자연어처리및응용 중간고사 대체과제

컴퓨터공학부(컴퓨터공학과) 202239903 황규한

1번 문제: FordA dataset를 이용하여 시계열 데이터 분류

첫실험

Conv layer를 간단히 5개만 쌓고 스케쥴러를 사용해서 학습률조정

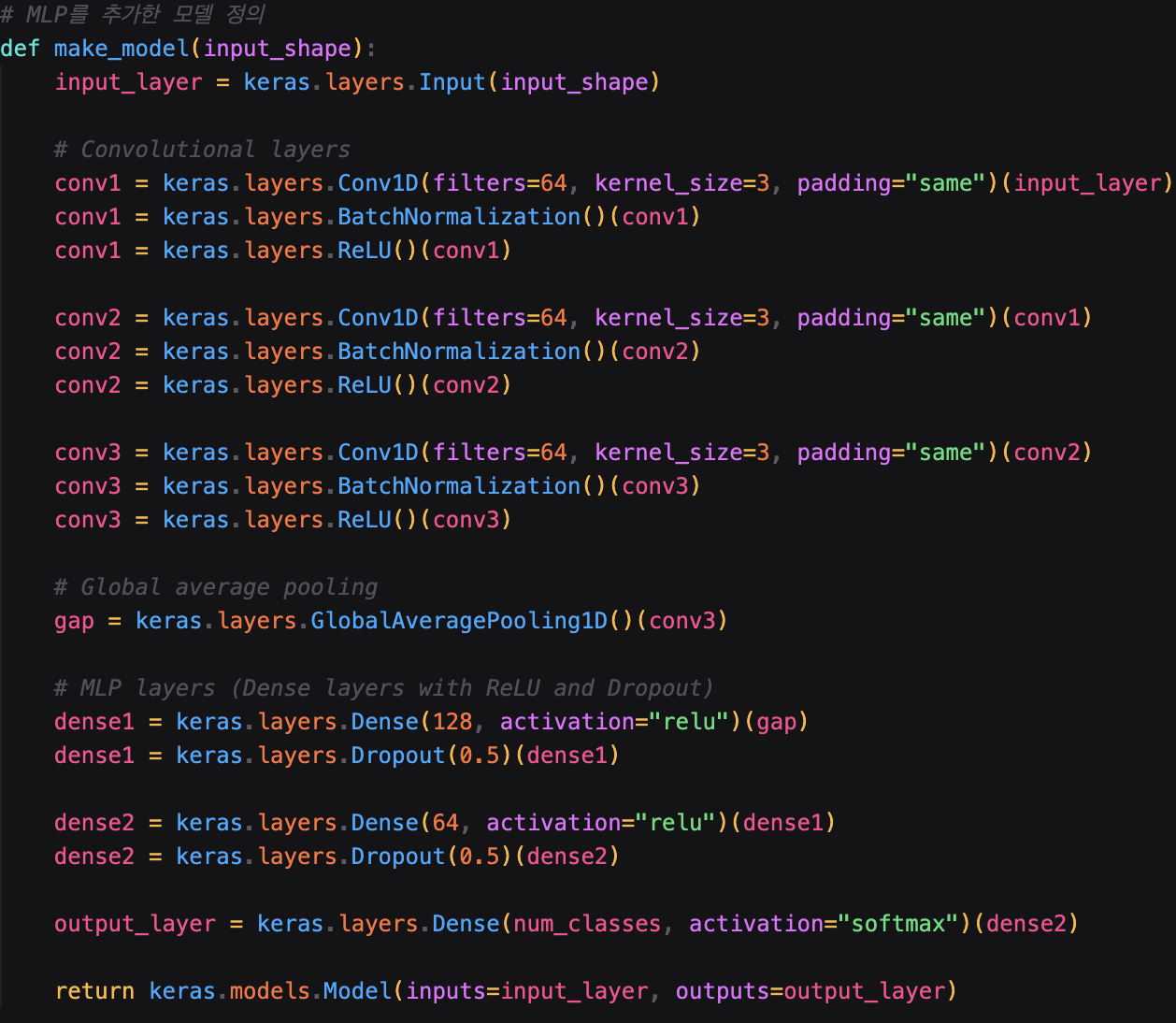
Epoch 500에 early stopping 적용 했더니 262에서 끊김  
Test accuracy: 0.967424213886261

