

[과제] Load the wine data from chapter 4, and create a new model with the appropriate number of input parameters.)

### 1. 실행환경.txt

### 2. 과제#4\_결과보고서

### 3. main.py : 소스코드

# Dataset : tabular-wine/winequality-white.csv

# train set : dataset 80%

# validation set : dataset 80%

# loss function : MES

# optimizer : Adam

# learning rate : lr = 1e-2(고정)

### 4. main\_img 폴더 : epoch에 따른 train loss 와 validation loss 그래프

# hidden1#32\_epoch50.png : hidden layer 1(32nodes)의 model에서 epoch50회 실시

# hidden1#64\_epoch50.png : hidden layer 1(64nodes)의 model에서 epoch50회 실시

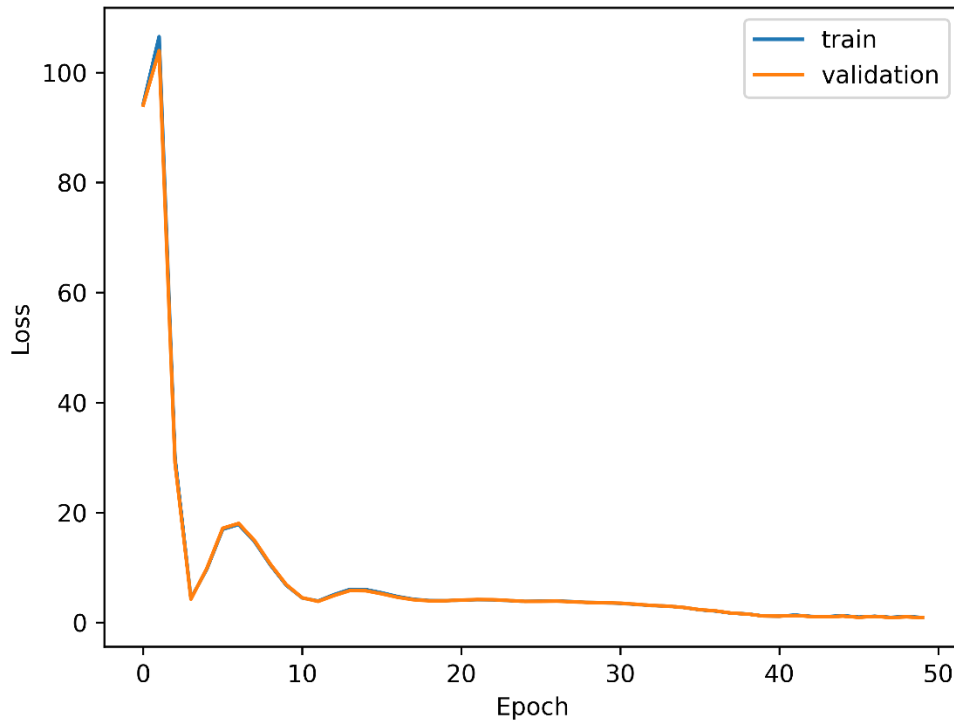
# hidden2#32#64\_epoch50.png : hidden layer 2(32, 64nodes)의 model에서 epoch50회 실시

# hidden2#64#128\_epoch50.png

# hidden2#256#256\_epoch50.

# hidden3#256#256#128\_epoch50.png

# hidden3#256#256#128\_epoch500.png



- hidden2#64#128\_epoch50.png -

와인 분석 데이터셋에서 모든 parameter를 사용하여, nn module 형태로 신경망을 구축하여 학습하면, train셋과 validation셋의 loss가 거의 일치한다. 따라서 제대로 학습함을 알 수 있다. 또한 epoch를 100이상 실행하지 않아도 충분히 loss가 감소한다.

다음 실험에서는 데이터셋을 변경해보자.

## 5. part\_wine\_dataset.xlsx

와인 분석을 위해 11개의 속성 값 fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol 중에서 눈으로 판단하여 연관있을 법한 속성값만 추려 데이터셋을 구성한다. 엑셀로 데이터셋을 불러온 후, quality에 대해 정렬한 다음, 해당 quality 수치에 따라 means를 구한다.

fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
7.6	0.33325	0.336	6.3925	0.0543	53.325	170.6	0.994884	3.1875	0.4745	10.345	3
7.129448	0.381227	0.304233	4.628221	0.050098	23.3589	125.2791	0.994277	3.182883	0.476135	10.15245	4
6.933974	0.302011	0.337653	7.334969	0.051546	36.43205	150.9046	0.995263	3.168833	0.482203	9.80884	5
6.837671	0.260564	0.338025	6.441606	0.045217	35.65059	137.0473	0.993961	3.188599	0.491106	10.57537	6
6.734716	0.262767	0.325625	5.186477	0.038191	34.12557	125.1148	0.992452	3.213898	0.503102	11.36794	7
6.657143	0.2774	0.326514	5.671429	0.038314	36.72	126.1657	0.992236	3.218686	0.486229	11.636	8
7.42	0.298	0.386	4.12	0.0274	33.4	116	0.99146	3.308	0.466	12.18	9

체크된, residual sugar, chlorides, free sulfur dioxide, density, pH, alcohol 속성들이 와인 퀄리티에 영향을 줄 것이라, 추측하여 학습을 진행한다. **part\_dataset.CSV**

