## Triplet auxiliary rework

August 6, 2023

## 1 run load\_data.ipynb BEFORE running this!

```
[1]: import numpy as np
     import torch
     device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
     s = {
                                : "regression",
          'problem'
          'approach'
                                : "metric learning/non-parametric",
                              : "triplet network",
          'input' : "samples from a distribution",
'input type' : "vectors",
'input meaning' : "spectrum",
'output' "
          'algorithm'
          'output' : "samples from a distribution",
'output type' : "one number",
'output meaning' : "temperature or pressure, depending on distribution",
'learning rate' : 1e-4,
          'input dimension' : 10000,
          'output dimension' : 1,
          'feature dimension' : 300,
                                 : 1000,
          'epoch'
          'epoch-development' : 1,
          'cross validation round': 16,
          'cross validation round-development' : 1,
          'batch size' : 64,
          'best model folder' : 'triplet_best_model/'
     # https://arxiv.org/pdf/1412.6622.pdf
     import data_accessor as acc
     data_names_list = [
          'temperature_230509_discrete',
          'pressure_230516_discrete'
     data_dictionary = acc.setup(data_names_list)
```

```
loading pressure_230516_discrete_____
            input shape (number, dimension): (5000, 10000)
            label shape (number, dimension): (5000, 1)
            there are 16 folds
            3500 for training, 500 for validating, 1000 for testing
[2]: import torch.nn as nn
     class TripletNetwork(torch.nn.Module):
         """ Input: pos, neg, anchor, anchor_label
             Output: pos_prediction, neq_prediction"""
        def __init__(self, device, input_dimension, feature_dimension,_
      →output_dimension):
             super().__init__()
             self.input_dimension = input_dimension
             self.feature_dimension = feature_dimension
             self.output_dimension = output_dimension
             self.device = device
             self.feature_sequential = torch.nn.Sequential(
                 torch.nn.Linear(self.input_dimension, 3000),
                 nn.ReLU(),
                 torch.nn.Linear(3000, 600),
                 nn.ReLU(),
                 torch.nn.Linear(600, 600),
                 nn.ReLU(),
                 torch.nn.Linear(600, 300),
                 nn.ReLU(),
                 torch.nn.Linear(300, self.feature_dimension)
             self.auxiliary_sequential = torch.nn.Sequential(
                 torch.nn.Linear(self.feature_dimension, 100),
                 nn.ReLU(),
                 torch.nn.Linear(100, 100),
                 nn.ReLU(),
                 torch.nn.Linear(100, self.output_dimension)
             )
             self.to(device)
             self.float()
        def forward(self, pos, neg, anchor, anchor_label):
             feature_pos = self.feature_sequential(pos)
             feature_neg = self.feature_sequential(neg)
             feature_anchor = self.feature_sequential(anchor)
             feature space difference pos anchor = feature pos - feature anchor
             feature_space_difference_neg_anchor = feature_neg - feature_anchor
             label_space_difference_pos_anchor = self.
      →auxiliary_sequential(feature_space_difference_pos_anchor)
```

