**실무 테스트**

1. 유전체에서 발생한 어떤 유전자 변이는 희귀질환을 일으킬 수 있고, 질병을 일으키는 유전자 변이를 “병원성 변이 (Pathogenic variant)”라고 한다. 그런데, 희귀질환 환자의 유전체에서는 질병을 일으키는 유전자 외에도 여러 “비병원성 변이 (Benign variant)”가 함께 발견된다. 다음 파일 (pathogenicity\_scores.csv) 에는 107명의 희귀질환 환자들의 유전체에서 발견된 여러 비병원성 변이와 질병 유발 변이가 기록되어 있다. 3개의 서로 다른 병원성 판독기 (A, B, C)는 유전자 변이의 병원성을 예측하여 점수로 나타낸다. 우리는 판독기 중 하나를 사용해서 한 환자의 여러 유전자 변이 가운데 질병을 유발하는 유전자 변이를 찾아 내려고 한다. 각 판독기의 성능을 여러 기준으로 비교, 평가하는 코드를 작성한 후, 어떤 판독기를 활용하겠는지 설명하라.

(파일 설명)

Pathogenicity\_scores.csv 파일의 각 row는 환자에게서 발견된 유전자 변이를 나타내고, 각 column은 다음을 의미한다.

Patient\_ID: 각 환자에게 부여되는 고유한 ID

Mutation\_ID: 각 유전자 변이에 부여되는 고유한 ID

LABEL: 변이의 질병 유발 여부 (1: 질병 유발 변이, 0: 비병원성 변이)

SCORE\_A: A 판독기의 병원성 예측값 (1에 가까울 수록 병원성이 높고 0에 가까울 수록 낮다)

SCORE\_B: B 판독기의 병원성 예측값  
SCORE\_C: C 판독기의 병원성 예측값

2. 합성 약물은 구조에 따라 인체 내에서 결합할 수 있는 단백질의 종류가 달라진다. 어떤 약물은 다수의 단백질과 결합할 수 있는 반면 어떤 약물은 어떠한 단백질에도 결합하지 못할 수도 있다. 약물의 구조적 특징은 0과 1로 이루어진 1000 차원의 feature vector로 나타낼 수 있다. 100 종의 단백질과 약물의 결합 여부는 0과 1로 이루어진 100차원의 feature vector로 나타낼 수 있다. 300,000 개의 합성 약물에 대해 각각의 구조적 특징 데이터와 단백질 결합 데이터가 주어졌다고 가정하자. (파일은 제공되지 않습니다)

위 데이터를 사용하여 어떤 약물이 100종의 단백질 가운데 어떤 단백질(들)과 결합할지 0과 1 사이의 점수로 예측하는 아래 주어진 구조의 딥러닝 네트워크를 작성하시오. 작성한 구조를 학습하기 위해 어떤 loss function을 사용할 것인지 이유와 함께 기술하시오. (구조만 작성하고 학습/평가 코드는 작성하지 않는다)

Diagram

Description automatically generated