/it]         Epoch 480: train_
cost = 1.4541866592726773, val_cost = 1.928159541063006,
         train_acc = 0.5964, val_acc = 0.427

 44%|████      | 21/48 [01:41<02:10,  4.84s/it]
 46%|████      | 22/48 [01:43<01:45,  4.08s/it]  Epoch 640: train_cost =
1.5561500386696006, val_cost = 2.009374862740857,
         train_acc = 0.5519, val_acc = 0.4012

 48%|████      | 23/48 [01:48<01:50,  4.43s/it]
 50%|████      | 24/48 [01:51<01:31,  3.79s/it]  Epoch 800: train_cost =
1.4784877888699233, val_cost = 1.9142285753212473,
         train_acc = 0.5887, val_acc = 0.4331

 52%|████      | 25/48 [01:56<01:36,  4.21s/it]         Epoch 960: train_
cost = 1.4846686483066012, val_cost = 1.9194862161447999,
         train_acc = 0.5781, val_acc = 0.4283

 54%|████      | 26/48 [02:01<01:39,  4.53s/it]
 56%|████      | 27/48 [02:04<01:21,  3.86s/it]         Epoch 1120: train
_cost = 1.395601140132363, val_cost = 1.8688040781333042,
         train_acc = 0.6285, val_acc = 0.4441

 58%|████      | 28/48 [02:09<01:25,  4.27s/it]         Epoch 1280: train
_cost = 1.3442816410528484, val_cost = 1.8685307194082539,
         train_acc = 0.6498, val_acc = 0.453

 60%|████      | 29/48 [02:14<01:26,  4.56s/it]
 62%|████      | 30/48 [02:16<01:09,  3.89s/it]         Epoch 1440: train
_cost = 1.3015136212924847, val_cost = 1.8543328270013073,
         train_acc = 0.6786, val_acc = 0.4574

 65%|████      | 31/48 [02:22<01:13,  4.30s/it]
 67%|████      | 32/48 [02:24<00:59,  3.70s/it]         Epoch 0: train_co
st = 1.2680864076944824, val_cost = 1.8337372239158287,
         train_acc = 0.6999, val_acc = 0.4689

 69%|████      | 33/48 [02:29<01:02,  4.17s/it]         Epoch 160: train_
cost = 1.2891128384946369, val_cost = 1.8590428050922025,
         train_acc = 0.6829, val_acc = 0.461

 71%|████      | 34/48 [02:34<01:02,  4.50s/it]
 73%|████      | 35/48 [02:37<00:49,  3.84s/it]         Epoch 320: train_
cost = 1.3249121115000004, val_cost = 1.9056595708670527,
         train_acc = 0.6649, val_acc = 0.4514

 75%|████      | 36/48 [02:42<00:51,  4.27s/it]  Epoch 480: train_cost =
1.4058488125438957, val_cost = 1.9714625005268254,
         train_acc = 0.624, val_acc = 0.4285

 77%|████      | 37/48 [02:47<00:50,  4.57s/it]
 79%|████      | 38/48 [02:50<00:38,  3.89s/it]         Epoch 640: train_
cost = 1.4495809267571658, val_cost = 1.9935169080713888,
         train_acc = 0.601, val_acc = 0.4261

 81%|████      | 39/48 [02:55<00:38,  4.30s/it]
 83%|████      | 40/48 [02:57<00:29,  3.70s/it]         Epoch 800: train_
cost = 1.76341675635293, val_cost = 2.1977398705527897,
         train_acc = 0.5007, val_acc = 0.3734
```

```
 85%|████████  | 41/48 [03:02<00:29,  4.16s/it]  Epoch 960: train_cost =
1.5560937250394797, val_cost = 2.0274742127621184,
         train_acc = 0.5677, val_acc = 0.4043

 88%|████████▊ | 42/48 [03:08<00:27,  4.51s/it]
 90%|█████████ | 43/48 [03:10<00:19,  3.85s/it]         Epoch 1120: train
_cost = 1.3711659002484047, val_cost = 1.8845309294586032,
         train_acc = 0.6404, val_acc = 0.4559

 92%|█████████▏| 44/48 [03:15<00:17,  4.26s/it]         Epoch 1280: train
_cost = 1.316900668441122, val_cost = 1.8680362023389925,
         train_acc = 0.6755, val_acc = 0.4604

 94%|█████████▍| 45/48 [03:20<00:13,  4.55s/it]
 96%|█████████▌| 46/48 [03:23<00:07,  3.89s/it]  Epoch 1440: train_cost =
1.2742164022843179, val_cost = 1.8621394085807315,
         train_acc = 0.7007, val_acc = 0.4659

 98%|█████████▊| 47/48 [03:28<00:04,  4.28s/it]
100%|██████████| 48/48 [03:30<00:00,  4.39s/it]

