

딥러닝 스터디

C3W1L3-12

이승목

2020.07.21

TensorFlow 학습 자료

- TensorFlow 홈페이지에서 머신러닝 단기 집중 과정을 제공해줌. 관촬아보임.
 - <https://developers.google.com/machine-learning/crash-course>

C3W1L3-12. 목표 설정하기

목표설정하기	사람 수준의 성과와 비교하기
 단일 실수 평가 기준 업데이트 : 2020.07.09 ♥ 5	 왜 사람 수준의 성능을 봐야할까요? 업데이트 : 2020.07.09 ♥ 5
 최적화 척도 만족시키기 업데이트 : 2020.07.09 ♥ 4	 회피 가능 편향 업데이트 : 2020.07.09 ♥ 4
 Train/Dev/Test 세트 분포 업데이트 : 2020.07.09 ♥ 4	 사람 수준의 성능 이해하기 업데이트 : 2020.07.09 ♥ 5
 Dev 와 Test 세트의 크기 업데이트 : 2020.07.09 ♥ 5	 사람 수준의 성능 뛰어넘기 업데이트 : 2020.07.09 ♥ 4
 언제 Dev/Test 세트를 바꿔야할까요? 업데이트 : 2020.07.09 ♥ 4	 모델 퍼포먼스 향상시키기 업데이트 : 2020.07.10 ♥ 5

단일 평가 기준의 필요

- 두 모델이 아래와 같이 서로 다른 방면으로 장점이 있는 경우 선택이 어렵다.

모델	Precision: 고양이로 분류된 사진들 중 실제 고양이의 비율	Recall: 실제 고양이 중 고양이로 분류된 사진의 비율
A	95%	90%
B	98%	85%

- 둘을 적절히 고려하는 하나의 평가 기준 필

단일 평가 기준 예시

- F1 점수: precision과 recall의 조화평균

모델	Precision: 고양이로 분류된 사진들 중 실제 고양이의 비율	Recall: 실제 고양이 중 고양이로 분류된 사진의 비율	F1 점수
A	95%	90%	92.4%
B	98%	85%	91.0%

- 평균 오답률

모델	American Cat	Chinese Cat	Indian Cat	Other Cat	Average
A	3%	7%	5%	9%	6%
B	5%	6%	5%	10%	6.5%
C	2%	3%	4%	5%	3.5%
D	5%	8%	7%	2%	5.25%

최적화 척도와 조건 척도

- 특정 척도에 대해서는 과락 조건을 적용.
 - Ex. 시간(조건 척도)이 100ms 이하로 소요되는 모델 중에서 정확도(최적화 척도)가 높은 모델을 고르자.

모델	Accuracy	Running Time
A	90%	80ms
B	92%	95ms
C	95%	1,500ms

- Ex. Noise contamination(조건 척도)이 1% 이하이면서 signal efficiency(최적화 척도)가 높은 모델을 만들자.

Train/Dev(Val)/Test Set의 분포

- Dev와 Test set은 실제로 수행할 과제에서 중요한 것들이 충분히 반영되도록 설정.
- Dev와 Test set의 분포는 동일해야 한다.
- 데이터가 부족했던 과거에는 Train:Test = 7:3, Train:Dev:Test = 6:2:2 정도로 사용.
- 현대 빅데이터 시대에는 Train, Dev set > 10K or 100K 정도가 되도록 설정하여 사용.

평가 척도, Dev/Test Set 교체

- 모델이 평가 척도 상으로는 훌륭하더라도 실제 이용에 있어서는 효과적이지 못할 수 있다.
 - 고양이 사진 분류기가 정확도는 높는데 간혹 음란물을 고양이로 분류
 - AI 의사가 정확도는 높는데 영유아에게 유독 약물을 처방
- 이 경우 (1) 평가 척도에 가중치를 부여하거나
- (2) Dev/Test set이 해당 케이스를 충분히 반영하도록

사람 수준 오차

- 많은 경우 사람을 뛰어넘는 것을 목표로 하며, 일부 뛰어넘기도 했다.
 - ex. 광고, 영화 추천, 유통 예측 등
 - 구조화된 빅데이터 문제의 경우 사람 수준을 뛰어넘기 쉽다.
 - 다만 최근에는 일부 음성인식, 파형 분석, 사물 인식 등에서도 사람 수준 오차를 뛰어넘기도

사람 수준 오차

- 그런데 사람 수준 오차는 어떻게 정의할까?