4200 for training, 600 for validating, 1200 for testing

```
label_space_difference_neg_anchor = self.

auxiliary_sequential(feature_space_difference_neg_anchor)

prediction_pos = anchor_label + label_space_difference_pos_anchor

prediction_neg = anchor_label + label_space_difference_neg_anchor

return prediction_pos, prediction_neg
```

```
[3]: from tools import SaveBestModel, PatienceEarlyStopping, Scheduler, plot_losses
     class Manager:
         """ DOES: train & evaluate a Siamese network
         def __init__(self, epoch, cross_validation_round):
             self._network = TripletNetwork(device, s['input dimension'], s['feature_

¬dimension'], s['output dimension'])
             self._network.apply(self.initializer)
             self._learning_rate = s['learning rate']
             self._optimizer = torch.optim.Adam(
                 params=self._network.parameters(), lr=self._learning_rate,
                 weight_decay=3e-3)
             self._energy = nn.MSELoss()
             self._train_loss = []
             self._valid_loss = []
             self._test_loss = []
             self. epoch = epoch
             self._stopper = PatienceEarlyStopping(patience=5, min_delta=5e-7)
             self. cross validation round = cross validation round
             self._saver = SaveBestModel(s['best model folder'])
             self. scheduler = Scheduler(optimizer=self. optimizer,
                 minimum_learning_rate=1e-6, patience=5, factor=0.5)
         def initializer(self, layer):
             if type(layer) == nn.Linear:
                 nn.init.kaiming_normal_(layer.weight) # normal version
         def _step(self, job):
             pos, pos_label, neg, neg_label, anchor, anchor_label = job
             pos_prediction, neg_prediction = self._network(pos, neg, anchor,_
      →anchor_label)
             pos_loss = self._energy(pos_prediction, pos_label)
             neg_loss = self._energy(neg_prediction, neg_label)
             loss = (pos_loss + neg_loss) / 2.0
             return loss
         def train(self, train_dataloader, valid_dataloader):
             """ DOES: calculate loss from tasks
                 NOTE: we have a BATCH of tasks here """
             for e in range(self._epoch):
                 batch_train_loss = []
                 for _, batch in enumerate(train_dataloader):
                     self._optimizer.zero_grad()
                     loss = self._step(batch)
```

```
loss.backward()
               self._optimizer.step()
               batch_train_loss.append(loss.item())
           self._train_loss.append(np.mean(batch_train_loss))
          batch_valid_loss = []
          with torch.no_grad():
               for _, batch in enumerate(valid_dataloader):
                   loss = self._step(batch)
                   batch valid loss.append(loss.item())
           self._valid_loss.append(np.mean(batch_valid_loss))
           # saving, early stopping, scheduler for EACH epoch!
          self._saver(current_loss=np.mean(batch_valid_loss), model=self.
→_network, round=self._cross_validation_round)
          self._scheduler(np.mean(batch_valid_loss))
           self._stopper(np.mean(batch_valid_loss))
           if self._stopper.early_stop == True:
              print(f"EARLY STOPPING @ epoch {e}")
              break
       # summary printout, after we're done with epochs
      print(f"min train loss: {np.min(self._train_loss)}")
      print(f"min valid loss: {np.min(self. valid loss)}")
      plot_losses(self._train_loss, self._valid_loss, self.
→_cross_validation_round)
      return np.min(self._valid_loss)
  def test(self, test_dataloader):
      with torch.no_grad():
          batch test loss = []
          for _, batch in enumerate(test_dataloader):
               loss = self._step(batch)
              batch_test_loss.append(loss.item())
           self._test_loss.append(np.mean(batch_test_loss))
      return np.min(self._test_loss)
```

```
[4]: import random
    class TripletDataset(torch.utils.data.TensorDataset):
        """ input: input data
            label: label
            indices: indices used e.g. training indices
            """

        def __init__(self, input, label, indices, device):
            self.input = torch.Tensor(input).to(device)
            self.label = torch.Tensor(label).to(device)
            self.access_indices = indices
            self.indices = range(len(self.access_indices))
        def __len__(self):
            return len(self.indices)
        def __getitem__(self, index):
```

```
index = self.access_indices[index]
anchor_index = random.choice(self.access_indices)
neg_index = random.choice(self.access_indices)
pos = self.input[index]
pos_label = self.label[index]
anchor = self.input[anchor_index]
anchor_label = self.label[anchor_index]
neg = self.input[neg_index]
neg_label = self.label[neg_index]
return pos, pos_label, neg, neg_label, anchor, anchor_label
```

```
[10]: from torch.utils.data import DataLoader
     from tools import SaveBestCrossValidationModel
     from Style import TextColor
     class CrossValidator:
         """ number of cross validation rounds
             \nnumber of epochs
             \nsaver (actual object)
             \ndataset (the function of the object)
         def __init__(self, number_of_cross_validation_rounds, number_of_epochs,__
       ⇒saver, dataset, datas, data_dictionary, settings):
             self.number_of_cross_validation_rounds =_
       →number_of_cross_validation_rounds
             self.number of epochs = number of epochs
             self.saver = saver
             self.cross validation loss = []
             self.dataset = dataset
             self.datas = datas
             self.data_dictionary = data_dictionary
             self.settings = settings
             print(f"{TextColor.Bold}{TextColor.BrightGreen text} CROSS__

¬VALIDATION______{TextColor.End}")
             print(f"Cross-validation rounds: {self.
       →number_of_cross_validation_rounds}")
             print(f"Epochs: {self.number_of_epochs}")
             print(f"Datas to learn: ")
             for index, data in enumerate(self.datas):
                 print(f"\t{index}: {data}")
         def complete_notify(self):
             # unicode https://www.geeksforgeeks.org/python-program-to-print-emojis/
             print(f"\U0001f607 {TextColor.Bold}{TextColor.BrightGreen text}TRAINing
       →COMPLETE _____{TextColor.End}")
         def single_task_train(self, data_index): # magenta
             print(f"{TextColor.Bold}{TextColor.Magenta_text}SINGLE_
       →TASK______{TextColor.End}")
```

```
print(f"we're learning: {self.datas[data_index]}")
      self.cross_validation_loss = []
      for round_index in range(self.number_of_cross_validation_rounds):
           print(f">round {round_index}")
          network_object = Manager(self.number_of_epochs, round_index)
          valid_loss = network_object.train( # DON'T do so in separate_
→ function
              DataLoader(self.dataset(
              self.data_dictionary[self.datas[data_index]]['data'],
              self.data_dictionary[self.datas[data_index]]['label'],
              self.data_dictionary[self.datas[data_index]]['train_
→indices'][round_index],
              device=device,), shuffle=False, batch_size=self.settings['batch_
⇔size']),
              DataLoader(self.dataset(
              self.data_dictionary[self.datas[data_index]]['data'],
              self.data_dictionary[self.datas[data_index]]['label'],
              self.data_dictionary[self.datas[data_index]]['valid__
⇔indices'][round_index],
              device-device,), shuffle-False, batch size-self.settings['batch,
⇔size']))
          self.saver(current_loss=valid_loss, round=round_index)
           self.cross_validation_loss.append(valid_loss)
      print(f"{TextColor.Bold}{TextColor.BrightGreen_text}BEST{TextColor.End}_
wmodel: {self.saver.best_model_name} with {self.saver.current_best_loss}")
      print(f"trained on {self.datas[data index]}")
      print(f"Aggregate performance: Valid loss mean {np.mean(self.
Gross_validation_loss)}, std {np.std(self.cross_validation_loss)}")
  def multi_task_train_sequential(self): # blue
       """ learn ONE data at a time, need to reset model in between"""
      print(f"{TextColor.Bold}{TextColor.Blue_text}MULTI TASK,__

Sequential {TextColor.End}")
      print(f"we're learning: multiple tasks")
      print(f"given [1, 2, 3], [a, b, c]: learn [1, 2, 3], reset model, learn [1, 2, 3]
# for number, data in enumerate(self.datas):
      # print(f'' \setminus t\{number\}: \{data\}'')
      self.cross_validation_loss = [[] for data in datas]
      # print(self.cross_validation_loss)
      for round_index in range(self.number_of_cross_validation_rounds):
          print(f">round {round_index}")
          network_object = Manager(self.number_of_epochs, round_index)
          for number, data in enumerate(self.datas):
              print(f"task {TextColor.Bold}{number}{TextColor.End}: {data}")
              valid_loss = network_object.train( # DON'T do so in separate_
→ function
```