  0%|          | 0/48 [00:00<?, ?it/s]   Epoch 0: train_cost = 3.135510407
719734, val_cost = 3.141959603178897,
         train_acc = 0.1181, val_acc = 0.11

  2%|▏         | 1/48 [00:05<04:10,  5.33s/it]  Epoch 160: train_cost =
2.3262084302795816, val_cost = 2.4224666352218724,
         train_acc = 0.3861, val_acc = 0.3489

  4%|▍         | 2/48 [00:10<04:03,  5.30s/it]
  6%|▋         | 3/48 [00:12<03:17,  4.39s/it]  Epoch 320: train_cost =
2.070174418268555, val_cost = 2.2459572783435746,
         train_acc = 0.4442, val_acc = 0.3843

  8%|▊         | 4/48 [00:19<03:42,  5.05s/it]  Epoch 480: train_cost =
1.870658837424456, val_cost = 2.1066322369218042,
         train_acc = 0.4833, val_acc = 0.3955

 10%|▉         | 5/48 [00:24<03:42,  5.17s/it]
 12%|█▎        | 6/48 [00:27<03:04,  4.38s/it]  Epoch 640: train_cost =
1.7307568279193317, val_cost = 2.0074762026463957,
         train_acc = 0.5157, val_acc = 0.4131

 15%|█▍        | 7/48 [00:33<03:15,  4.77s/it]
 17%|█▋        | 8/48 [00:35<02:42,  4.05s/it]  Epoch 800: train_cost =
1.659125201866435, val_cost = 1.957902310275101,
         train_acc = 0.5238, val_acc = 0.4144

 19%|█▉        | 9/48 [00:40<02:51,  4.41s/it]  Epoch 960: train_cost =
1.5440122907143021, val_cost = 1.8912368784213933,
         train_acc = 0.5556, val_acc = 0.4246

 21%|██        | 10/48 [00:45<02:57,  4.67s/it]
 23%|██▎       | 11/48 [00:48<02:26,  3.96s/it]         Epoch 1120: train
_cost = 1.4630948375131791, val_cost = 1.851045993780677,
         train_acc = 0.5889, val_acc = 0.4382

 25%|██▌       | 12/48 [00:53<02:36,  4.35s/it]  Epoch 1280: train_cost =
1.4138567510080646, val_cost = 1.845184706148546,
         train_acc = 0.6106, val_acc = 0.4514
```

```
 27%|██▋       | 13/48 [00:58<02:43,  4.66s/it]
 29%|██▉       | 14/48 [01:01<02:14,  3.96s/it]         Epoch 1440: train
_cost = 1.3667008057887702, val_cost = 1.8296109864309973,
         train_acc = 0.6338, val_acc = 0.4535

 31%|███       | 15/48 [01:06<02:23,  4.35s/it]
 33%|███▎      | 16/48 [01:08<01:59,  3.73s/it]         Epoch 0: train_co
st = 1.3295481153126056, val_cost = 1.8105912472947534,
         train_acc = 0.6579, val_acc = 0.4658

 35%|███▌      | 17/48 [01:15<02:21,  4.57s/it]  Epoch 160: train_cost =
1.3435494616194148, val_cost = 1.8360880593714914,
         train_acc = 0.6447, val_acc = 0.4571

 38%|███▊      | 18/48 [01:22<02:36,  5.23s/it]
 40%|███▉      | 19/48 [01:24<02:06,  4.35s/it]         Epoch 320: train_
cost = 1.3644137898282607, val_cost = 1.873020880640674,
         train_acc = 0.6396, val_acc = 0.4487

 42%|████      | 20/48 [01:29<02:09,  4.63s/it]         Epoch 480: train_
cost = 1.4352795314500495, val_cost = 1.9355486228026066,
         train_acc = 0.602, val_acc = 0.4296

 44%|████▍     | 21/48 [01:34<02:09,  4.79s/it]
 46%|████▌     | 22/48 [01:37<01:45,  4.04s/it]  Epoch 640: train_cost =
1.5709562856337014, val_cost = 2.047092874212683,
         train_acc = 0.5545, val_acc = 0.3995

 48%|████▊     | 23/48 [01:42<01:50,  4.40s/it]
 50%|█████     | 24/48 [01:44<01:30,  3.76s/it]  Epoch 800: train_cost =
1.5043647209920061, val_cost = 1.9335360779780624,
         train_acc = 0.5721, val_acc = 0.4236

 52%|█████▏    | 25/48 [01:49<01:36,  4.19s/it]         Epoch 960: train_
cost = 1.4905583898992028, val_cost = 1.9438429390604033,
         train_acc = 0.5745, val_acc = 0.421

