

데이터 불균형 (Data Imbalance): Oversampling, Undersampling

practice 인공지능,머신러닝 · 2024. 9. 28. 15:27

데이터 불균형 (Data Imbalance): Oversampling, Undersampling이 뭔가요.

- 데이터가 class별로 상당히 불균형(갯수차이가 심한경우) 하게 있는것입니다.

해결책

- oversampling : 갯수가 적은 class를 복사하듯이 여러개 뽑은것이고

- undersampling: 갯수가 많은 class를 일부만 뽑아서 활용

- 이두개가 완벽한 해결책은아님

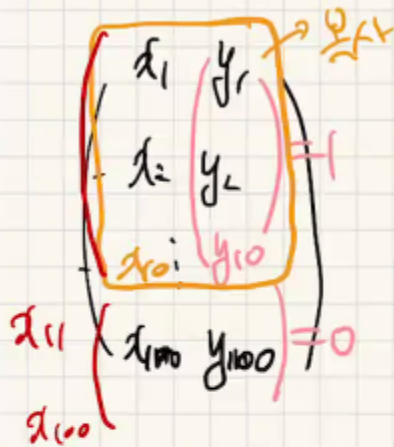
Binary Classification

0이 90개, 1이 10개일때

-oversampling : 1이 적으므로, 1이 10개를 9번 반복해서 90개를 만든다. 0이 90개라서 비율을 맞추려고

-undersampling : 0이 많으므로 0이 90개중에 10개만 뽑아서 1이 10개와의 비율을 맞춤

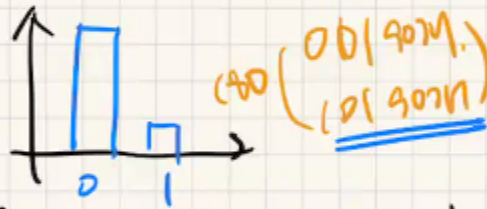
• Data Imbalance



$$y = 0, 1$$

Binary Classification

0이 90개, 1이 10개



Multi class classification



• oversampling

• undersampling

90개 중 10개

20개 (0이 10개, 1이 10개)

♡ 공감



...

구독하기

'practice 인공지능,머신러닝' 카테고리의 다른 글

교차검증 (0)

2024.10.11

전이학습 (Transfer Learning) (0)

2024.10.11

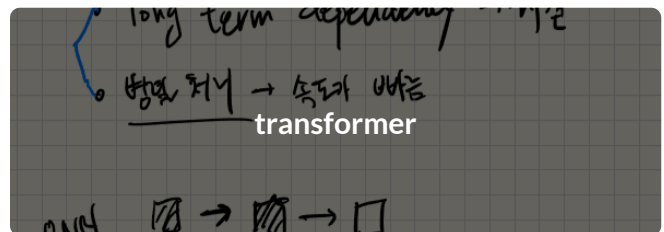
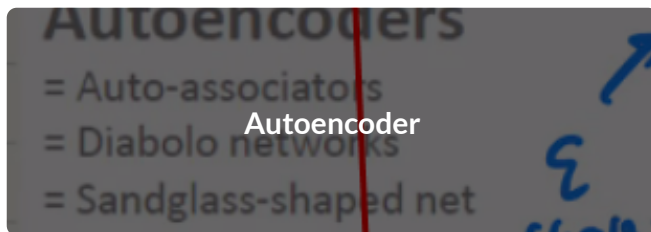
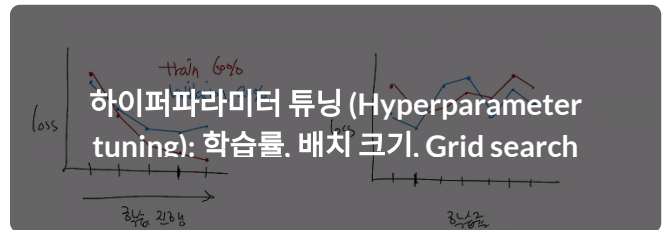
하이퍼파라미터 튜닝 (Hyperparameter tuning): 학습률, 배치 크기, Grid search (0)

2024.10.05

Autoencoder (0)

2024.10.04

관련글

[관련글 더보기](#)

자연어(NLP)

네이쳐2024 님의 블로그입니다.

[구독하기 +](#)

댓글 1



익명

비밀댓글입니다.

2024. 9. 28. 15:30

⋮



이름

비밀번호

내용을 입력하세요.

