

Week2.

PyTorch 2-3.

DataLoader

신승진

DataLoader

How DataLoader makes batch?

```
DataLoader(dataset, batch_size=1, shuffle=False, sampler=None,  
          batch_sampler=None, num_workers=0, collate_fn=None,  
          pin_memory=False, drop_last=False, timeout=0,  
          worker_init_fn=None, *, prefetch_factor=2,  
          persistent_workers=False)
```

- dataset
 - Map-style dataset
 - Iterable dataset
- sampler
 - dataset Index Iterator
- collate_fn
 - Callable

DataLoader

How DataLoader makes batch?

map-style dataset → get index [0,1,2,3,...] → shuffle index → make index

iterator [4,10,8,0,...] **sampler** → batch sampler [[4,10], [8,0], ...] →

```
for indices in batch_sampler:  
    yield collate_fn([dataset[i] for i in indices])
```

yield **collate_fn**([4,10]) ... → iterator

<batch 0>

“코딩/을/하면/시간/가는/줄/몰라” (7개)

“진심/이니/?” (3개) → “진심/이니/?/PAD/PAD/PAD” (7개)

<batch 1>

“오징어/게임/봤/니/?” (5개)

“그걸/지금/봤/니/?” (5개)