 54%|█████▍    | 26/48 [01:56<01:46,  4.82s/it]
 56%|█████▋    | 27/48 [01:58<01:25,  4.06s/it]         Epoch 1120: train
_cost = 1.4115196830938657, val_cost = 1.8914111652663628,
         train_acc = 0.6203, val_acc = 0.4438

 58%|█████▊    | 28/48 [02:03<01:28,  4.41s/it]         Epoch 1280: train
_cost = 1.3393807431428844, val_cost = 1.8836607964653742,
         train_acc = 0.6513, val_acc = 0.4516

 60%|██████    | 29/48 [02:08<01:28,  4.67s/it]
 62%|██████▎   | 30/48 [02:11<01:11,  3.96s/it]         Epoch 1440: train
_cost = 1.3017317498386052, val_cost = 1.8727422633601605,
         train_acc = 0.6832, val_acc = 0.4542

 65%|██████▍   | 31/48 [02:16<01:13,  4.33s/it]
 67%|██████▋   | 32/48 [02:18<00:59,  3.74s/it]         Epoch 0: train_co
st = 1.2627530993567853, val_cost = 1.8469044385716837,
         train_acc = 0.7068, val_acc = 0.4676

 69%|██████▉   | 33/48 [02:23<01:02,  4.18s/it]         Epoch 160: train_
cost = 1.280751136626574, val_cost = 1.868741113182531,
         train_acc = 0.6931, val_acc = 0.4604

 71%|███████   | 34/48 [02:29<01:02,  4.49s/it]
```

```
 73%|███████   | 35/48 [02:31<00:49,  3.84s/it]         Epoch 320: train_
cost = 1.3244850459786988, val_cost = 1.9176256709088357,
         train_acc = 0.6625, val_acc = 0.4506

 75%|███████   | 36/48 [02:36<00:50,  4.24s/it]  Epoch 480: train_cost =
1.4003498683042883, val_cost = 1.9828512528369524,
         train_acc = 0.6236, val_acc = 0.4262

 77%|███████▋  | 37/48 [02:42<00:51,  4.64s/it]
 79%|███████▉  | 38/48 [02:44<00:39,  3.93s/it]         Epoch 640: train_
cost = 1.5021323806904863, val_cost = 2.0460730807304053,
         train_acc = 0.5814, val_acc = 0.405

 81%|████████  | 39/48 [02:49<00:38,  4.31s/it]
 83%|████████▎ | 40/48 [02:52<00:29,  3.70s/it]         Epoch 800: train_
cost = 1.7263578473647079, val_cost = 2.1919609225771244,
         train_acc = 0.5199, val_acc = 0.3829

 85%|████████▌ | 41/48 [02:57<00:29,  4.15s/it]  Epoch 960: train_cost =
1.6225726002225989, val_cost = 2.110062211861977,
         train_acc = 0.5549, val_acc = 0.3959

 88%|████████▊ | 42/48 [03:02<00:26,  4.48s/it]
 90%|████████▉ | 43/48 [03:04<00:19,  3.81s/it]         Epoch 1120: train
_cost = 1.365743286136856, val_cost = 1.8946923759910284,
         train_acc = 0.6477, val_acc = 0.4519

 92%|█████████▏| 44/48 [03:10<00:17,  4.28s/it]         Epoch 1280: train
_cost = 1.3149331507003337, val_cost = 1.876550451162703,
         train_acc = 0.6794, val_acc = 0.4579

 94%|█████████▍| 45/48 [03:15<00:13,  4.57s/it]
 96%|█████████▌| 46/48 [03:17<00:07,  3.89s/it]  Epoch 1440: train_cost =
1.271372419176326, val_cost = 1.8731515719273835,
         train_acc = 0.7045, val_acc = 0.4668

 98%|█████████▊| 47/48 [03:22<00:04,  4.27s/it]
100%|██████████| 48/48 [03:25<00:00,  4.27s/it]
```

In [7]:

```
# Load all data
```

In [24]:

```
importlib.reload(m)
X_train_whole, y_train_whole, Y_train_whole = u.load_data('/data_batch_1', clipping=Tru
e)
for i in range(2,6):
    X, y, Y = u.load_data('/data_batch_'+str(i), clipping=True)
    X_train_whole = np.concatenate((X, X_train_whole), axis=1)
    y_train_whole = np.concatenate((y, y_train_whole))
    Y_train_whole = np.concatenate((Y, Y_train_whole), axis=1)

X_val_small, y_val_small, Y_val_small = X_train_whole[:,-5000:], y_train_whole[-5000:],
Y_train_whole[:,-5000:]
X_train_whole, y_train_whole, Y_train_whole = X_train_whole[:,:-5000], y_train_whole[:-
5000], Y_train_whole[:,:-5000]

