

Chatpor3. 분류

(3.4 다중분류~3.7 다중 출력 분류)

22.04.27

화학안전연구센터 최지원

스터디 목차

1. 이중 및 다중 분류

2. 다중 분류

3. 에러 분석

4. 예제 코드


1. 이중 및 다중 분류

Pick one

Label 1	✓
Label 2	

Binary

Binary Classification




```
graph TD; Root[ ] --> Dog[Dog 0.9]; Root --> NotDog[Not Dog 0.1];
```

Pick one

Label 1	
Label 2	
Label 3	
Label 4	✓
...	
...	
Label L	

Multi-class

Multiclass Classification




```
graph TD; Root[ ] --> Dog[Dog 0.5]; Root --> Cat[Cat 0.09]; Root --> Bus[Bus 0.01]; Root --> Plant[Plant 0.4];
```

Pick all applicable

Label 1	
Label 2	✓
Label 3	
Label 4	✓
...	
...	
Label L	✓

Multi-label

Multilabel Classification



```
graph TD; Root[ ] --> Dog[Dog 0.8]; Root --> Cat[Cat 0.2]; Root --> Bus[Bus 0.04]; Root --> Plant[Plant 0.7];
```

Pick all applicable

<Class1>


Label 1	✓
Label 3	

<Class2>

Label 4	
Label L	✓

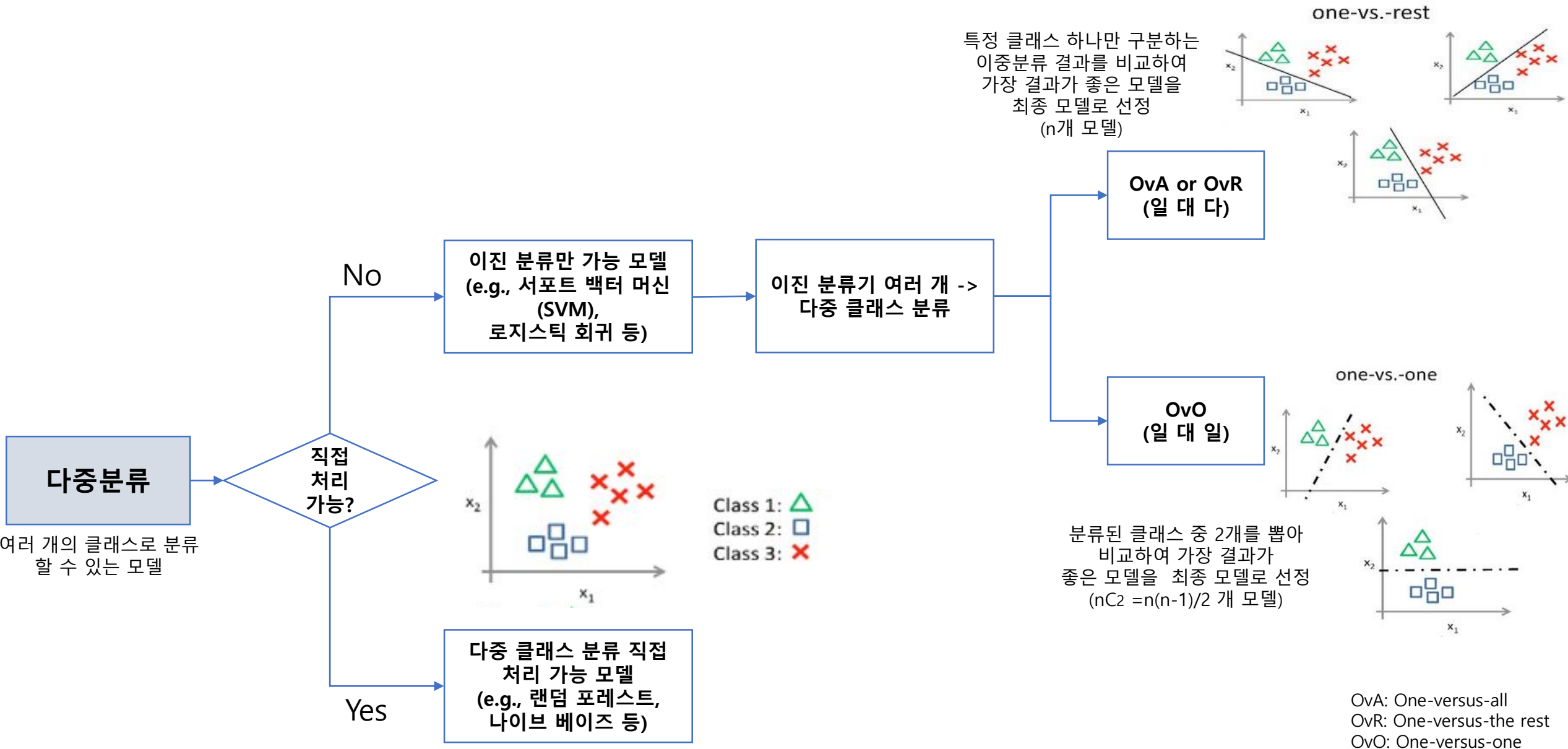
Multioutput-multiclass

Multioutput-multiclass classification



```