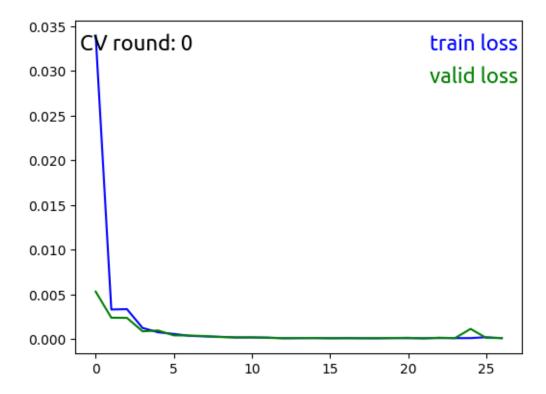
```
DataLoader(self.dataset(
                   self.data_dictionary[data]['data'],
                   self.data_dictionary[data]['label'],
                   self.data_dictionary[data]['train indices'][round_index],
                   device=device,), shuffle=False, batch_size=self.
⇔settings['batch size']),
                   DataLoader(self.dataset(
                   self.data dictionary[data]['data'],
                   self.data_dictionary[data]['label'],
                   self.data_dictionary[data]['valid indices'][round_index],
                   device=device,), shuffle=False, batch_size=self.
⇔settings['batch size']))
               # reset the network's saver, stopper, and losses (otherwise, it_{\square}
⇔early-stops rightaway)
              network_object._saver.reset()
              network_object._stopper.reset()
              network_object._train_loss = []
              network_object._valid_loss = []
               # reset auxiliary network
              network_object._network.auxiliary_sequential.
→apply(network_object.initializer)
               self.cross validation loss[number].append(valid loss)
           # picking best model by Current performance
           self.saver(current_loss=valid_loss, round=round_index)
      print(self.cross_validation_loss)
      print(f"{TextColor.Bold}{TextColor.BrightGreen_text}BEST{TextColor.End}_
_model: {self.saver.best_model_name} with {self.saver.current_best_loss}")
      print(f"trained datas sequentially")
      print(f"Aggregate performance:")
      for index, cv_loss in enumerate(self.cross_validation_loss):
          print(f"{self.datas[index]}: Valid loss mean {np.mean(self.
⇔cross_validation_loss[index])}, std {np.std(self.
⇔cross validation loss[index])}")
  def multi_task_train_weave(self, weave): # blue
       """ learn altogether, using 'super dataset' woven from datasets
           weave: function for weaving"""
      print(f"\U0001f9f5{TextColor.Bold}{TextColor.Magenta_text}MULTI TASK,__
→Interweave_____{TextColor.End}")
      print(f"we're learning: multiple tasks")
      print(f"given [1, 2, 3], [a, b, c]: learn [1, a, 2, b, 3, c], simple_{\sqcup}
⇔handling of different counts")
       # for number, data in enumerate(self.datas):
            print(f"\t{number}: {data}")
      self.cross_validation_loss = []
      for round_index in range(self.number_of_cross_validation_rounds):
          print(f">round {round_index}")
```

```
network_object = Manager(self.number_of_epochs, round_index)
          # t =
weave([data_dictionary[data]['data'][data_dictionary[data]['train_
# print(f"shape of woven: {t.shape}")
          # i = sum([len(data dictionary[data]['train indices'][0]) for data___
\rightarrow in datas 1)
          # print(f"length of woven: {i}")
          valid_loss = network_object.train( # DON'T do so in separate_
→ function
              DataLoader(self.dataset(
              weave([self.data dictionary[data]['data'][self.
-data_dictionary[data]['train indices'][round_index]] for data in self.
⇔datas]),
              weave([self.data_dictionary[data]['label'][self.
-data_dictionary[data]['train indices'][round_index]] for data in self.
→datas]),
              range(sum([len(self.data dictionary[data]['train indices'][0])
⇔for data in self.datas])),
              device=device,), shuffle=False, batch_size=self.settings['batch_u
⇔size']),
              DataLoader(self.dataset(
              weave([self.
odata_dictionary[data]['data'][data_dictionary[data]['validu
⇔indices'][round_index]] for data in self.datas]),
              weave([self.
odata_dictionary[data]['label'][data_dictionary[data]['valid⊔
range(sum([len(self.data dictionary[data]['valid indices'][0])

¬for data in self.datas])),
              device=device,), shuffle=False, batch_size=self.settings['batch_
⇔size']))
          self.saver(current loss=valid loss, round=round index)
          self.cross_validation_loss.append(valid_loss)
      print(f"{TextColor.Bold}{TextColor.BrightGreen_text}BEST{TextColor.End}_
→model: {self.saver.best_model_name} with {self.saver.current_best_loss}")
      print(f"trained datas by weaving them")
      print(f"Aggregate performance: Valid loss mean {np.mean(self.
across_validation_loss)}, std {np.std(self.cross_validation_loss)}")
  def test_all(self):
      """ test data one by one"""
      print(f"{TextColor.Bold}{TextColor.
→Blue_text}TEST_____{TextColor.End}")
      retained loss = {}
      network_object = Manager(None, None)
```

```
network_object._network.load_state_dict(torch.load(self.settings['best_L
  for data in self.datas:
            print(f"Testing {data}, loss: ", end=" ")
            test_loss = network_object.test(
                DataLoader(self.dataset(
                self.data_dictionary[data]['data'],
                self.data dictionary[data]['label'],
                self.data_dictionary[data]['test indices'],
                device=device,), shuffle=False, batch_size=self.settings['batch_
 ⇔size']))
            # record results
            retained_loss[data] = test_loss
            print(f"{test_loss}")
            # print(f"{data} Test loss: {test_loss}")
datas = ['temperature_230509_discrete', 'pressure_230516_discrete']
CVtor = CrossValidator(s['cross validation round'],
                       s['epoch'],
                      SaveBestCrossValidationModel(s['best model folder']),
                      TripletDataset,
                      datas,
                      data_dictionary,
# CVtor.single_task_train(0)
CVtor.multi_task_train_sequential()
# from data import alternate rows itertools
# CVtor.multi task train weave(alternate rows itertools)
CVtor.complete_notify()
CVtor.test all()
# from IPython.display import Audio # auto-play doesn't work!
# sound file = 'sound/IRWTS@UH.mp3'
# aud = Audio(sound_file, autoplay=True)
# display(aud)
_____CROSS VALIDATION_____
Cross-validation rounds: 16
Epochs: 1000
Datas to learn:
       0: temperature_230509_discrete
       1: pressure_230516_discrete
MULTI TASK, Sequential_____
we're learning: multiple tasks
given [1, 2, 3], [a, b, c]: learn [1, 2, 3], reset model, learn [a, b, c]
>round 0
task 0: temperature_230509_discrete
EARLY STOPPING @ epoch 26
```