## normalize with mean and std of train set
mean = np.mean(X_train_whole, axis=1)
std = np.std(X_train_whole, axis=1)

X_train_whole -= np.outer(mean, np.ones(X_train_whole.shape[1]))
X_train_whole /= np.outer(std, np.ones(X_train_whole.shape[1]))

X_val_small -= np.outer(mean, np.ones(X_val_small.shape[1]))
X_val_small /= np.outer(std, np.ones(X_val_small.shape[1]))
```

In [25]:

```
data = {'X_train':X_train_whole, 'Y_train':Y_train_whole, 'y_train':y_train_whole,'X_va
l':X_val_small, 'Y_val':Y_val_small, 'y_val':y_val_small}
```

Do lambda search

In [39]:

```
importlib.reload(m)
GD_params = {"epochs":2, "n_batch":100, "eta_min":1e-5,'eta_max':1e-1, 'ns':2*np.floor(
X_train_whole.shape[1]/100), 'n_cycles':2, 'freq':10}
search = m.LambdaSearch(-5, -1,10)
model = search.lambda_search(data=data,GDparams=GD_params)
```

```
  0%|          | 0/8 [00:00<?, ?it/s]
 12%|▉         | 1/8 [00:54<06:18, 54.04s/it]
 25%|██        | 2/8 [01:29<04:50, 48.35s/it]
 38%|███▏      | 3/8 [02:11<03:52, 46.43s/it]
 50%|████      | 4/8 [02:40<02:44, 41.21s/it]
 62%|████▉     | 5/8 [03:18<02:00, 40.31s/it]
 75%|██████    | 6/8 [03:48<01:14, 37.34s/it]
 88%|███████▏  | 7/8 [04:28<00:38, 38.19s/it]
100%|████████| 8/8 [05:02<00:00, 37.79s/it]

  0%|          | 0/8 [00:00<?, ?it/s]
 12%|▉         | 1/8 [00:42<04:55, 42.16s/it]
 25%|██        | 2/8 [01:14<03:54, 39.15s/it]
 38%|███▏      | 3/8 [01:53<03:15, 39.10s/it]
 50%|████      | 4/8 [02:22<02:24, 36.01s/it]
 62%|████▉     | 5/8 [03:00<01:50, 36.67s/it]
 75%|██████    | 6/8 [03:32<01:10, 35.44s/it]
 88%|███████▏  | 7/8 [04:12<00:36, 36.78s/it]
100%|████████| 8/8 [04:42<00:00, 35.32s/it]

  0%|          | 0/8 [00:00<?, ?it/s]
 12%|▉         | 1/8 [00:40<04:43, 40.53s/it]
 25%|██        | 2/8 [01:09<03:43, 37.21s/it]
 38%|███▏      | 3/8 [01:52<03:14, 38.84s/it]
 50%|████      | 4/8 [02:25<02:27, 36.99s/it]
 62%|████▉     | 5/8 [03:09<01:57, 39.29s/it]
 75%|██████    | 6/8 [03:43<01:14, 37.47s/it]
 88%|███████▏  | 7/8 [04:25<00:38, 38.78s/it]
100%|████████| 8/8 [04:57<00:00, 37.21s/it]

  0%|          | 0/8 [00:00<?, ?it/s]
 12%|▉         | 1/8 [00:38<04:31, 38.76s/it]
 25%|██        | 2/8 [01:08<03:36, 36.15s/it]
 38%|███▏      | 3/8 [01:48<03:06, 37.24s/it]
 50%|████      | 4/8 [02:21<02:23, 35.79s/it]
 62%|████▉     | 5/8 [02:59<01:49, 36.46s/it]
 75%|██████    | 6/8 [03:29<01:09, 34.72s/it]
 88%|███████▏  | 7/8 [04:08<00:35, 35.86s/it]
100%|████████| 8/8 [04:39<00:00, 34.96s/it]

  0%|          | 0/8 [00:00<?, ?it/s]
 12%|▉         | 1/8 [00:36<04:16, 36.70s/it]
 25%|██        | 2/8 [01:05<03:26, 34.35s/it]
 38%|███▏      | 3/8 [01:42<02:55, 35.01s/it]
 50%|████      | 4/8 [02:12<02:15, 33.76s/it]
 62%|████▉     | 5/8 [02:50<01:44, 34.95s/it]
 75%|██████    | 6/8 [03:19<01:06, 33.05s/it]
 88%|███████▏  | 7/8 [03:59<00:35, 35.18s/it]
100%|████████| 8/8 [04:33<00:00, 34.18s/it]

