



deeplearning.ai

트레이닝 셋 distribution이 dev/test set의 distribution과 다를 때는 bias and variance를 분석하는 방법을 바꿔야 한다.

Mismatched training and dev/test data

Bias and Variance with mismatched data distributions

Cat classifier example

Assume humans get $\approx 0\%$ error.

Training error 1.0%

Dev error 10.0%

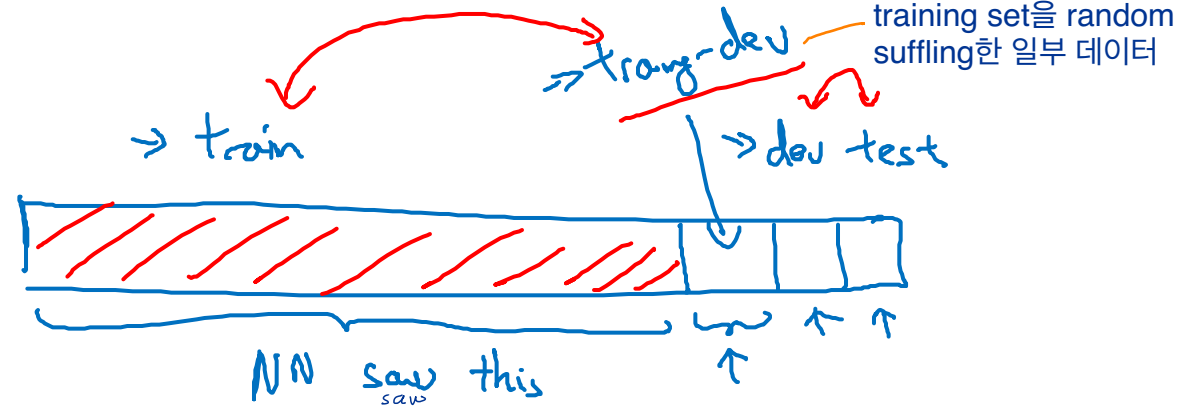
이 경우 보통은 high variance 문제로 볼 수 있다. 단, training, dev set이 같은 distribution에서 왔다면 말이다. 같은 distribution이 아니라면 단순히 variance 문제가 원인이라고만 할 수는 없다. 단순히 training set의 이미지들이 너무 쉬워서 그럴 수도 있다. 이 경우 variance problem이라기 보다는 단지 dev set의 이미지들이 더 분류하기 어려운 놈들인 것을 반영하는 것일 수도 있다.

Training-dev set: Same distribution as training set, but not used for training

training-dev와 dev 모두 NN이 명시적으로 학습에 사용하지 않았지만 dev는 다른 distribution에서 왔다. training-dev에는 잘 동작하는데 dev에는 아니라는 것은 니가 진짜 care하는 distribution 말고 다른 distribution에는 generalization의 문제 없이 잘 학습되었다는 얘기 -> data mismatch problem이라는 얘기

training set은 다른 distribution에서

dev/test는 같은 distribution에서



training set과 same distribution이지만 NN이 보지 못했던 training-dev set에서 정확도가 낮게 나왔으니 generalization이 문제라는 것을 알 수 있으므로 variance problem 때문인 것을 알 수



Bias/variance on mismatched training and dev/test sets

Key Quantities

Human level
Training set error
Training-dev set error
→ Dev error
→ Test error

4%
7%
10%
12%
12%

↑ avoidable bias
↑ variance
↑ data mismatch
↑ degree of overfitting to dev set.

만약 dev error랑 test error가 차이가 많이 나면 dev set에 overfitting 된거고 dev set을 늘려주는 수 밖에 없지 뭐

4%
7%
10%
6%
6%

these two were evaluated on training set distribution
these two were evaluated on dev/test set distribution

training dataset이 더 어려운 데이터일 때 이런 현상이 발생할 수도 있음.

More general formulation

Rearview mirror

speech activation rearview mirror 예제를 다시 생각해보자.

	General speech recognition	Rearview mirror speech data.	
Human level	"Human level" 4%	6%	↑ avoidable bias
Error on examples trained on	"Training error" 7%	6%	
Error on examples <u>not</u> trained on	"Training-dev error" 10%	"Dev/Test error" 6%	↑ Variance

주로 빨간 영역에서 유용한 정보를 얻을 수 있지만 때로는 나머지 영역도 채우는데서 추가적인 인사이트를 얻을 수도 있다. 예를 들어 휴먼 레벨 행에서 4%, 6%로 차이가 나는 것으로 봤을 때 rear view mirror speech data가 더 어렵다는 것을 알 수 있다.

data mismatch

앞에 두 개는 많이 다뤘는데 data mismatch 문제는 체계적인 해결책이 없는게 사실이다. 하지만 몇 가지 도움이 될만한 방법들이 있다. 다음 강의에서 알아보자.