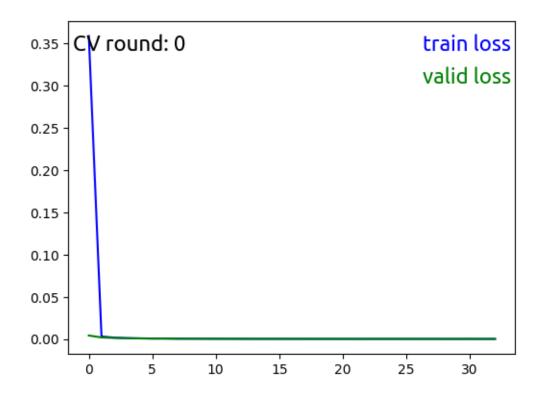
min train loss: 0.00010869141844909102 min valid loss: 9.737271830090322e-05



task 1: pressure\_230516\_discrete

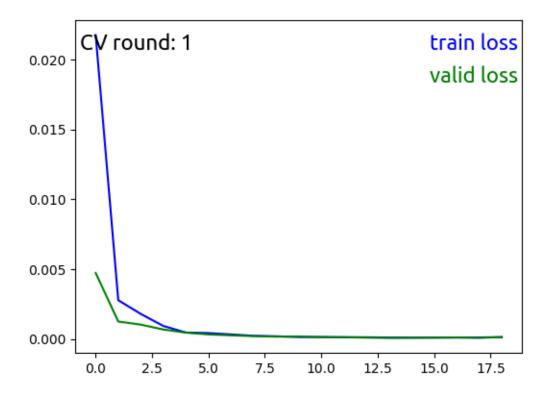
EARLY STOPPING @ epoch 32

min train loss: 0.00024692749297669666 min valid loss: 0.00024825766195135657



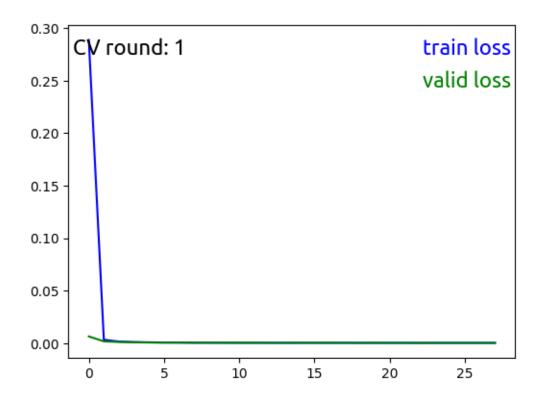
>round 1
task 0: temperature\_230509\_discrete
EARLY STOPPING @ epoch 18

min train loss: 0.00010494319455271277 min valid loss: 8.265267169917933e-05



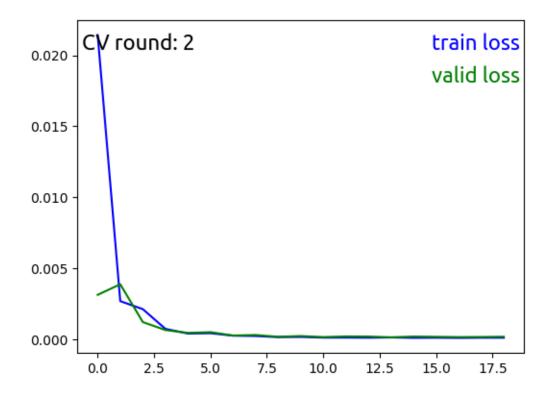
task 1: pressure\_230516\_discrete

min train loss: 0.00019081774358214302 min valid loss: 0.0001892097880045185



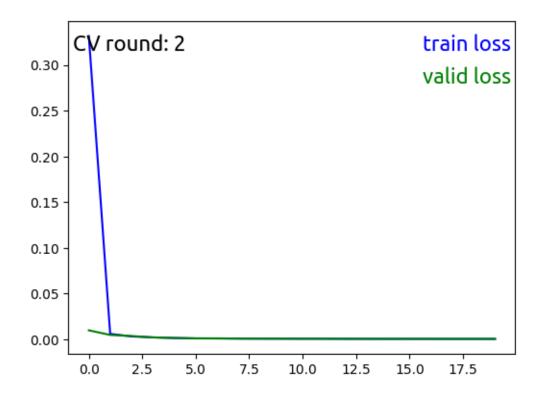
>round 2
task 0: temperature\_230509\_discrete
EARLY STOPPING @ epoch 18

min train loss: 9.678274970307636e-05 min valid loss: 0.00014083173009566964



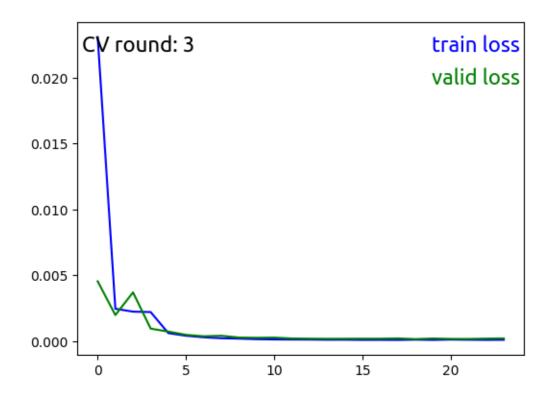
task 1: pressure\_230516\_discrete

min train loss: 0.00038998132195874034 min valid loss: 0.0004620359941327479



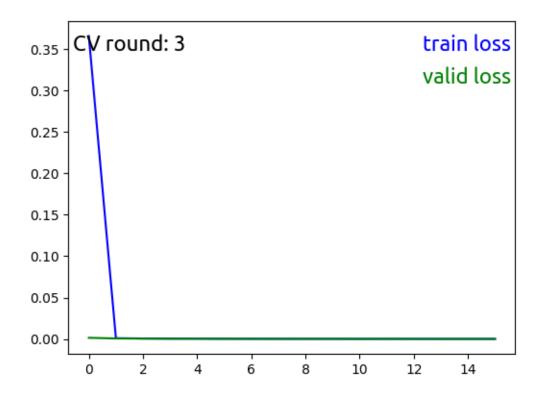
>round 3
task 0: temperature\_230509\_discrete