  0%|          | 0/8 [00:00<?, ?it/s]
 12%|▉         | 1/8 [00:42<04:57, 42.45s/it]
 25%|██        | 2/8 [01:11<03:50, 38.47s/it]
 38%|███▏      | 3/8 [01:50<03:13, 38.69s/it]
 50%|████      | 4/8 [02:18<02:21, 35.41s/it]
 62%|████▉     | 5/8 [02:54<01:46, 35.66s/it]
 75%|██████    | 6/8 [03:25<01:08, 34.10s/it]
 88%|███████▏  | 7/8 [04:00<00:34, 34.57s/it]
100%|████████| 8/8 [04:28<00:00, 33.62s/it]

  0%|          | 0/8 [00:00<?, ?it/s]
```

```
 12%|█▎          | 1/8 [00:35<04:08, 35.54s/it]
 25%|██▌         | 2/8 [01:03<03:19, 33.24s/it]
 38%|███▊        | 3/8 [01:39<02:49, 33.99s/it]
 50%|█████       | 4/8 [02:07<02:08, 32.16s/it]
 62%|██████▎     | 5/8 [02:42<01:39, 33.23s/it]
 75%|███████▌    | 6/8 [03:17<01:07, 33.73s/it]
 88%|████████▊   | 7/8 [03:54<00:34, 34.73s/it]
100%|██████████| 8/8 [04:23<00:00, 32.90s/it]

  0%|          | 0/8 [00:00<?, ?it/s]
 12%|█▎          | 1/8 [00:36<04:14, 36.38s/it]
 25%|██▌         | 2/8 [01:04<03:23, 33.97s/it]
 38%|███▊        | 3/8 [01:41<02:54, 34.91s/it]
 50%|█████       | 4/8 [02:10<02:12, 33.05s/it]
 62%|██████▎     | 5/8 [02:46<01:42, 34.06s/it]
 75%|███████▌    | 6/8 [03:19<01:06, 33.48s/it]
 88%|████████▊   | 7/8 [03:57<00:34, 34.88s/it]
100%|██████████| 8/8 [04:26<00:00, 33.32s/it]

  0%|          | 0/8 [00:00<?, ?it/s]
 12%|█▎          | 1/8 [00:36<04:13, 36.22s/it]
 25%|██▌         | 2/8 [01:03<03:21, 33.64s/it]
 38%|███▊        | 3/8 [01:52<03:10, 38.13s/it]
 50%|█████       | 4/8 [02:25<02:27, 36.75s/it]
 62%|██████▎     | 5/8 [03:17<02:03, 41.08s/it]
 75%|███████▌    | 6/8 [03:51<01:18, 39.17s/it]
 88%|████████▊   | 7/8 [04:30<00:39, 39.06s/it]
100%|██████████| 8/8 [05:00<00:00, 37.55s/it]

  0%|          | 0/8 [00:00<?, ?it/s]
 12%|█▎          | 1/8 [00:41<04:47, 41.06s/it]
 25%|██▌         | 2/8 [01:13<03:50, 38.41s/it]
 38%|███▊        | 3/8 [01:59<03:23, 40.80s/it]
 50%|█████       | 4/8 [02:35<02:37, 39.42s/it]
 62%|██████▎     | 5/8 [03:13<01:56, 38.93s/it]
 75%|███████▌    | 6/8 [03:42<01:11, 35.77s/it]
 88%|████████▊   | 7/8 [04:18<00:36, 36.00s/it]
100%|██████████| 8/8 [04:47<00:00, 35.97s/it]
```

In [45]:

```
importlib.reload(m)
for key in model.keys():
    train_loss_ = np.array(model[key].train_loss)
    val_loss_ = np.array(model[key].val_loss)
    train_acc_ = np.array(model[key].train_acc)
    val_acc_ = np.array(model[key].val_acc)
    train_cost_ = np.array(model[key].train_cost)
    val_cost_ = np.array(model[key].val_cost)
    hist = {'train_loss':train_loss_, 'val_loss':val_loss_, 'train_acc':train_acc_, 'va
l_acc':val_acc_,'train_cost':train_cost_,'val_cost':val_cost_}
    np.save(f"Models/lambda/hist_{model[key].lambda_}.npy",hist)
```

In [22]:

```
lmda = [0.1,0.03593813663804626,0.012915496650148827,0.004641588833612777,0.00166810053
72000592,0.0005994842503189409,7.742636826811278e-05,2.782559402207126e-05,1e-05]
print(f"{np.round(lmda[0],5)} & {np.round(lmda[1],5)} & {np.round(lmda[2],5)} & {np.rou
nd(lmda[3],5)} & {np.round(lmda[4],5)} & {np.round(lmda[5],5)} & {np.round(lmda[6],5)}
 & {np.round(lmda[7],5)} & {np.round(lmda[8],5)} ")
```

```
0.1 & 0.03594 & 0.01292 & 0.00464 & 0.00167 & 0.0006 & 8e-05 & 3e-05 & 1e-
05
```

In [23]:

