

2_Classification_Model의_성능_평가_방법

문제 정의

- 분류 모델의 성능을 평가하는 것은 **베이지안 결정 이론**에 기반하여 입력값 X 에 대해 가장 올바르게 예측한 모델을 선택하는 것에 있어서 아주 중요하다. 이러한 성능을 평가하기 위한 방법으로 먼저 **제거 옵션(Rejection Option)**이 있다. 제거 옵션은 특정한 상황에 대해서 계산한 Posterior의 값이 임계값 보다 낮아 명확하지 않은 부분을 제거 범위로 지정하여 예측을 진행하지 않는 것이다. 이 외에도 클래스 혼동 매트릭스가 있다. 이는 각 클래스에 대한 모델의 예측값과 실제값을 비교하여 하나의 표로 나타낸 것이다. 이러한 혼동 매트릭스에서 가장 많이 사용되는 성능 지표는 바로 **정확도**이다. 정확도는 **전체 상황 중 정답을 맞춘 비율**을 뜻한다. 그러나 이러한 정확도는 **데이터가 불균형**하게 되면 정확하지 않다. 만약 False Sample이 압도적으로 많은 불균형 데이터셋이 있다면, 모델의 예측값은 False로 편향되어 정확도 수치는 높게 나오지만 실제로 True인 데이터를 놓칠 수 있기에 **정확도만으로 모델의 성능을 평가하는 것은 부적절하다**.

해당 문제에 대한 일반적인 접근

- **정밀도는 모델이 True라고 분류한 것 중에서 실제 True인 것의 비율**을 말한다. 이는 실제 정답이 False인 데이터를 True라고 잘못 예측하면 안되는 경우에 중요한 성능 지표가 된다. 이러한 정밀도를 높이기 위해서는 모델이 True라고 예측했지만 답이 False인 경우를 낮추는 것이 중요하며 0~1 사이의 값을 가지며 1에 가까울수록 좋다. **재현율은, 실제 True인 것 중에서 모델이 True라고 예측한 비율**을 말한다. 이는 실제 정답이 True인 데이터를 False라고 잘못 예측하면 안되는 의료계 쪽에서 중요한 성능 지표가 된다. 재현율을 높이기 위해서는 모델이 False라고 예측했는데 정답이 True인 경우를 낮추는 것이 중요하다. 이는 0~1 사이의 값을 가지며 1에 가까울수록 좋다. 이 두 성능지표는 각각 모델의 입장과 실제 정답의 입장에서 본 관점이기에 **두 수치를 복합적으로 사용한다**. 이를 **정밀도-재현율 곡선**이라고 하며 **우상단으로 갈수록 모델의 성능이 좋음**을 의미한다.

일반적인 접근법의 제한 사항

- 이러한 정밀도-재현율 곡선은 **데이터 불균형에 영향을 받는다**. 특히 모델이 True라고 예측했는데 실제값이 False라고 했을 경우 정밀도 값에 큰 영향을 끼친다. 정밀도의 경우, True 데이터가 매우 적을 때 적은 수의 예측만으로도 그 값이 크게 변할 수 있다. 그리고 이렇게 **데이터가 불균형할 경우, 정밀도-재현율 곡선은 정밀도가 높더라도 재현율이 낮을 수 있고 반대로 재현율이 높더라도 정밀도 값이 낮을 수 있다**. 그렇기에 데이터가 불균형한 경우 정밀도-재현율 곡선은 이러한 상황을 충분히 설명하지 못한다.

제한 사항에 대한 해결 방안

- **F1 Score**는 이러한 상황에서 모델의 성능을 정확하게 평가하기 위한 성능지표이다. F1-Score는 **정밀도와 재현율의 조화 평균**인 다음 식으로 정의된다. $F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$. 이는 두 성능지표의 균형을 맞추는 역할을 한다. 데이터가 불균형할 경우 정밀도 혹은 재현율 중 하나의 수치가 훨씬 높아져도 균형을 맞춰주기에 **F1 Score는 더욱 효과적인 성능지표**가 될 수 있다. 이러한 F1 Score는 데이터가 불균형할 때에도 성능을 정확하게 평가할 수 있으며 최종적인 **분류 모델의 성능을 하나의 숫자로 표현 가능**하다는 점에서도 장점을 지니고 있다.