<DNN 연구 현황 정리>

**1. 데이터 수집 과정**

1만명의 암환자 조직 세포와 정상 조직 세포를 얻어(비율은 거의 1 대 1)

2000명의 Test set과 8000명의 Training set으로 나눈다. 이때 랜덤 셔플링으로 데이터를 나눈다. Five-fold cross validation또한 가능하게 했다.

**2. 전처리 과정**

로그 스케일 되지 않은 데이터를 로그 스케일 시켜주고, na 가 있는 사람의 데이터는 제거 했으며 중복되는 유전자 이름이 없는 유전자 또한 제거 했다.

또한 각 사람의 유전자 발현량으로 standard-normalize를 했다.

**3. 모델 형성 과정**

DNN 모델을 형성 하였고 다양한 노드와 레이어를 실험해 보아 그 중 적합한 것을 골랐다.

모든 변수를 다 쓰지 않고 분산이 큰 유전자 30퍼센트만 활용하였다. 또한 weight 값을 초기화 할때 Xavier initializer을 활용했다.

Hypothesis 함수로 Relu 함수를 사용 했으며 소프트 맥스를 cost function으로 사용했다.

Adam decent algoritms을 사용하였다.

**4. 모델 최적화 과정**

over fitting을 막기 위해 l2 regulazation과 drop out을 하였다.