```
importlib.reload(m)
val_acc = []
for lamda in lmda:
    GD_params = {"epochs":10, "n_batch":100, "eta_min":0.001,'eta_max':1e-1, 'ns':2*np.
floor(49000/100), 'n_cycles':2, 'freq':10,'lambda':lamda,'seed':42}
    hist = np.load(f'Models/lambda/hist_{lamda}.npy',allow_pickle=True)
    val_acc.append(np.max(hist.item()['val_acc']))
print(f"{np.round(val_acc[0],5)} & {np.round(val_acc[1],5)} & {np.round(val_acc[2],5)}
 & {np.round(val_acc[3],5)} & {np.round(val_acc[4],5)} & {np.round(val_acc[5],5)} & {n
p.round(val_acc[6],5)} & {np.round(val_acc[7],5)} & {np.round(val_acc[8],5)} ")
```

```
0.385 & 0.459 & 0.487 & 0.503 & 0.5 & 0.506 & 0.495 & 0.497 & 0.502
```

In [26]:

```
importlib.reload(m)
GD_params = {"epochs":10, "n_batch":100, "eta_min":1e-5,'eta_max':1e-1, 'ns':2*np.floor
(X_train_whole.shape[1]/100), 'n_cycles':3, 'freq':10}
search = m.LambdaSearch(-5, -2,8)
model_narrow = search.lambda_search(data=data,GDparams=GD_params)
```

```
100%|██████████| 12/12 [07:00<00:00, 35.08s/it]
100%|██████████| 12/12 [07:13<00:00, 36.11s/it]
100%|██████████| 12/12 [07:07<00:00, 35.61s/it]
100%|██████████| 12/12 [08:05<00:00, 40.47s/it]
100%|██████████| 12/12 [06:28<00:00, 32.34s/it]
100%|██████████| 12/12 [05:48<00:00, 29.03s/it]
100%|██████████| 12/12 [05:24<00:00, 27.01s/it]
100%|██████████| 12/12 [05:30<00:00, 27.58s/it]
```

In [27]:

```
importlib.reload(m)
for key in model_narrow.keys():
    train_loss_ = np.array(model_narrow[key].train_loss)
    val_loss_ = np.array(model_narrow[key].val_loss)
    train_acc_ = np.array(model_narrow[key].train_acc)
    val_acc_ = np.array(model_narrow[key].val_acc)
    train_cost_ = np.array(model_narrow[key].train_cost)
    val_cost_ = np.array(model_narrow[key].val_cost)
    hist = {'train_loss':train_loss_, 'val_loss':val_loss_, 'train_acc':train_acc_, 'va
l_acc':val_acc_,'train_cost':train_cost_,'val_cost':val_cost_}
    np.save(f"Models/lambda/second_run_hist_{model_narrow[key].lambda_}.npy",hist)
```

In [41]:

```
print(f"{np.round(model_narrow[keys[0]].lambda_,5)} & {np.round(model_narrow[keys[1]].l
ambda_,5)} & {np.round(model_narrow[keys[2]].lambda_,5)} & {np.round(model_narrow[keys
[3]].lambda_,5)} & {np.round(model_narrow[keys[4]].lambda_,5)} & {np.round(model_narrow
[keys[5]].lambda_,5)} & {np.round(model_narrow[keys[6]].lambda_,5)} & {np.round(model_n
arrow[keys[7]].lambda_,5)}  ")
```

```
1e-05 & 3e-05 & 7e-05 & 0.00019 & 0.00052 & 0.00139 & 0.00373 & 0.01
```

In [38]:

```
print(f"{np.round(keys[0],5)} & {np.round(keys[1],5)} & {np.round(keys[2],5)} & {np.rou
nd(keys[3],5)} & {np.round(keys[4],5)} & {np.round(keys[5],5)} & {np.round(keys[6],5)}
 & {np.round(keys[7],5)}  ")
```

```
0.5146 & 0.5158 & 0.5142 & 0.5176 & 0.516 & 0.5206 & 0.5182 & 0.5064
```

train model with best lambda

In [30]:

```
importlib.reload(m)
X_train_whole, y_train_whole, Y_train_whole = u.load_data('/data_batch_1', clipping=Tru
e)
for i in range(2,6):
    X, y, Y = u.load_data('/data_batch_'+str(i), clipping=True)
    X_train_whole = np.concatenate((X, X_train_whole), axis=1)
    y_train_whole = np.concatenate((y, y_train_whole))
    Y_train_whole = np.concatenate((Y, Y_train_whole), axis=1)