min train loss: 0.00010253816303289072 min valid loss: 0.00016388599324272944



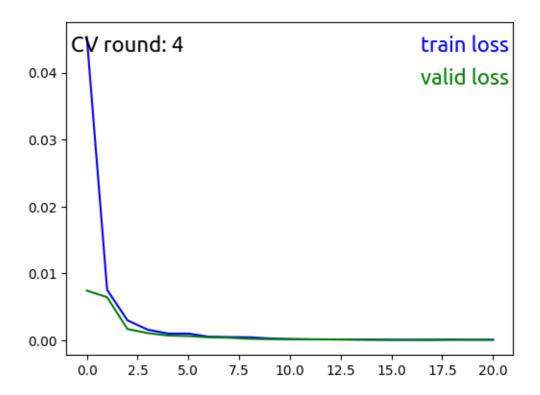
task 1: pressure\_230516\_discrete

min train loss: 0.00019498267211019992 min valid loss: 0.00017604994536668528



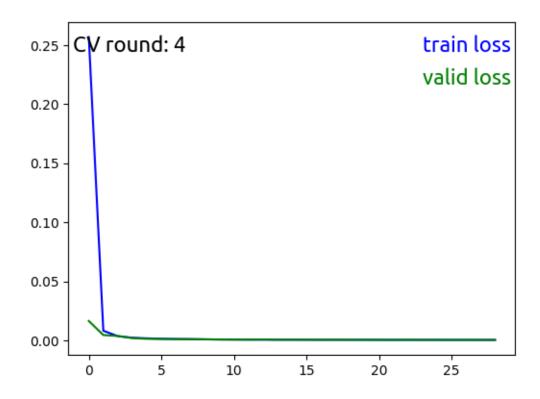
>round 4
task 0: temperature\_230509\_discrete

min train loss: 0.00010347300072184721 min valid loss: 8.235134155256673e-05



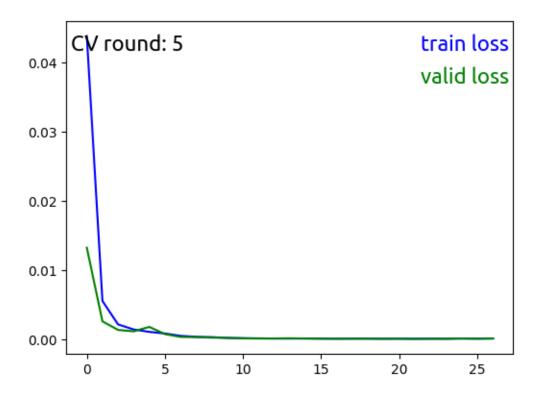
task 1: pressure\_230516\_discrete

min train loss: 0.00036523225241002035 min valid loss: 0.0003884993893734645



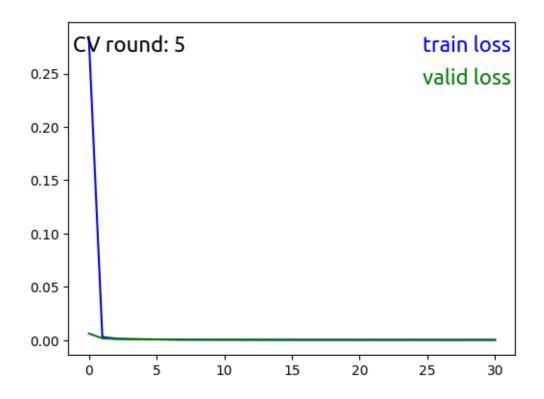
>round 5
task 0: temperature\_230509\_discrete

min train loss: 0.00010554670955047406 min valid loss: 8.296059495478403e-05



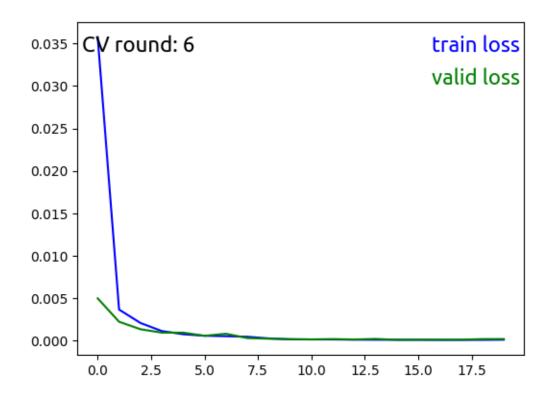
task 1: pressure\_230516\_discrete

min train loss: 0.00027385312688155947 min valid loss: 0.00023006127412372734



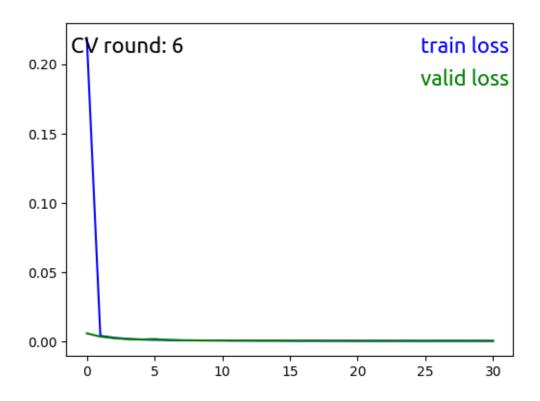
>round 6
task 0: temperature\_230509\_discrete

min train loss: 0.00010665822937329666 min valid loss: 0.00011839915605378337



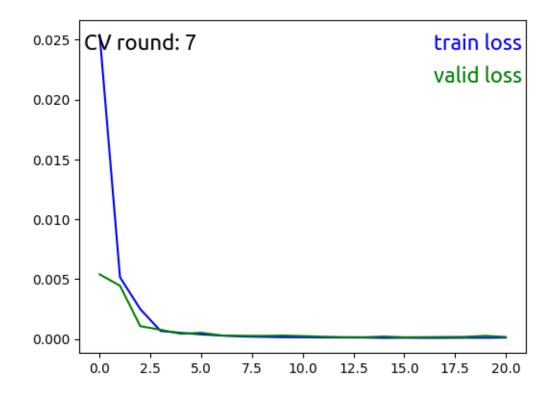
task 1: pressure\_230516\_discrete

min train loss: 0.00032262610356238757 min valid loss: 0.00036560107764671557



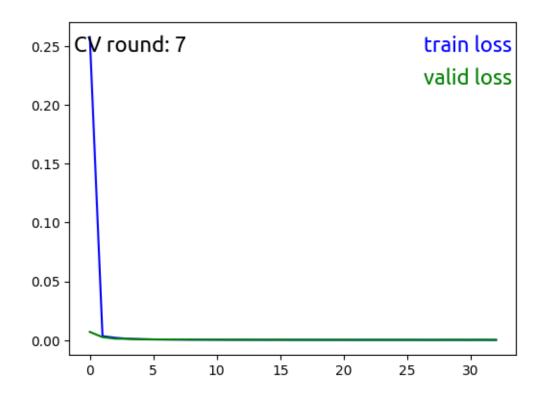
>round 7
task 0: temperature\_230509\_discrete
EARLY STOPPING @ epoch 20

min train loss: 9.943244957398609e-05 min valid loss: 0.00015313469120883382



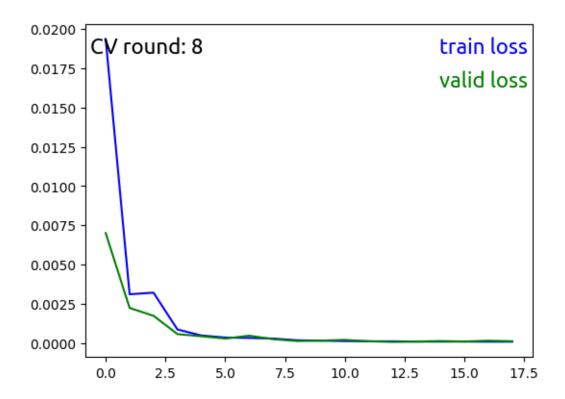
task 1: pressure\_230516\_discrete

min train loss: 0.0002284721955551173 min valid loss: 0.00022982150767347775



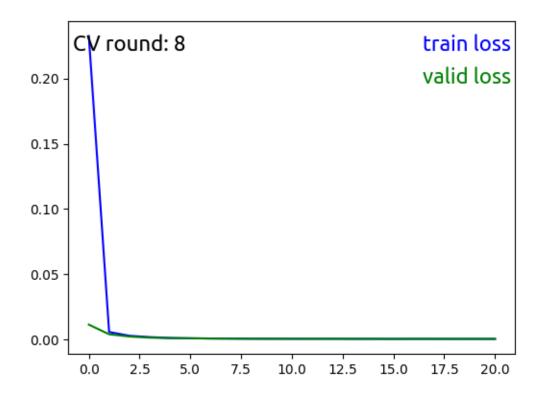
>round 8 task 0: temperature\_230509\_discrete EARLY STOPPING @ epoch 17

min train loss: 0.0001030636224944222 min valid loss: 8.87308062374359e-05



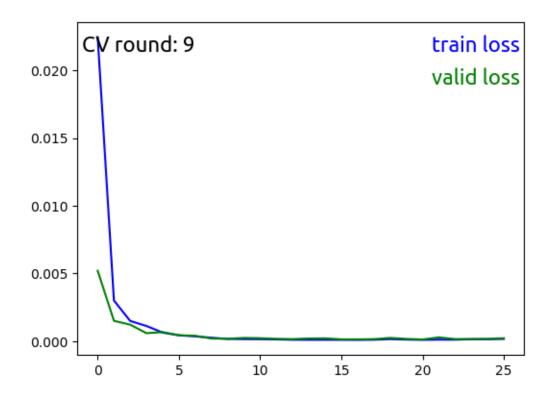
task 1: pressure\_230516\_discrete

min train loss: 0.00028455696936527435 min valid loss: 0.00028703051430056803



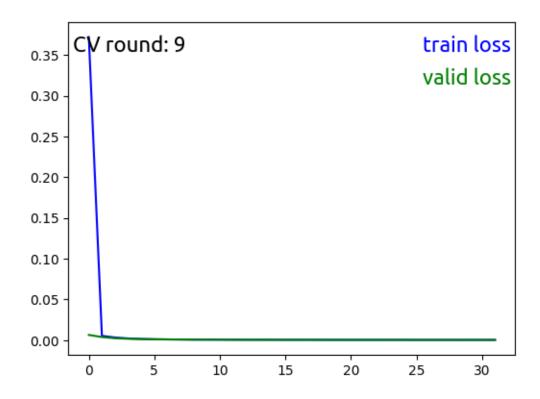
>round 9
task 0: temperature\_230509\_discrete
EARLY STOPPING @ epoch 25

min train loss: 9.773058550446436e-05 min valid loss: 0.00012975300196558238



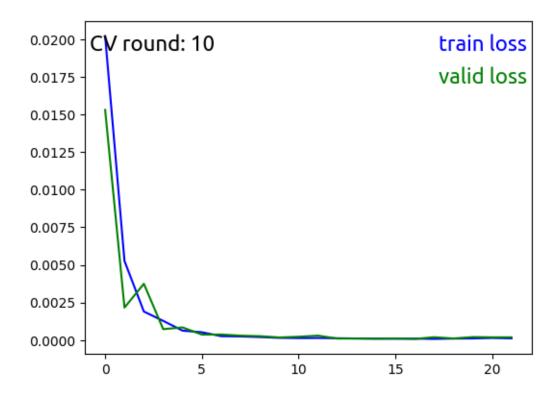
task 1: pressure\_230516\_discrete

min train loss: 0.00036854941546070303 min valid loss: 0.0003501349929138087



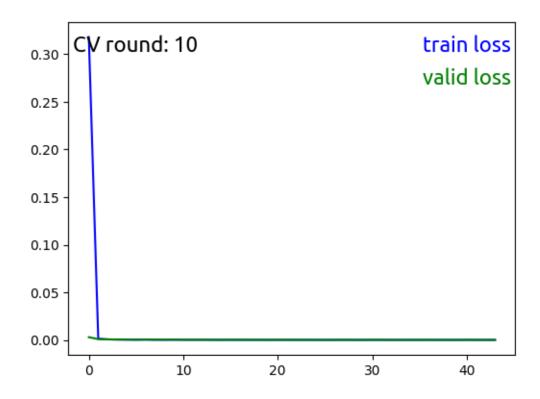
>round 10
task 0: temperature\_230509\_discrete

min train loss: 8.701189984673296e-05 min valid loss: 8.450537388853263e-05



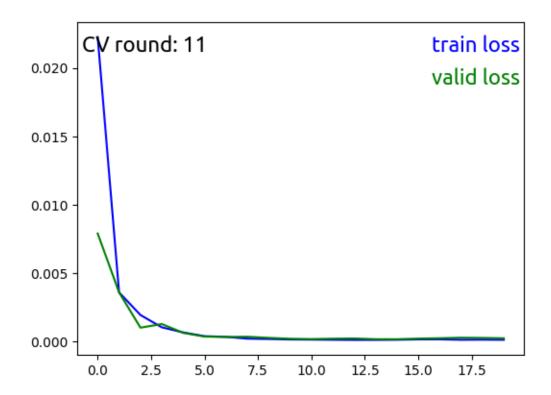
task 1: pressure\_230516\_discrete

min train loss: 0.00011170285261786458 min valid loss: 0.0001023190652631456



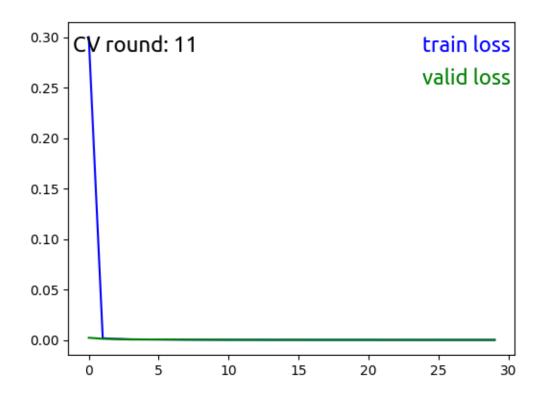
>round 11
task 0: temperature\_230509\_discrete
EARLY STOPPING @ epoch 19

min train loss: 0.0001205099045667642 min valid loss: 0.0001703827096207533



task 1: pressure\_230516\_discrete

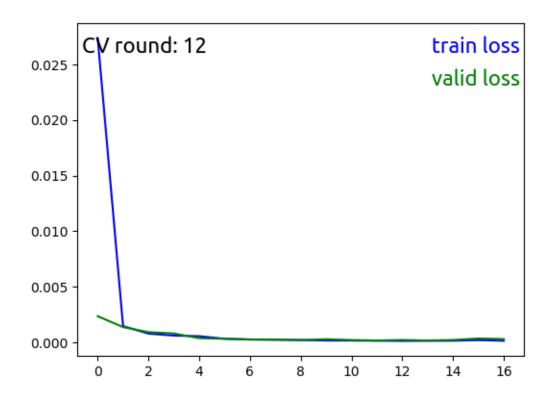
min train loss: 0.00016422359794060785 min valid loss: 0.00016678247266099788



>round 12 task 0: temperature\_230509\_discrete EARLY STOPPING @ epoch 16

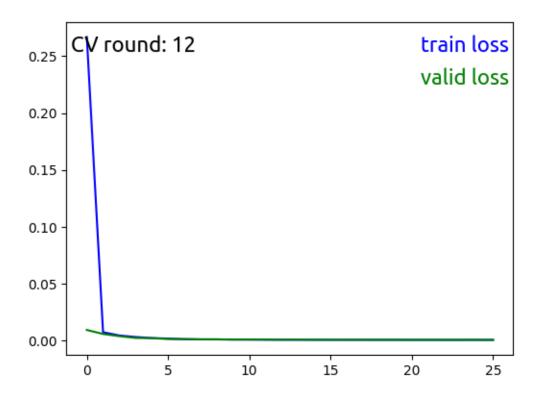
min train loss: 0.00012534673420084414

min valid loss: 0.00016402994297095574



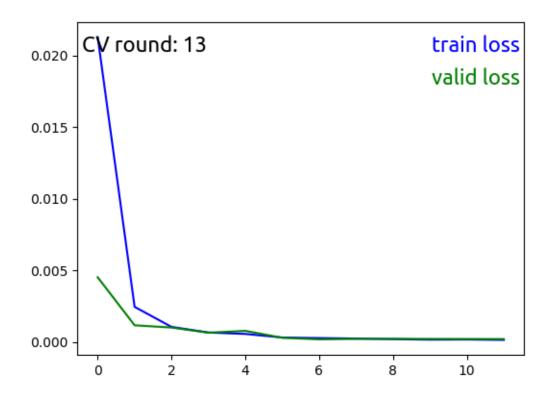
task 1: pressure\_230516\_discrete

min train loss: 0.00044018910874993626 min valid loss: 0.0004875907106907107



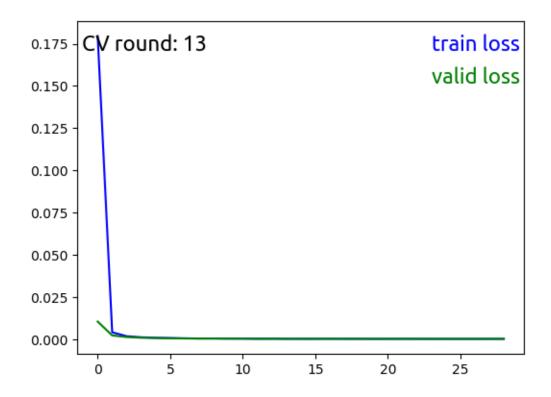
>round 13
task 0: temperature\_230509\_discrete
EARLY STOPPING @ epoch 11

