

프로젝트 평가가이드

미션 정보

평가 미션

프로젝트 ID	프로젝트 이름	미션 ID	미션 이름
2	뇌 CT 분석	1	CT 이미지 분석
코드 자동 평가	퀴즈	동료 평가	멘토 평가
	√	√	

미션 설명

- 구글드라이브에 환자 진료 데이터를 저장하고 구글 드라이브를 구글 Collaboratory에서 사용할 있도록 마운트하라.
- 그리고 데이터들을 다양한 그래프로 출력하여 분석한다. 이미지는 학습을 위해 128, 128 크기로 저장한다.
- 또한 학습을 위해 데이터를 train 80%, validation 10%, test 10%로 나누어라.
- 데이터 변환으로 추가데이터를 확보한다는 것은 train 과 validation 데이터를 가공하여 추가 데이터를 확보한다는 의미이다.
- test 데이터는 가공할 이유가 없으므로 추가데이터 확보는 반드시 데이터를 train 과 test 로 구분하고 난 뒤에 train에 대해서만 가공한다.

결과물

- Image Data Generator 의 발표 동영상

평가 항목

구분	평가 항목
Data Store	1. try except 의 사용으로 데이터 액세스 에러에 대처했는가?
Data Store	2. 구글드라이브에 저장했는가?
Data Analysis	3. 그래프를 많이 만들어서 데이터 분석을 많이 할수록 높은 점수 부여
Data Analysis	4. 이미지는 128, 128 크기로 저장했는가?
Data Split	5. 데이터를 train, validation, test 로 나누었는가?
Image Data Generator	6. Image Data Generator 를 사용하였는가?
Image Data Generator	7. 발표를 잘할 수록 높은 점수 부여

퀴즈 문항

분류	문제	정답
Data Analysis	이미지를 pyplot 에서 출력하기 위해서 사용하는 메소드는?	imshow()
Data Split	퀴즈 1: train_test_split 을 사용할 때, 실행 시 마다 같은 학습 데이터를 얻고 싶다. 어떤 인자 값을 어떻게 해야 하는가?	random_state 에 임의의 값을 넣어준다. 시드를 고정하기 때문에 실생 시마다 같은 학습 데이터를 얻는다.

출제자의 의도

해당 미션을 통해 학생은 '왜 데이터 포맷에 대한 이해가 필요한지' 그리고 '데이터를 왜 가공하는지에 대한 이유'를 파악해야 한다.

또한 더 나은 학습 모델링을 위해 추가 데이터를 생성하는 방법을 이해하고 있어야 하며, 이를 위한 기술적인 부분을 알고 있어야 한다.

어떤 관점으로 학생의 미션을 평가해야 하는가?

해당 미션에서 가장 우선되어야 하는 것은 데이터를 분석하여 데이터의 포맷을 이해하는 것이다. 따라서 데이터의 포맷을 적절하게 이해하였는지를 중점적으로 판단해야 한다. 데이터를 분석하여 처리하는 궁극적인 목적은 머신러닝을 활용하여 모델을 만들기 위함이다. 따라서 '원하는(혹은 사용할) 머신러닝에 맞게 데이터를 가공하였는가'를 기준으로 '데이터의 포맷을 적절하게 이해하였는지' 판단한다.

또한 학습을 위해 추가 데이터를 생성하는 경우가 있는데, 이러한 상황을 대비하여 추가적인 데이터를 생성하는 기술적인 방법을 숙지하고 있는지 평가한다. 이를 위해 'Image Data Generator'와 같은 기술을 적절하게 사용하였는지 평가한다.

데이터를 training set 과 test set 으로 '왜 구분하는지', '어떤 방법으로 구분하는지'에 대한 이해도를 평가한다.