DataLoader

دیتا لودر چیست ؟

ابزاری است که مدیریت بارگذاری، پیش‌پردازش و ارائه داده‌ها به مدل را بهینه می‌کند

اصلی‌ترین وظایف دیتالودر:

1.*بارگذاری داده*

داده‌ها را از منابع مختلف می‌خواند

مثال :

```
from torch.utils.data import DataLoader, Dataset
dataset = MyCustomDataset()  # داده‌های شما
dataloader = DataLoader(dataset, batch_size=32)
```

2.*ایجاد باتچ‌ها*

داده‌ها را به دسته‌های کوچک یا بچ تقسیم می‌کند تا پردازش آن‌ها توسط مدل بهینه شود

```
dataloader = DataLoader(dataset, batch_size=32)
for batch in dataloader:
    inputs, labels = batch    # هر batch شامل 32 نمونه است
(batch_size=32)
```

3.*تصادفی‌سازی*

داده‌ها را قبل از هر دوره تصادفی می‌کند تا مدل به ترتیب داده‌ها حساس نشود

```
dataloader = DataLoader(dataset, batch_size=32, shuffle=True)
```

چرا دیتالودر مهم است؟

داده‌ها را به جای بارگذاری یکجا، به صورت باتچ‌های کوچک لود می‌کند

دلیل حرفای بالا ?!

در کد ما دیتالودر به صورت دستی با متد get_batch پیاده سازی شده است

DataLoader -> **get_batch** method

```
  def get_batch(split:str):
    ids = train_ids if split == 'train' else val_ids
    ix = torch.randint(0,len(ids) - cfg.seq_len - 1,(cfg.batch_size,))
    x = torch.stack([ids[i : i + cfg.seq_len] for i in ix])
    y = torch.stack([ids[i+1 : i+1 + cfg.seq_len] for i in ix])
    return x.to(device),y.to(device)
```

▶ توضیح کد

■ *انتخاب تصادفی ایندکس‌ها*

```
ix = torch.randint(0, len(ids) - cfg.seq_len - 1, (cfg.batch_size,))
```

ایندکس‌های تصادفی در محدوده [0, len(ids) - seq_len - 1] تولید می‌کند

seq_len -> دلیل تفریق

اطمینان از اینکه هنگام برش دنباله (تایی seq_len) از محدوده خارج نشویم

■ *ساخت باتچ‌های ورودی و هدف*

```
x = torch.stack([ids[i : i + cfg.seq_len] for i in ix])
y = torch.stack([ids[i+1 : i+1 + cfg.seq_len] for i in ix])
```

```
  def get_batch(split:str):
    ids = train_ids if split == 'train' else val_ids
    ix = torch.randint(0,len(ids) - cfg.seq_len - 1,(cfg.batch_size,))
    x = torch.stack([ids[i : i + cfg.seq_len] for i in ix])
    y = torch.stack([ids[i+1 : i+1 + cfg.seq_len] for i in ix])
    return x.to(device),y.to(device)
```