X_val_small, y_val_small, Y_val_small = X_train_whole[:,-1000:], y_train_whole[-1000:],
Y_train_whole[:,-1000:]
X_train_whole, y_train_whole, Y_train_whole = X_train_whole[:,:-1000], y_train_whole[:-
1000], Y_train_whole[:,:-1000]

## normalize with mean and std of train set
mean = np.mean(X_train_whole, axis=1)
std = np.std(X_train_whole, axis=1)

X_train_whole -= np.outer(mean, np.ones(X_train_whole.shape[1]))
X_train_whole /= np.outer(std, np.ones(X_train_whole.shape[1]))

X_val_small -= np.outer(mean, np.ones(X_val_small.shape[1]))
X_val_small /= np.outer(std, np.ones(X_val_small.shape[1]))
```

In [31]:

```
data = {'X_train':X_train_whole, 'Y_train':Y_train_whole, 'y_train':y_train_whole,'X_va
l':X_val_small, 'Y_val':Y_val_small, 'y_val':y_val_small}
```

In [33]:

```
importlib.reload(m)
mlp = m.MLP(lambda_=0.0013894954943731374)
GD_params = {"epochs":10, "n_batch":100, "eta_min":1e-5,'eta_max':1e-1, 'ns':2*np.floor
(X_train_whole.shape[1]/100), 'n_cycles':4, 'freq':10}
mlp.cyclicLearning(data, GD_params, 'best_lambda', True, True)
```

```
  0%|          | 0/16 [00:00<?, ?it/s]   Epoch 0: train_cost = 2.616085820
1226027, val_cost = 2.58126356702231,
         train_acc = 0.09991836734693878, val_acc = 0.112
         Epoch 196.0: train_cost = 1.86949565245215, val_cost = 1.87571623
66967464,
         train_acc = 0.36753061224489797, val_acc = 0.353
         Epoch 392.0: train_cost = 1.7639565516343176, val_cost = 1.802083
143383545,
         train_acc = 0.4019795918367347, val_acc = 0.394
  6%|▋         | 1/16 [00:53<13:25, 53.71s/it]  Epoch 588.0: train_cost =
1.7457656211710768, val_cost = 1.782695384730414,
         train_acc = 0.4101836734693878, val_acc = 0.401
         Epoch 784.0: train_cost = 1.6361154315150774, val_cost = 1.659805
9943282635,
         train_acc = 0.4512244897959184, val_acc = 0.437
 12%|█▎        | 2/16 [01:21<10:43, 45.94s/it]  Epoch 980.0: train_cost =
1.6427514088023318, val_cost = 1.6868616777578815,
         train_acc = 0.4440816326530612, val_acc = 0.432
         Epoch 1176.0: train_cost = 1.5661862566520661, val_cost = 1.63609
3671684625,
         train_acc = 0.4795510204081633, val_acc = 0.452
         Epoch 1372.0: train_cost = 1.4966358502579293, val_cost = 1.59470
89483827286,
         train_acc = 0.5009591836734694, val_acc = 0.447
 19%|█▉        | 3/16 [02:02<09:37, 44.40s/it]  Epoch 1568.0: train_cost
= 1.437022971339872, val_cost = 1.5337799254931426,
         train_acc = 0.524469387755102, val_acc = 0.486
         Epoch 1764.0: train_cost = 1.3972301604972852, val_cost = 1.50347
4884857446,
         train_acc = 0.5427755102040817, val_acc = 0.488
 25%|██▌       | 4/16 [02:31<07:57, 39.81s/it]   Epoch 0.0: train_cost =
1.371762459969887, val_cost = 1.4782462681607274,
         train_acc = 0.5504897959183673, val_acc = 0.499
         Epoch 196.0: train_cost = 1.380787495508573, val_cost = 1.4887590
133927593,
         train_acc = 0.5463877551020409, val_acc = 0.494
         Epoch 392.0: train_cost = 1.4004398685600141, val_cost = 1.539155
924261773,
         train_acc = 0.5393673469387755, val_acc = 0.491
 31%|███▏      | 5/16 [03:11<07:18, 39.84s/it]  Epoch 588.0: train_cost =
1.4771292621669907, val_cost = 1.6334167691219326,
         train_acc = 0.5053877551020408, val_acc = 0.46
         Epoch 784.0: train_cost = 1.467689943496293, val_cost = 1.5947260
58096818,
         train_acc = 0.5083673469387755, val_acc = 0.466
 38%|███▊      | 6/16 [03:48<06:30, 39.04s/it]  Epoch 980.0: train_cost =
1.4712238771370583, val_cost = 1.5708636126842854,
         train_acc = 0.5153877551020408, val_acc = 0.463
         Epoch 1176.0: train_cost = 1.462874646294769, val_cost = 1.599210
2717415948,
         train_acc = 0.5088571428571429, val_acc = 0.446
         Epoch 1372.0: train_cost = 1.4006421237893818, val_cost = 1.50826
0511663802,
         train_acc = 0.5364285714285715, val_acc = 0.475
 44%|████▍     | 7/16 [04:22<05:38, 37.66s/it]  Epoch 1568.0: train_cost
= 1.339939597653143, val_cost = 1.491497596588065,
         train_acc = 0.5612244897959183, val_acc = 0.496
         Epoch 1764.0: train_cost = 1.3131486552930944, val_cost = 1.48191
1591150807,
         train_acc = 0.5724081632653061, val_acc = 0.509
 50%|█████     | 8/16 [04:49<04:35, 34.42s/it]   Epoch 0.0: train_cost =
```

