

딥러닝을 이용한 이상치 검출 방법

<https://youtu.be/hm27RaJVq3k>

접근 방식

모델 평가 방법

실습

접근 방식

1. 규칙 설정

Examples

1. 일상적으로 구매하던 지역과 다른 지역에서 사용 되었는가?
2. 다른 날과 다르게 얼마나 빈번하게 사용 되었는가?
3. 신뢰할 수 없는 계좌에서 큰 금액을 송금 받거나 했는가?
4. 동일 IP에서 단기간에 여러 계좌가 생성되어 송금에 사용되었는가?

→수집된 Data를 활용하여 이상 거래를 예측해야 함.

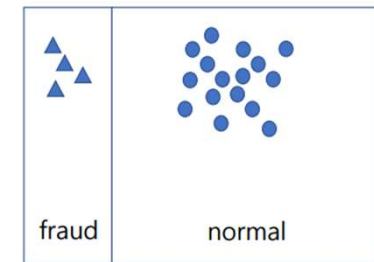
But. 전체 Data 중에서 이상 거래는 아주 소량의 데이터 and 빠른 반응

접근 방식

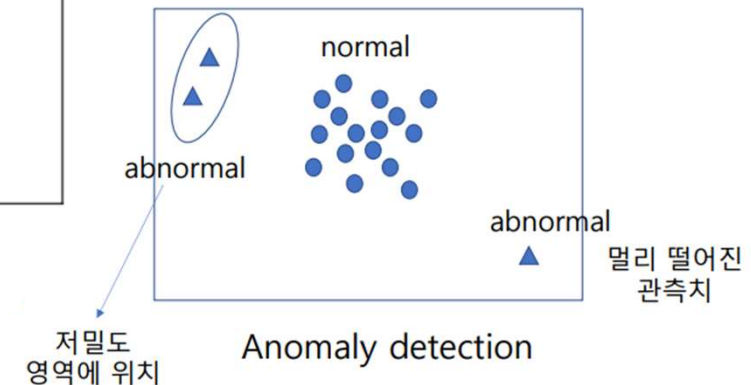
이상감지 (Anomaly Detection)

| | |
|----------------------------------|--|
| Outlier detection (이상치 감지) | <ul style="list-style-type: none"> • 훈련데이터에 다른 관측치와 멀리 떨어진 관측치로 정의되는 이상치가 포함. • 따라서 이상치 탐지기는 비정상 관찰을 무시하고 훈련 데이터가 가장 집중된 영역을 맞추려고 한다. • 저밀도 영역에 위치한다고 가정. 따라서, 조밀한 클러스터 형성 않음. |
| Novelty detection (특이치 감지) | <ul style="list-style-type: none"> • 훈련데이터는 정상 데이터로만 구성. 따라서 새로운 관측치가 훈련 데이터와 비교해서 많이 떨어져 있는지 여부를 감지. → 특이치 라고도 함. • 정상으로 간주되는 학습데이터의 저밀도 영역에 있는 한 조밀한 클러스터를 형성할 수 있다. |

label 기반 지도 학습



이진 분류



접근 방법

Anomaly Detection 방법 분류

| 구분 | 접근 방법 | 측정 기준 | 알고리즘 |
|----------------------------|-------------------------|-------------------------------------|---|
| 전통적 Machine Learning | 지도 학습 방법 | Label 데이터 | 이진 분류 |
| | Random split 방법 | Number of random splits | Isolated Forest |
| | Proximity(근접도) 기반 방법 | clustering, density, distance | DBSCAN, LOF(Local Outlier Factor), K-nearest Kernel method |
| Deep Learning | Deviation(편차) 기반 방법 | 재구성 오류 | Autoencoder |
| | 통계적 방법 | 확률 분포의 차이 | Variational Autoencoder |

모델 평가 방법

Confusion Matrix

| | | True condition | |
|---------------------|------------------------------|--------------------------|--------------------------|
| Total population | | Condition positive | Condition negative |
| Predicted Condition | Predicted condition positive | TP True positive | FP False positive |
| | Predicted condition negative | FN False negative | TN True negative |

TP → 1을 1로 제대로 분류, FP → 0을 1로 잘못 분류

FN → 1을 0으로 잘못 분류, TN → 0을 0으로 제대로 분류

정확도(Accuracy) = $(TP+TN) / (TP+TN+FP+FN)$

모델 평가 방법

- Precision = $TP / (TP + FP)$

- 정밀성. → Model 이 sample 을 True 로 분류했을 때 얼마나 자주 맞추었는가 ?
- positive 분류의 정확성 측정 (1 에 가까울 수록 좋음)

| | | True condition | |
|---------------------|------------------------------|--------------------|--------------------|
| Total population | | Condition positive | Condition negative |
| Predicted Condition | Predicted condition positive | True positive | False positive |
| | Predicted condition negative | False negative | True negative |

ex) 포르노 영상 검출기 – 포르노로 분류했을 때 실제 포르노인 비율
security check 영상 탐지기 – 통과 승인된 사람 중 실제 직원 비율

모델 평가 방법

• **Recall** (Sensitivity/ True positive Rate) = $TP / (TP + FN)$

- 민감도. **전체 Positive 데이터 중에서 Positive로 분류한 비율**
(1 에 가까울 수록 좋음)
- Positive case 를 놓치고 싶지 않은 경우의 성능 측정

| | | True condition | |
|---------------------|------------------------------|--------------------------------|-------------------------------|
| Total population | | Condition positive | Condition negative |
| Predicted Condition | Predicted condition positive | True positive | False positive → Type I error |
| | Predicted condition negative | False negative → Type II error | True negative |

ex) 포르노 영상 검출기 – **전체 포르노 중 포르노로 분류된 비율**
 security check 영상 탐지기 – **전체 직원 중에서 통과로 분류된 비율**

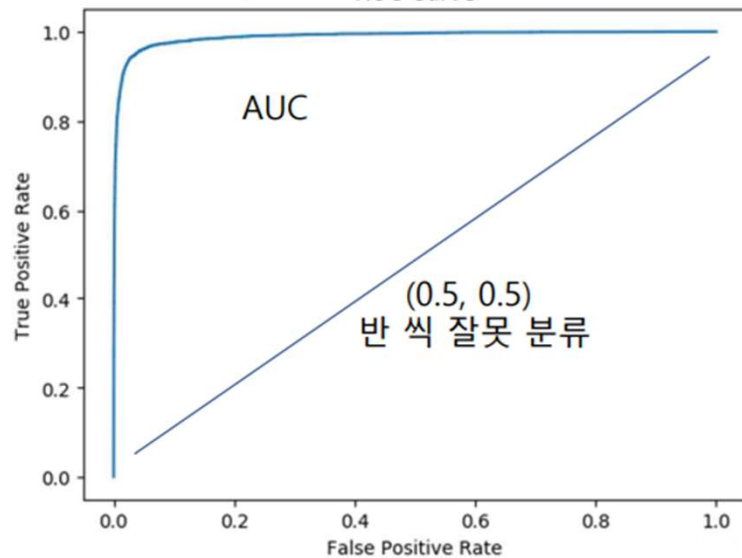
모델 평가 방법

ROC Curve

Recall

$$TPR = \frac{TP}{TP+FN} = \frac{TP}{\text{al positive}}$$

ROC Curve



$$FPR = \frac{FP}{FP+TN} = \frac{FP}{\text{al negative}}$$

Negative 를 Positive 로 잘못 분류한 비율

→ TPR은 1에 가까울 수록 좋고 FPR은 0에 가까울 수록 좋다

실습

실습진행.

Q & A