## 6. part\_wine\_data.py

# Dataset : part\_dataset.CSV

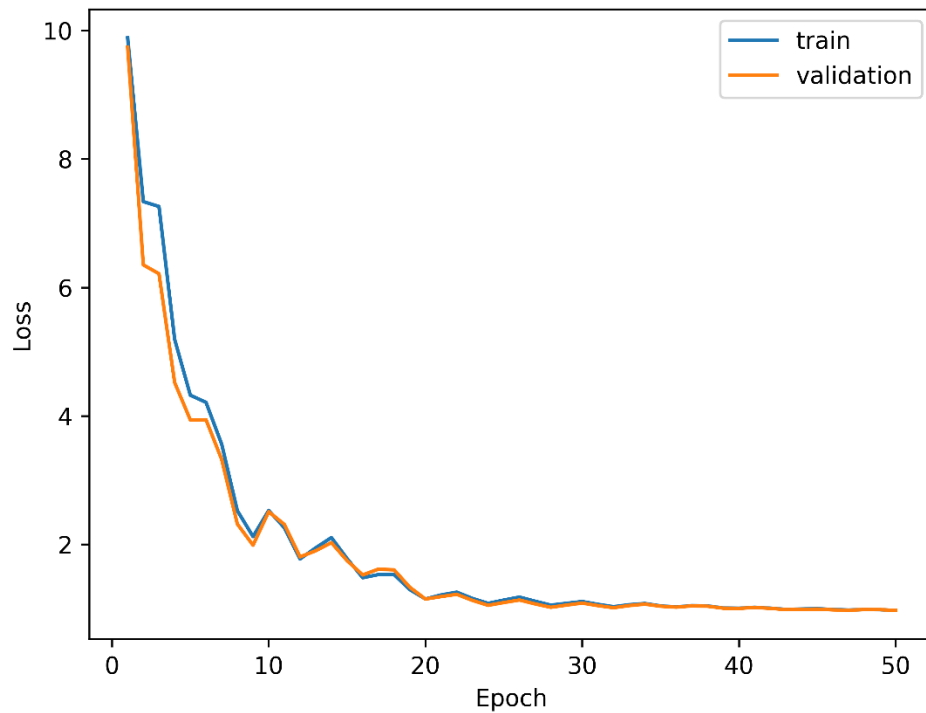
## 7. part\_wine\_data\_img 폴더

# hidden2#8#16\_epoch50.png

# hidden2#16#32\_epoch50.png

# hidden3#8#16#8\_epoch50.png

# hidden3#128#256#128\_epoch50.png



- hidden2#16#32\_epoch50.png -

데이터셋 크기가 작아지니까, 프로그램 실행속도가 조금 빨라졌다. 또한 충분히 loss를 감소하는데 손색없는 결과가 나타났다.

## 8. red\_wine\_data\_analysis.CSV

[참고1]<https://bibinmjose.github.io/RedWineDataAnalysis/> 를 참고하여, 실제 red wine에 영향 미치는 속성 세가지를 찾았다. **volatile acidity**, **sulphates**, **alcohol** 속성들로 dataset을 생성하여 학습을 진행하였다.

## 9. red\_wine\_data\_analysis.py

# Dataset : red\_wine\_data\_analysis.CSV

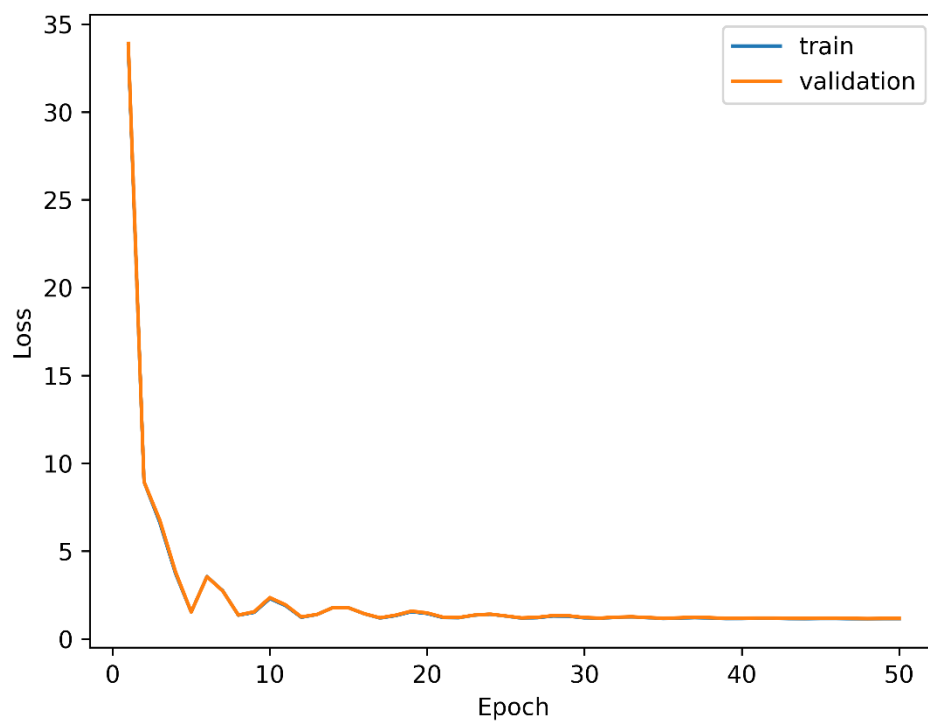
## 7. red\_wine\_data\_analysis\_img 폴더

# hidden1#8\_epoch50.png

# hidden1#8\_epoch500.png

# hidden2#64#128\_epoch50.png

# hidden3#64#128#64\_epoch50.png



# hidden3#64#128#64\_epoch50.png

volatile acidity, sulphates, alcohol 세가지 속성값만으로도 quality를 판단하기에 적합했다. 또한, 적절한 epoch로 조절하면 overfitting 없는 신경망을 구축할 수 있다. 또한 deep한 신경망일수록 초기 loss를 감소하는데 유리하다고 생각한다.