```
1.2857004726266956, val_cost = 1.450464034663028,
         train_acc = 0.5838775510204082, val_acc = 0.504
         Epoch 196.0: train_cost = 1.3054636352034463, val_cost = 1.477600
795862976,
         train_acc = 0.5742244897959183, val_acc = 0.494
         Epoch 392.0: train_cost = 1.3370897980730934, val_cost = 1.526711
635764653,
         train_acc = 0.5605918367346939, val_acc = 0.472
 56%|█████▋     | 9/16 [05:23<03:59, 34.16s/it]  Epoch 588.0: train_cost =
1.3629607762205427, val_cost = 1.5647183234038882,
         train_acc = 0.5502040816326531, val_acc = 0.485
         Epoch 784.0: train_cost = 1.4231046441973076, val_cost = 1.578166
1120314556,
         train_acc = 0.528734693877551, val_acc = 0.491
 62%|██████▎    | 10/16 [05:49<03:10, 31.69s/it]         Epoch 980.0: trai
n_cost = 1.5686155093189584, val_cost = 1.7305642848749874,
         train_acc = 0.4839387755102041, val_acc = 0.454
         Epoch 1176.0: train_cost = 1.4017662990810629, val_cost = 1.56161
45700489796,
         train_acc = 0.5333469387755102, val_acc = 0.467
         Epoch 1372.0: train_cost = 1.3570861890635697, val_cost = 1.52306
38507652372,
         train_acc = 0.5546122448979592, val_acc = 0.494
 69%|██████▉    | 11/16 [06:22<02:41, 32.26s/it]         Epoch 1568.0: tra
in_cost = 1.3055310707432275, val_cost = 1.4718039901534536,
         train_acc = 0.5737551020408164, val_acc = 0.509
         Epoch 1764.0: train_cost = 1.271172613028134, val_cost = 1.472538
6680435613,
         train_acc = 0.5923469387755103, val_acc = 0.51
 75%|███████▌   | 12/16 [06:52<02:05, 31.32s/it]  Epoch 0.0: train_cost =
1.2416485197742533, val_cost = 1.4284529060869031,
         train_acc = 0.602530612244898, val_acc = 0.517
         Epoch 196.0: train_cost = 1.2558463337120211, val_cost = 1.458765
1074635972,
         train_acc = 0.5946734693877551, val_acc = 0.498
         Epoch 392.0: train_cost = 1.2758371881963755, val_cost = 1.459744
1761538388,
         train_acc = 0.5839183673469388, val_acc = 0.518
 81%|████████▏  | 13/16 [07:30<01:40, 33.38s/it]         Epoch 588.0: trai
n_cost = 1.3522525887567214, val_cost = 1.537571869645093,
         train_acc = 0.5533061224489796, val_acc = 0.47
         Epoch 784.0: train_cost = 1.4108921717739804, val_cost = 1.592357
2172204055,
         train_acc = 0.534530612244898, val_acc = 0.472
 88%|████████▊  | 14/16 [07:56<01:02, 31.37s/it]         Epoch 980.0: trai
n_cost = 1.4375464398690245, val_cost = 1.5501894628720034,
         train_acc = 0.5287755102040816, val_acc = 0.477
         Epoch 1176.0: train_cost = 1.4650695972471066, val_cost = 1.61034
53400426042,
         train_acc = 0.5156122448979592, val_acc = 0.473
         Epoch 1372.0: train_cost = 1.3496783411249784, val_cost = 1.51882
73148116962,
         train_acc = 0.5613877551020409, val_acc = 0.503
 94%|█████████▍ | 15/16 [08:32<00:32, 32.53s/it]         Epoch 1568.0: tra
in_cost = 1.285356725123175, val_cost = 1.4760540845608656,
         train_acc = 0.5871836734693877, val_acc = 0.513
         Epoch 1764.0: train_cost = 1.2422394034688782, val_cost = 1.42502
02549780109,
         train_acc = 0.6001632653061224, val_acc = 0.525
100%|██████████| 16/16 [08:58<00:00, 33.68s/it]
```

In [ ]: