

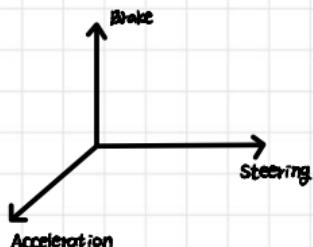
Orthogonalization

원하는 효과를 위해 변수를 조정하는 과정

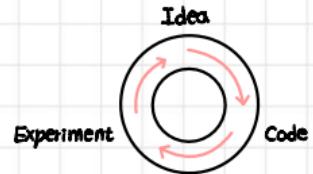


Car function

- Steering
- Acceleration
- Brake



→ 제어요소가 분리되어 있어야
원하는대로 조정하기 쉬움



chain of assumptions in ML

Fit training set well on cost function

\downarrow
Bigger network
Better optimization (ex. Adam)

Fit dev set well on cost function

\downarrow
Regularization
Bigger training Set

Fit test set well on cost function

\downarrow
Bigger dev set

performs well in real world

Change dev set or cost function

Using a Single number evaluation metric

평가 기준이 여러개일 경우,
어떤 모델이 좋은 모델인지 선택하기 어려움

classifier	trade off		Single F1 Score
	precision	Recall	
A	95%	90%	92.4%
B	98%	85%	91.0%

precision : 모델이 분류한 정답 중 실제 정답의 비율

recall : 실제 정답 중 모델이 정답으로 분류한 것의 비율

$$\text{F-1 score} = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{Recall}}}$$

Dev Set과 하나의 정량적 평가 기준을 사용하게 되면 모델 선택이 쉬움
 → 알고리즘 개선의 순환 속도 향상

많은 것을 고려해야 할 때, 성능을 최대로 높이고
 싶은 하나를 최적화 속도로 두고, 조금 더 중요한
 목표를 조건 속도로 설정하여 평가

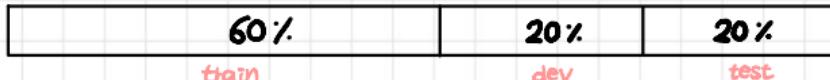
Classifier	Accuracy	Running time
A	90%	80ms
B	92%	95ms
C	95%	1,500ms

Accuracy를 높이면서 Running time이 100ms를 넘지 않도록!
 최적화 속도 조건 속도

Train / Dev / Test Set Distributions

Old way : data의 양이 적을 때,

data size : 1,000 or 10,000



Big data : 1%로도 검증하기에 충분한

data size : 1,000,000



Train set : 모델을 학습시키기 위해 사용하는 데이터

Dev set : 학습된 모델을 검증하기 위한 데이터

Test set : 최종 성능 평가하는데 사용하는 데이터

Dev set과 Test set의 분포는 같아야 함 → 무작위로 섞은 뒤 추출하기

When to change Dev/Test sets

[마신러닝 과정]

- 1) 모델을 평가할 적절한 척도 설정
- 2) 해당 척도를 기준으로 좋은 성능을 이끌어 내는 것

좋은 성능을 내지 못한다면 ?

- ① 평가 척도 변경
- ② Dev/Test set을 실제 application 훈련 데이터로 변경

Cat data examples

Metric : classification error

Algorithm A : 3% error → pornography를 잘못 보류

Algorithm B : 5% error

→ application에서 사용하려면 Algorithm B가 적합

[평가 척도 변경]

$$\frac{1}{M_{\text{dev}}} \sum_{i=1}^m I\{y_{\text{pred}}^{(i)} \neq y^{(i)}\}$$

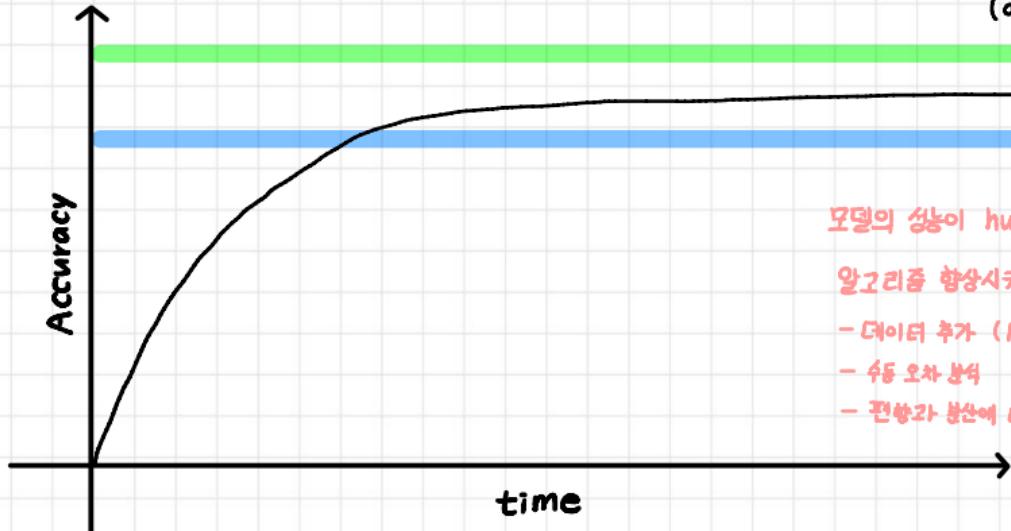
↓

$$\frac{1}{\sum w^{(i)}} \sum_{i=1}^m w^{(i)} I\{y_{\text{pred}}^{(i)} \neq y^{(i)}\}$$

→ 원치 않는 데이터에 가중치 부여

$$w^{(i)} = \begin{cases} 1 & \text{if } x^{(i)} \text{ is Not pornography} \\ 10 & \text{if } x^{(i)} \text{ is pornography} \end{cases}$$

Comparing to human-level performance



모델의 이론상 가능한 최적의 오차값
(overfitting이 아닌 경우, 이 값을 뛰어넘을 수 X)
base optimal

모델의 성능이 human-level 보다 좋지 않다면 ?

알고리즘 향상시키기

- 데이터 추가 (labeled data)
- 수동 오차 분석
- 전통과 분산에 대한 더 나은 분석

많은 머신러닝 모델들이 사람 수준의 성능은 빠르게 뛰어넘지만,
베이지안 최적 오차까지는 줄이는 데 시간이 많이 소요