Test loss: 0.0918307602405581

--------------------------------------------------------------------------------------------------------------

두번째 실험

아무생각없이 train acc,loss를 안나타내서 다시 구현함, 또 살면서 이런저런 논문 읽어보니 이유는 알 수 없으나 모델 마지막에 간단한 mlp를 세우기만해도 성능이 꽤 올라간다는 논문들이 많아 나도 간단하게 세워봄



간단하게 층 두개만 추가했을뿐인데 오히려 성능이 더 떨어짐

Train accuracy 0.9322410225868225

Train loss 0.17408673465251923

Test accuracy 0.9189394116401672

Test loss 0.19506072998046875

--------------------------------------------------------------------------------------------------------------

세번째 실험

다시 MLP빼고 conv 층도 3개로 줄임 그랬더니 성능이 매우 좋아짐!

하이퍼파라미터   
lr = 1.0000e-04

Epoch = 314 (Early Stopping)

Batch = 32

Classification Report

Train accuracy : 0.9869480729103088

Train loss : 0.04948848485946655

Test accuracy : 0.9704545736312866

Test loss : 0.09247986972332001

Class 0 (정상)

Precision: 0.9706744868035191

Recall: 0.9720998531571219

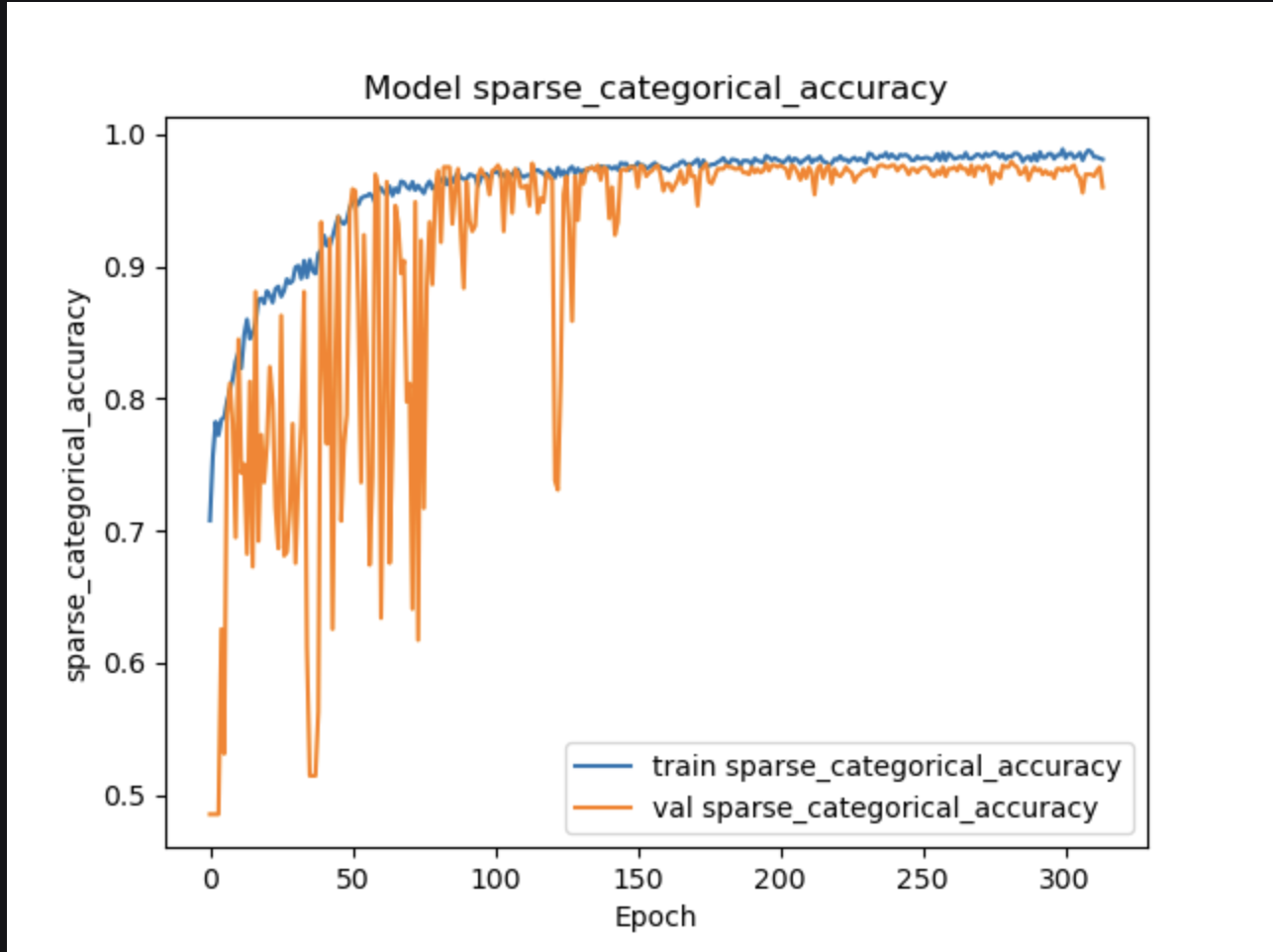
F1-score: 0.9713866471019809

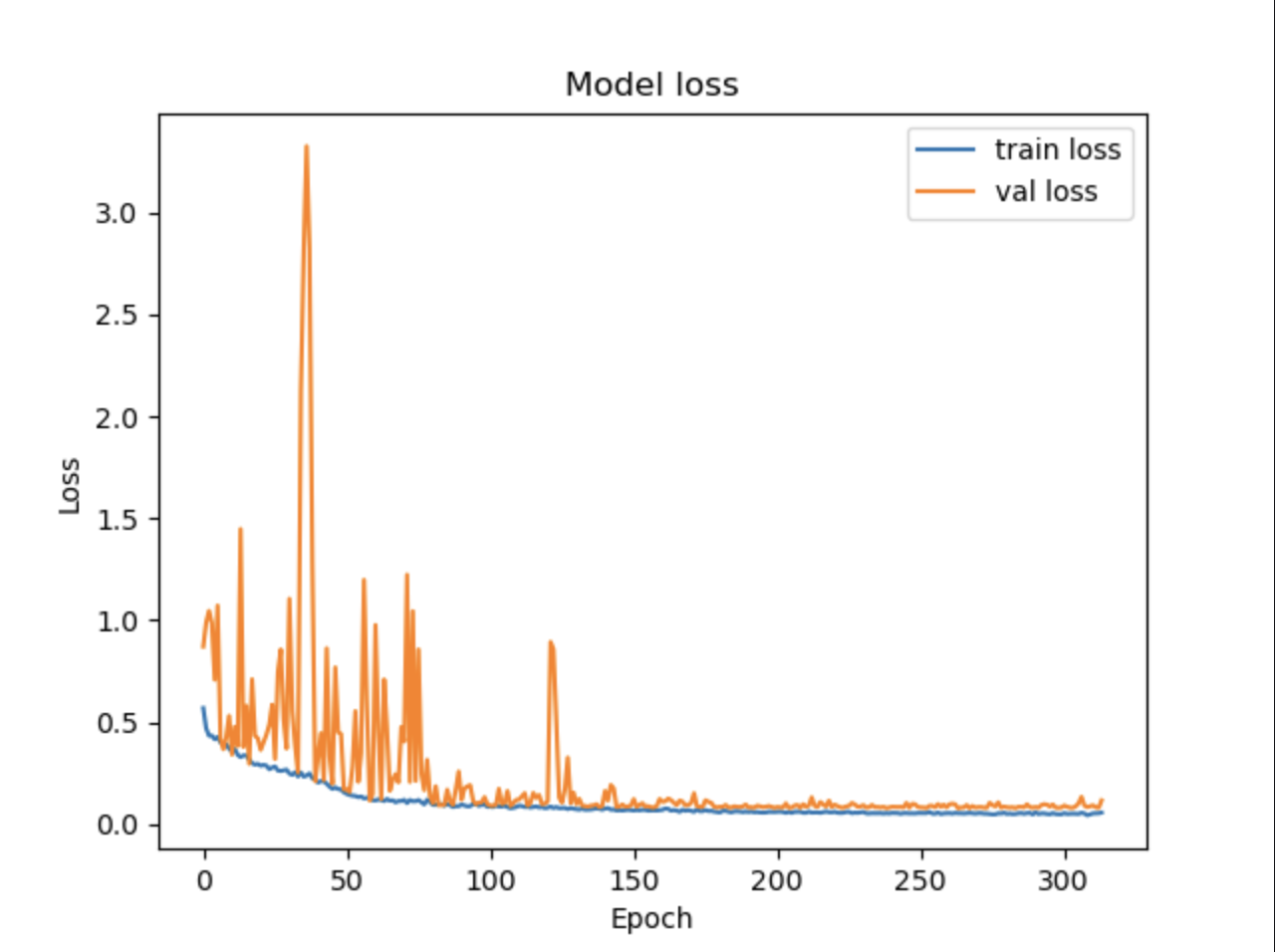
Class 1(비정상)

Precision: 0.9702194357366771

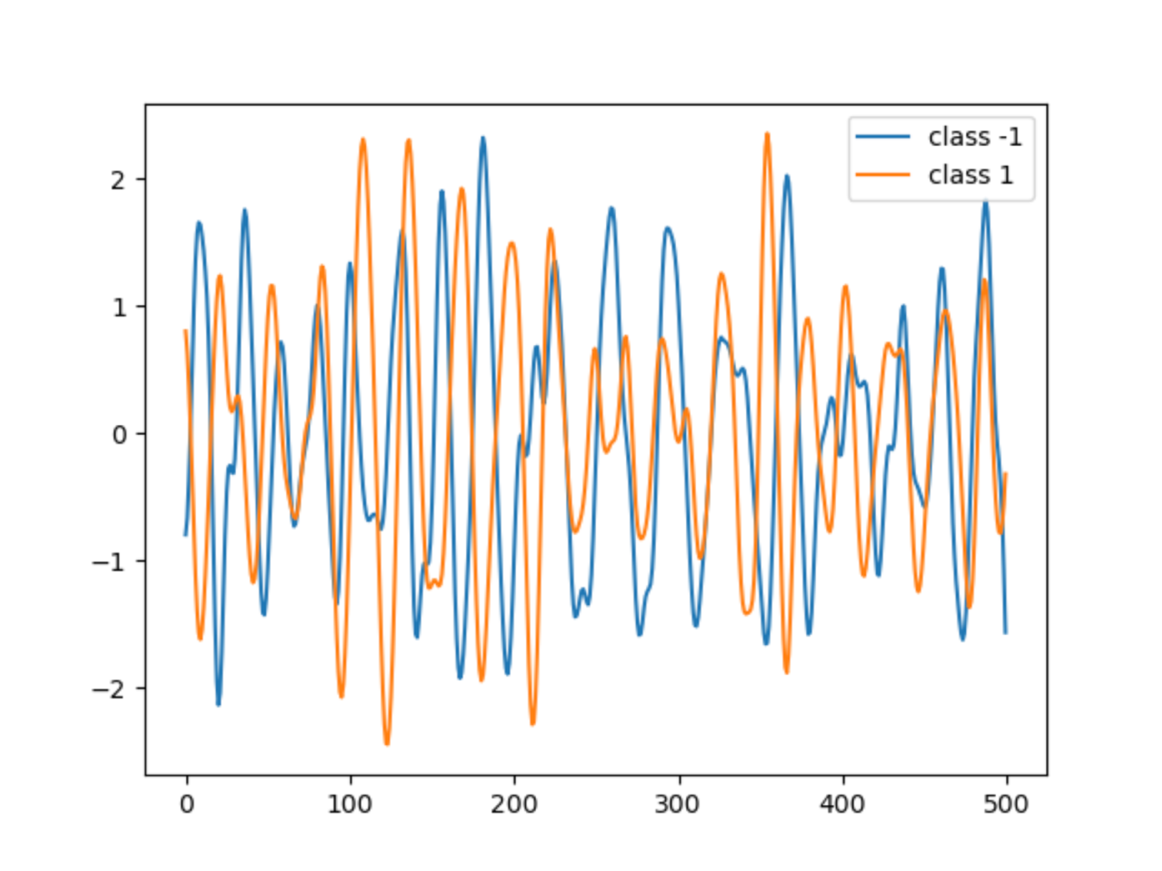
Recall: 0.9687010954616588

