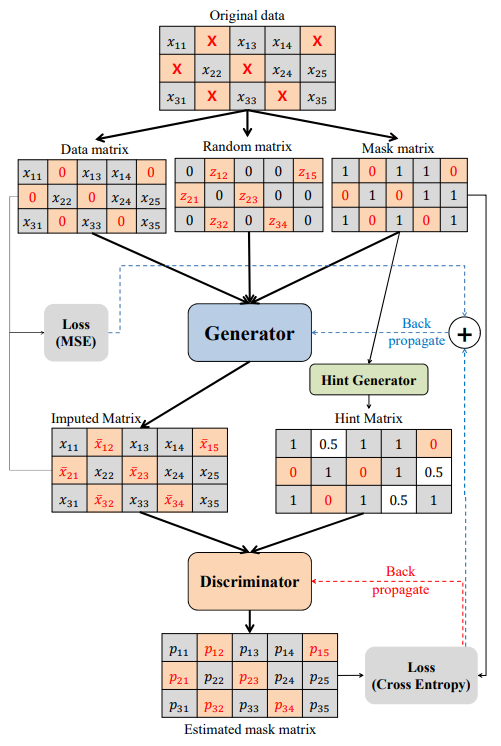
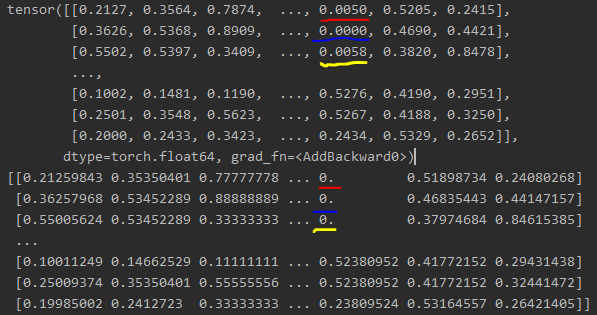
**Report for Survival Analysis**

2020020592 한정우

이번에 AI, Healthcare Assignment로 Survival Analysis을 진행하였다. 데이터는 ADNI을 사용하였으며, missing data를 Imputation 하는 것이 STEP1, 결과로 Linear regression based classification을 하는 것이 STEP2, 결과를 Kaplan-Meier Estimator를 하는 것이 STEP3이고, STEP2와 STEP3를 비교하는 것이 마지막 STEP4 이다.

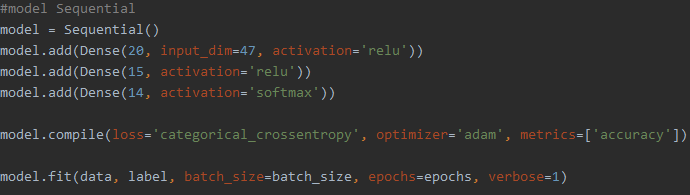
STEP1 에서는 교수님이 추천해주신 [1]GAIN 을 적용하였다.

1. Generator는 Missing 부분을 완전히 채우고, Discriminator의 오차를 최대화 하는게 목표
2. Discriminator는 imputed 값과 관측 값 사이에서 분별, 분류 로스를 최소화 하는게 목표
3. Hint Generator는 Discriminator의 학습 방향을 도와주는 역할

Result : 처음에는 Loss가 떨어지지 않아 진행 도중에 어려움이 많았으나, 기존 Model을 가져와 Tuning하는 과정에서 실수가 발생하여 어려움을 겪었음

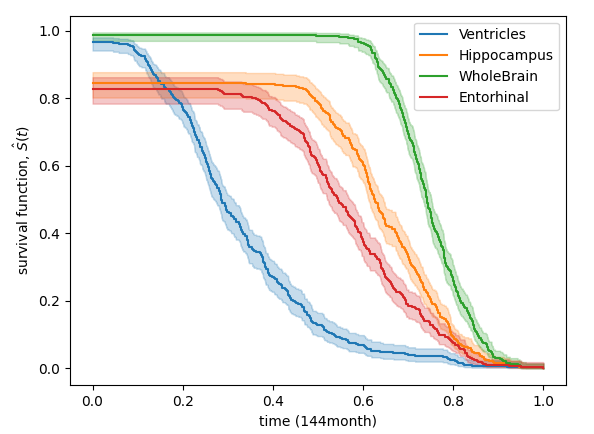
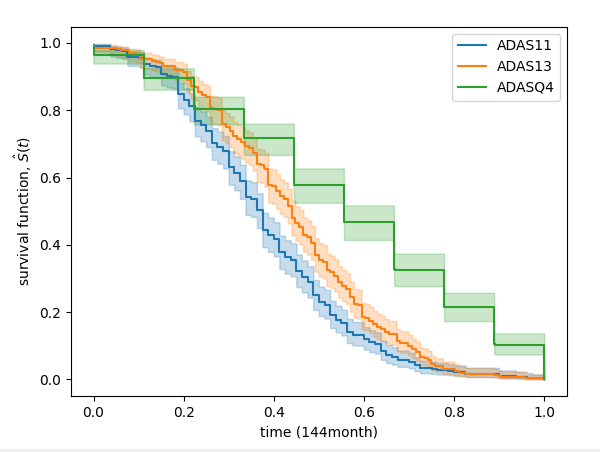
Step2 에서는 기본적인 Linear Layer인 Keras의 Dense를 사용



Result : 어려움은 없었으나 GAIN의 결과 데이터(Imputed data)가 나오지 않아, 오래 걸림, 기존의 Missing data를 사용하여도 꽤 높은 정확도가 나오는데 Overfitting이라고 생각함.

Step1과 Step2의 모델을 통하여, 이후의 개별 데이터를 적용하게 된다면, 정확하지 않거나 찾지 못하였던 데이터를 추정 값으로 찾아주고 높은 확률로 classification이 되므로 예상 Conv\_Time을 측정할 수 있음.

Step3는 Imputed 된 Data를 Kaplan-Meier Estimation을 적용하면 된다.



각각의 Feature를 확인할 수 있고, 더 Survival Function에 영향을 많이 주는 Feature를 파악하기에 좋은 기법인 것 같다.

[1]Jinsung Yoon, 2018, GAIN : Missing Data Imputation using Generative Adversarial Nets. ICML