

# Avoidable Bias

## # Case 1

Human(≈Base) 1%  
 ↓  
 Training error 8%  
 ↓  
 Dev error 10%

→ 모델이 충분히 훈련되지 않을

Avoidable Bias를 줄여야 함

## # Case 2

Human(≈Base) 7.5%  
 ↓  
 Training error 8%  
 ↓  
 Dev error 10%

Variance를 줄여야 함

human-level error를 Base optimal error로 추정할 때,

human-level error가 train error 보다 높지 않도록 주의해야 함

# Surpassing Human-Level Performance

구조화된 데이터 + 방대한 양

- Online advertising
- Product recommendations
- Logistics
- Loan approvals

# Reducing (avoidable) bias and variance

Human-level



Avoidable Bias

Training error



Variance

Dev error

{ Train bigger model  
Train longer/better optimization algorithms  
→ momentum, RMSprop, Adam  
NN architecture / hyperparameters search

{ More data  
Regularization  
NN architecture / hyperparameters search

# Carrying Out Error Analysis

오차를 줄이려면?

Cat classifier

Accuracy : 90%  
→ Error : 10%



오차분석!

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%	

model이 잘못 분류한 예시들을 분석하고, 그 오류를 고치기  
→ 비중이 큰 것이 가장 중요!

ex) 잘못 분류한 데이터 100개 중에서 50개가 강아지 사진  
→ 강아지 사진에 대한 오류를 줄이는 것이 효율적!

# Training and Testing on Different Distributions

Train	Dev	Test
-------	-----	------

딥러닝은 학습 데이터가 많이 필요하기 때문에,

Train set와 Dev/Test set의 분포가 달라지는 경우가 생김

→ 분포가 다른 데이터가 추가된다면 ?

① 기존의 데이터와 섞어서 다시 Train/Dev/Test로 분리하고 학습

→ 동일한 분포에서 얻어져서 다루기 쉽지만, Dev/Test에 새로운 데이터의 비중이 적어서 성능을 올리기 어려움

② 새로운 데이터의 일부를 Train set에 넣고, Dev/Test를 새로운 데이터로만 구성

→ 새로운 데이터에서 좋은 결과를 얻는 목표에 도달