

11/23

1. GNN - 학습데이터

However, if you are working on DOCKGROUND dataset, it should be simpler. All the correct decoys typically with large numbers like 365130, and those decoys named with r-l-1.pdb to r-l-100.pdb are typically incorrect structures.

숫자가 1 ~ 100까지는 incorrect -> $Y = 0$.으로 설정

숫자 100 초과는 correct -> $Y = 1$.으로 설정

1. GNN - 결과_데이터 전체

데이터 처리를 한 후에 데이터 개수 **1837**개로 학습을 했습니다.
그랬더니 **10%**를 벗어나지 못했습니다.

```
grad_fn=<DivBackward0>)
12
0.2719958921287069 tensor([0.0799, 0.0800, 0.0804, 0.0796], device='cuda:0',
grad_fn=<DivBackward0>)
13
0.27732021958507885 tensor([0.0825, 0.0825, 0.0817, 0.0826], device='cuda:0',
grad_fn=<DivBackward0>)
14
0.27679847811526687 tensor([0.0825, 0.0818, 0.0829, 0.0818], device='cuda:0',
grad_fn=<DivBackward0>)
```

1. GNN - 결과_데이터 전체

그래서 논문을 보니 correct가 incorrect보다 훨씬 많은 데이터라서 학습을 할때, correct를 여러번 sample해서 학습을 진행하였습니다.

데이터를 살펴보니 incorrect 1700, correct 137개라서
 $\text{incorrect } 1700 + \text{correct } (137 * 13 = 1781) = 3481$ 개의 데이터로 학습을 했습니다.

Since the dataset was highly imbalanced with more incorrect decoys than acceptable ones, we balanced the training data by sampling the same number of acceptable and incorrect decoys in each batch. We sampled the same number of correct and incorrect decoys. To achieve this, a positive (i.e., correct) decoy may be sampled multiple times in one epoch of training.

1. GNN - 결과_데이터 전체

```
tensor([0.6233], device='cuda:0', grad_fn=<ViewBackward>)  
tensor([0.6122], device='cuda:0', grad_fn=<ViewBackward>)  
tensor([0.6160], device='cuda:0', grad_fn=<ViewBackward>)  
tensor([0.6120], device='cuda:0', grad_fn=<ViewBackward>)  
tensor([0.5836], device='cuda:0', grad_fn=<ViewBackward>)  
tensor([0.6060], device='cuda:0', grad_fn=<ViewBackward>)
```

학습을 하고 정답이 1인 데이터에 대해서 테스트를 해보니 0.6정도의 정확성이 나왔습니다.

1. GNN - 개선

데이터 종류를 늘리는 등의 데이터 처리를 하거나

각종 parameter 조정

정답과 오답 데이터간의 비율을 조정을 해봐야 할 것 같습니다.

비율을 약 1:1 정도로 진행을 했는데, 정답 데이터의 비율을 늘리면 결과가 달라질지 해보겠습니다.

아니면 오답과 정답의 값의 차이를 벌리는 형식으로도 해봐야 할 것 같습니다.

지금은 오답을 0, 정답을 1으로 했는데, 정답을 2로 올리던가 해서 값의 차이를 많이 줘도 괜찮을 것 같습니다.