Medical image classification example:

Suppose:

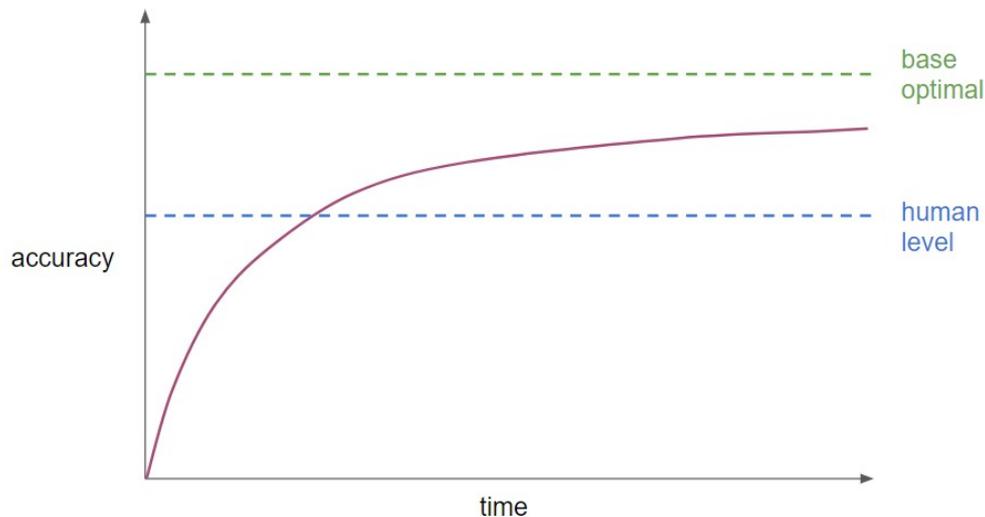
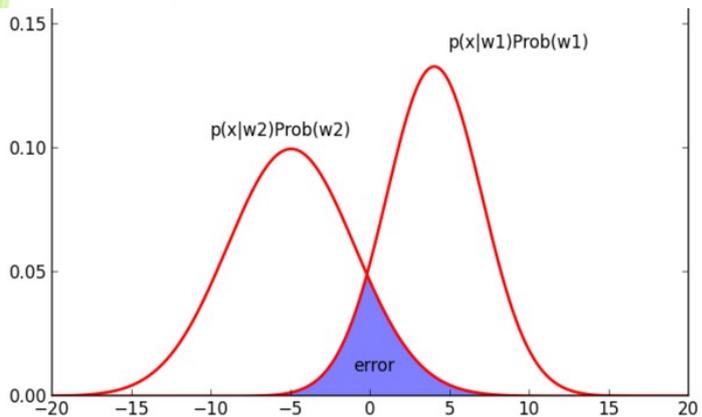
- (a) Typical human 3 % error
- (b) Typical doctor 1 % error
- (c) Experienced doctor 0.7 % error
- (d) Team of experienced doctors .. 0.5 % error



What is “human-level” error?

베이지스 오차와 사람 수준 오차

- 베이지안 오차: 이론상 가능한 오차의 최저값.
- 사람 수준의 오차를 베이지안 오차의 근삿값으로 자주 사용한다.



회피 가능 편향과 분산

- 회피 가능 편향: 학습 오차 - 베이지안 오차
 - 이론상 개선할 수 있는 오차율
- 분산: 개발 오차 - 학습 오차
 - Over fitting으로 인해 발생.

사람 수준 오차

- 베이지안 오차의 근삿값으로는 (d)
 - 프로그램 성능을 자랑하기 위해서는 (b), (c)도 가능

Suppose:

- (a) Typical human 3 % error
- (b) Typical doctor 1 % error
- (c) Experienced doctor 0.7 % error
- (d) Team of experienced doctors .. 0.5 % error



What is “human-level” error?

Andrew Ng

지도학습 모델 오차 분석

- 지도학습의 목적은
 - Train set에 잘 맞고 (회피 가능 편향 줄이기)
 - Dev/Test set에서도 잘 맞기 (분산 줄이기)

모델	A	B	C
학습 오차	5%	1%	0.7%
개발 오차	6%	5%	0.8%
베이스 오차가 0.7%	회피 가능 편향 4.3% 분산 1%	회피 가능 편향 0.3% 분산 4%	회피 가능 편향 0% 분산 0.1%
베이스 오차가 0.5%	회피 가능 편향 4.5% 분산 1%	회피 가능 편향 0.5% 분산 4.5%	회피 가능 편향 0.2% 분산 0.1%

지도학습 모델 최적화

- 회피 가능 편향이 있는 경우
 - 더 큰 모델 이용
 - 더 오래 훈련, 더 나은 optimizer
 - 다른 구조 탐색, 초 매개변수 튜닝
- 분산이 큰 경우
 - 더 많은 Train set 사용
 - 정규화 진행 (Dropout, Data augmentation)
 - 다른 구조 탐색, 초 매개변수 튜닝