graph TD; Root[ ] --> Animal[Animal]; Root --> NonAnimal[Non-animal]; Animal --> Dog[Dog 0.8]; Animal --> Cat[Cat 0.2]; NonAnimal --> Plant[Plant 0.9]; NonAnimal --> Bus[Bus 0.1];
```

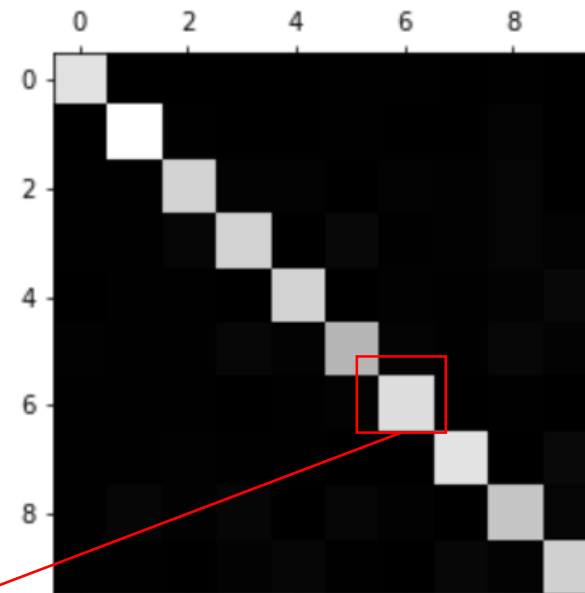
2. 다중 분류



3. 에러 분석

- 모델의 성능을 향상시킬 한 가지 방법으로, 훈련 데이터 셋에 대해 만들어 낸 예측의 에러 종류를 분석
- 분류모델 성능을 평가하는데 좋은 방법인 **오차 행렬**을 활용
 - 행: 훈련 데이터(클래스별 이미지); 열: 훈련 데이터 예측값(클래스별 분류된 이미지)
- 오차행렬의 주대각선 -> 데이터 셋이 올바르게 분류된 수

```
array([[5728, 2, 25, 10, 12, 48, 43, 10, 40, 5],  
       [ 1, 6501, 44, 25, 6, 35, 5, 8, 105, 12],  
       [ 47, 35, 5378, 86, 82, 21, 88, 57, 152, 12],  
       [ 47, 46, 142, 5333, 3, 225, 36, 54, 147, 98],  
       [ 14, 26, 39, 10, 5354, 9, 56, 34, 86, 214],  
       [ 67, 47, 38, 179, 80, 4602, 114, 33, 178, 83],  
       [ 35, 29, 48, 4, 42, 85, 5622, 7, 46, 0],  
       [ 27, 26, 68, 29, 47, 10, 6, 5794, 17, 241],  
       [ 47, 159, 73, 163, 18, 153, 59, 25, 5018, 136],  
       [ 41, 32, 25, 84, 153, 34, 3, 198, 86, 5293]])
```