min train loss: 0.0001334311608375036 min valid loss: 0.00017775970918592064



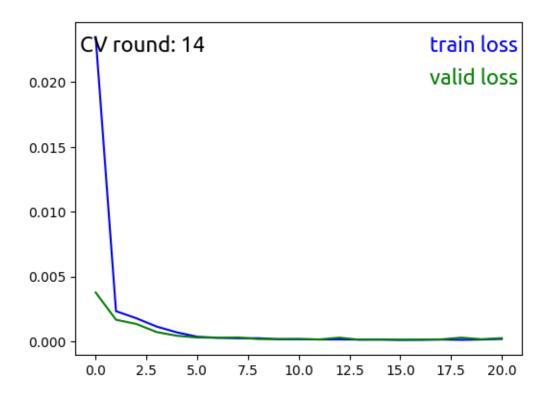
task 1: pressure\_230516\_discrete

min train loss: 0.00023313848730388353 min valid loss: 0.0002489888465788681



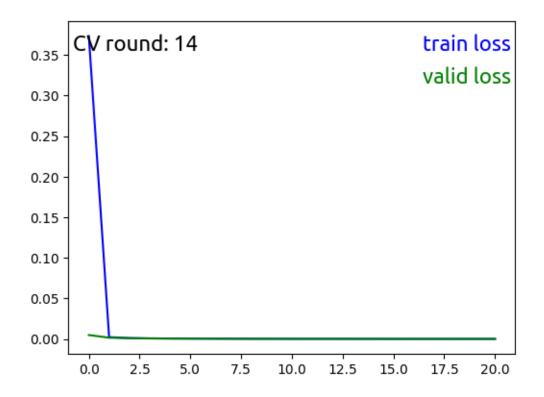
>round 14
task 0: temperature\_230509\_discrete

min train loss: 0.00010138779973865233 min valid loss: 0.00011621202211244963



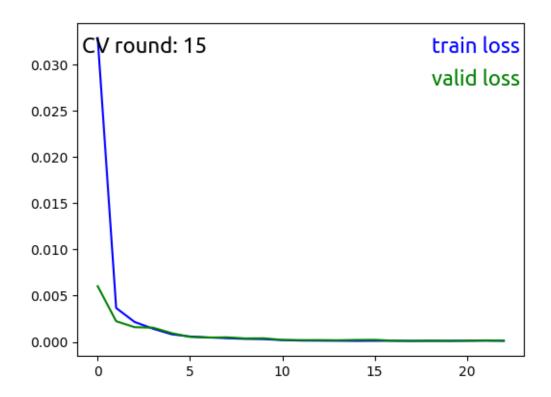
task 1: pressure\_230516\_discrete

min train loss: 0.00018253046609143813 min valid loss: 0.0002107188111040159



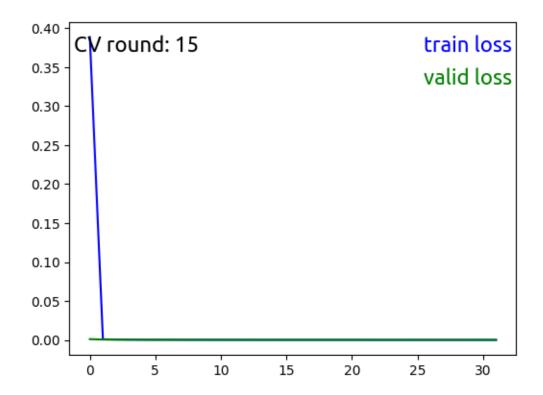
>round 15
task 0: temperature\_230509\_discrete

min train loss: 0.00010468159841063122 min valid loss: 0.00010580955749901477



task 1: pressure\_230516\_discrete

min train loss: 0.00015992978733265774 min valid loss: 0.00012046246501995483



```
[[9.737271830090322e-05, 8.265267169917933e-05, 0.00014083173009566964,
0.00016388599324272944, 8.235134155256673e-05, 8.296059495478403e-05,
0.00011839915605378337, 0.00015313469120883382, 8.87308062374359e-05,
0.00012975300196558238, 8.450537388853263e-05, 0.0001703827096207533,
0.00016402994297095574, 0.00017775970918592064, 0.00011621202211244963,
0.00010580955749901477], [0.00024825766195135657, 0.0001892097880045185,
0.0004620359941327479, 0.00017604994536668528, 0.0003884993893734645,
0.00023006127412372734, 0.00036560107764671557, 0.00022982150767347775,
0.00028703051430056803, 0.0003501349929138087, 0.0001023190652631456,
0.00016678247266099788, 0.0004875907106907107, 0.0002489888465788681,
0.0002107188111040159, 0.00012046246501995483]
BEST model: CV=10.pth with 0.0001023190652631456
trained datas sequentially
Aggregate performance:
temperature 230509 discrete: Valid loss mean 0.00012242325128681842, std
3.393405375557098e-05
pressure_230516_discrete: Valid loss mean 0.0002664727823002977, std
0.00011125465624891548
 TRAINing COMPLETE_____
Testing temperature_230509_discrete, loss: 0.36138172839817245
```

Testing temperature\_230509\_discrete, loss: 0.36138172839817245 Testing pressure\_230516\_discrete, loss: 0.0001522663201285468