



deeplearning.ai

Error Analysis

Cleaning up
Incorrectly labeled
data

Incorrectly labeled examples

(cf. mislabeled example : 러닝 알고리즘의 output이 틀린 녀석을 말함)

an example
with an incorrect label

x



y

1

0

1

1

0

1

o Training set.

-> 딥러닝 알고리즘은 트레이닝 세트에 존재하는 랜덤 에러에 대해서 꽤 robust하다. 그래서 이런 에러들이 랜덤에서 크게 벗어나지 않는다면 광 뒤도 괜찮으니 이거 고치는데 시간을 많이 쓰지 마라.

DL algorithms are quite robust to random errors in the training set.

Systematic errors에는 좀 덜 robust하다는 점에 주의해라.

예를 들어, 레이블 작업하는 사람이 흰 아기 강아지를 고양이라고 하는 경향이 있다면 이 분류기는 흰 강아지를 고양이로 분류하도록 학습되어질 것이다.

Error analysis

트레이닝 셋이 아니라 dev/test set에 오류가 있는 경우는 어떻게 할까? error analysis할 때 이렇게 따로 컬럼을 마련해서 같이 분석한다.

Image	Dog	Great Cat	Blurry	Incorrectly labeled	Comments
...					
98				✓	Labeler missed cat in background
99		✓			
100				✓	Drawing of a cat; Not a real cat.
% of total	8%	43%	61%	6%	

이렇게 label이 잘못된 애들이 있을 때 트레이닝과 마찬가지로 애들을 고치는게 잘 하는 짓인지는 아래와 같이 판단.

Overall dev set error

Errors due to incorrect labels

Errors due to other causes

2% 이런 경우엔 label 잘못된 비중이 꽤 크기 때문에 여기에도 신경을 써야 할 상황일 수도 있다.

0.6%

1.4%

2.1% 1.9% 특히 이런 경우 A가 더 좋은 모델이라고 할 수 없는게 잘못된 label의 영향 때문에 이런 결과가 나왔을 수도 있기 때문이다.

↑ 이런 경우엔 여기에 포커스를 두는게 맞지

Goal of dev set is to help you select between two classifiers A & B.

이런 경우 incorrectly labeled example 고치는게 의미있는 작업이 될 것이다. Andrew Ng

Correcting incorrect dev/test set examples

- Apply same process to your dev and test sets to make sure they continue to come from the same distribution

잘못한 놈만 살펴보고 고치면 이게 bias로 작용해서 unfair advantage가 될 수 있으니 잘한 놈도 살펴보자.

- Consider examining examples your algorithm got right as well as ones it got wrong.
근데 쉽지는 않아서 항상 하지는 않음. 왜냐하면 모델 성능이 좋을 수록 잘하는 비율이 높아지니까(98:2) 살펴봐야 할 잘한 example의 수가 엄청 많아지기 때문이다.
- Train and dev/test data may now come from slightly different distributions.

근데 괜찮아. dev/test는 같은 distribution에서 와야되는게 super important한데 training은 좀 달라도 robust하거든.
게다가 training set은 엄청 많을거라서 고치기도 힘들걸