F1-score: 0.9694596711041503



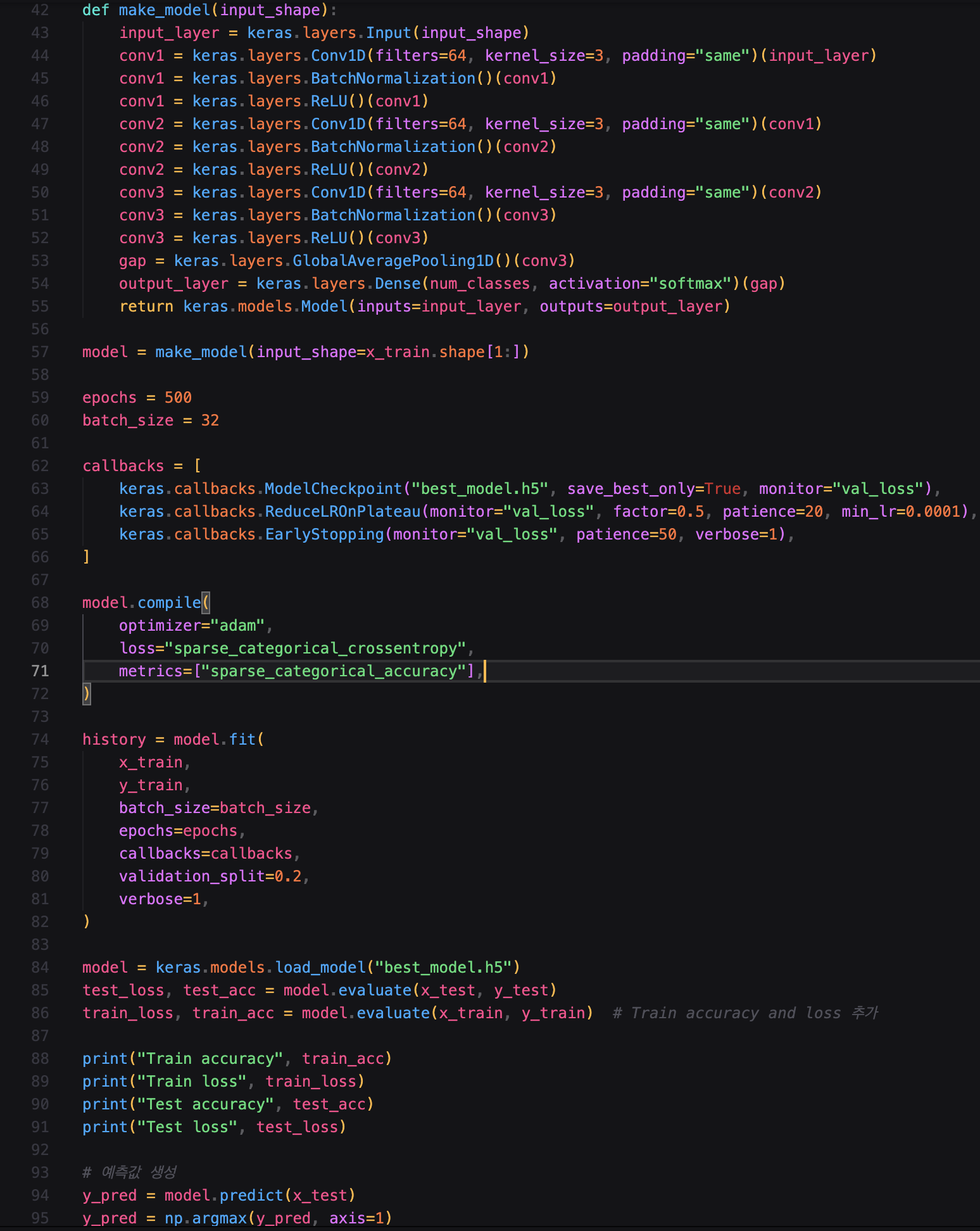


FordA\_class\_plot



Code







2번문제: 날씨 데이터를 LSTM으로 처리