숫자 5 에 행렬이 상대적으로 어두움

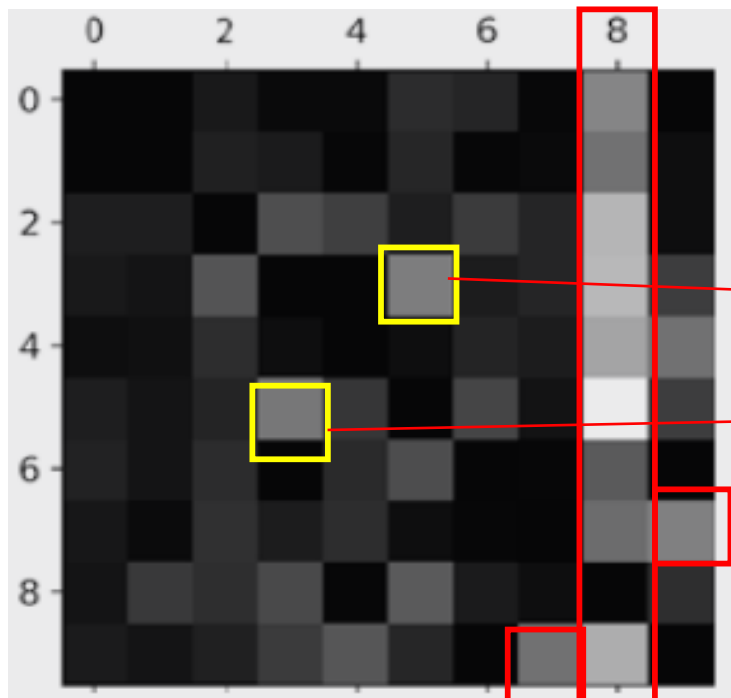
- 1) 숫자 5의 이미지가 상대적으로 적음
- 2) 숫자 5를 다른 숫자 만큼 잘 분류하지 못함



에러 분석 적합하지 않음(클래스마다 이미지 개수 동일하지 않음)

3. 에러 분석

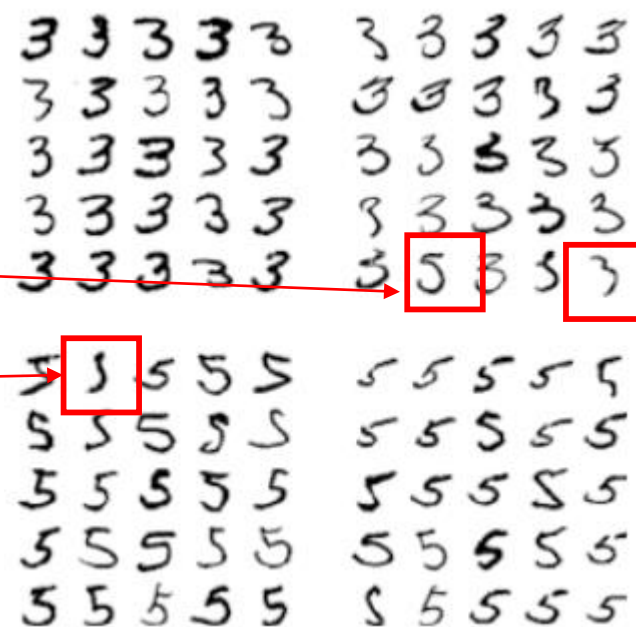
- 에러 분석을 위해 데이터의 오분류를 비율로 변환
 - 오분류 비율 = 클래스별 분류된 이미지 수 / 클래스별 총 이미지 수
- 오류만 집중적으로 분석하기 위해 오차행렬의 대각원소 값들은 0으로 변환



<종합적인 에러 분석>

- 클래스 8의 열 상대적 밝음 -> 다른 숫자가 8로 잘못 분류
- 숫자 3과 5** 및 7과 9가 서로 많이 혼동

➡ (개선방안) 8처럼 보이는 훈련 데이터로 분류 모델 학습, 동심원같은 숫자 패턴이 드러나게 이미지 전처리



<개개의 에러 분석>

- (우측상단) 클래스 3 -> 클래스 5 로 잘못 분류
- (좌측하단) 클래스 5 -> 클래스 3으로 잘못 분류

➡ (개선방안) 이미지에서 숫자를 중앙에 위치시켜 회전되지 않도록 전처리(3, 5 이미지 위치 및 회전방향 고려)

4. 예제 코드

- <https://colab.research.google.com/drive/1tLoFSg1nS3xxSsHfuAtJsbldgPvR-LWr#scrollTo=U0VEFO6QPDxy>