Week 2-1 Error Analysis, Incorrectly Labeled Examples, Data Mismatch (Training Set vs. Dev/Test Set)

笔记本: DL 3 - Structuring ML Projects
创建时间: 2021/1/11 13:47 **更新时间:** 2021/1/11 14:16

ceiling

Look at dev examples to evaluate ideas

90% accuracy
→ 10% error

Should you try to make your cat classifier do better on dogs?

Error analysis:
- Get ~100 mislabeled dev set examples.
- Count up how many are dogs.

5-10 min

"ceiling"

5% → 5/100

10% → 9.5%

50% → 50/100

10% → 5%

Andrew Ng

Evaluate multiple ideas in parallel

Ideas for cat detection:
- Fix pictures of dogs being recognized as cats
- Fix great cats (lions, panthers, etc..) being misrecognized
- Improve performance on blurry images

Image	Dog	Great Cats	Blurry	Instagram	Comments
1	✓			✓	Pitbull
2			✓	✓	
3		✓	✓		Rainy day at zoo
⋮	⋮	⋮	⋮		
% of total	8%	43%	61%	12%	

Andrew Ng

Incorrectly labeled example

Incorrectly labeled examples

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

Systematic errors

Error analysis

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%	

Overall dev set error	10%	2%
Errors due incorrect labels	0.6%	0.6%
Errors due to other causes	9.4%	1.4%
		2.1% 1.9%

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

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
- Consider examining examples your algorithm got right as well as ones it got wrong. (98.1) (2%)
- Train and dev/test data may now come from slightly different distributions.

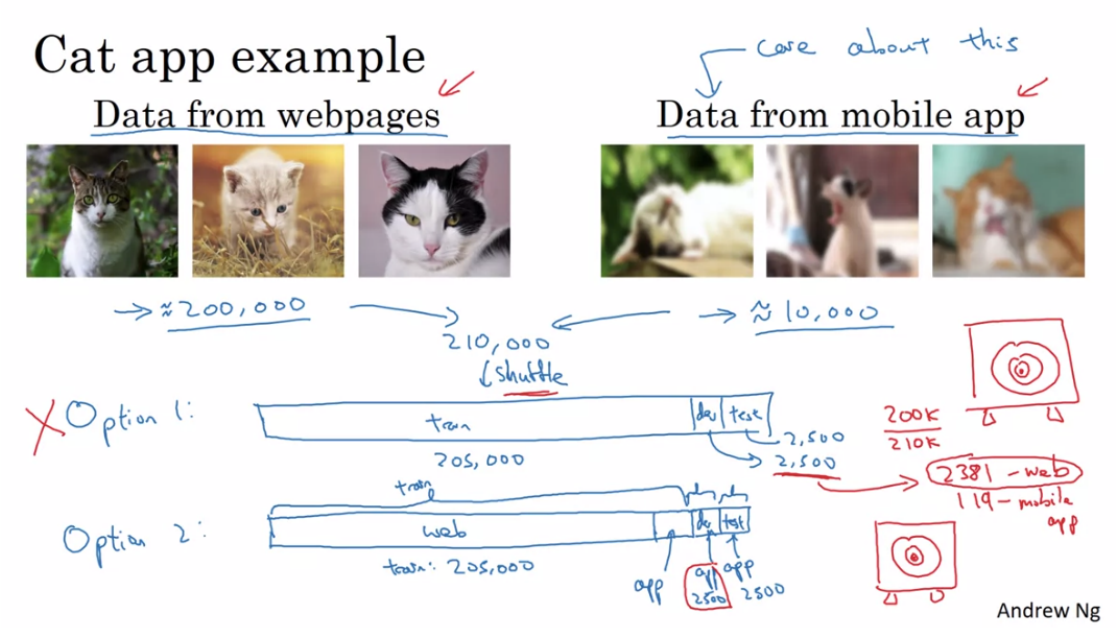
it is okay that train set is from different distribution

Build your first system quickly, then iterate

Speech recognition example

- Noisy background
  - Café noise
  - Car noise
- Accented speech
- Far from microphone
- Young children's speech
- Stuttering uh, ah, um, ...
- ...

- Set up dev/test set and metric
- Build initial system quickly
- Use Bias/Variance analysis & Error analysis to prioritize next steps.



Estimating the **bias and variance** of your learning algorithm really helps you prioritize what to work on next. But the way you analyze **bias and variance changes** when your training set comes from a different distribution than your dev and test sets. Let's see how.

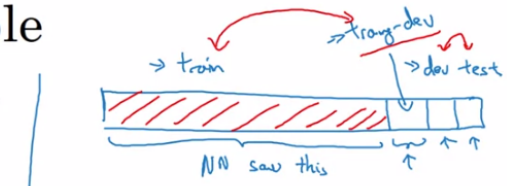
train-dev set

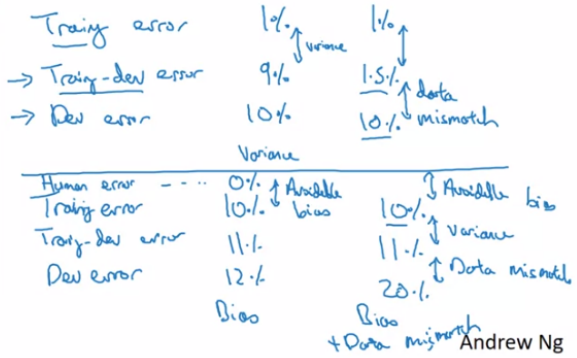
Cat classifier example

Assume humans get ≈ 0% error.

Training error 1%
Dev error 10%

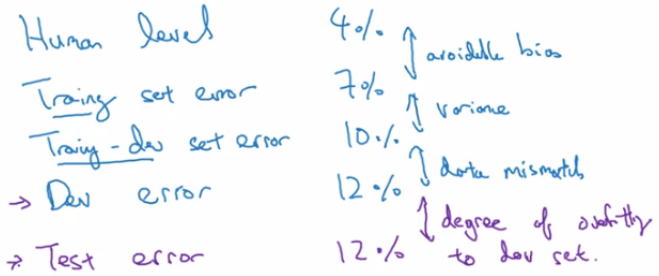
Training-dev set: Same distribution as training set, but not used for training



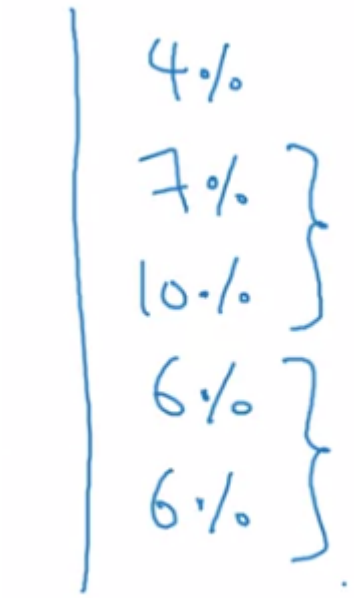


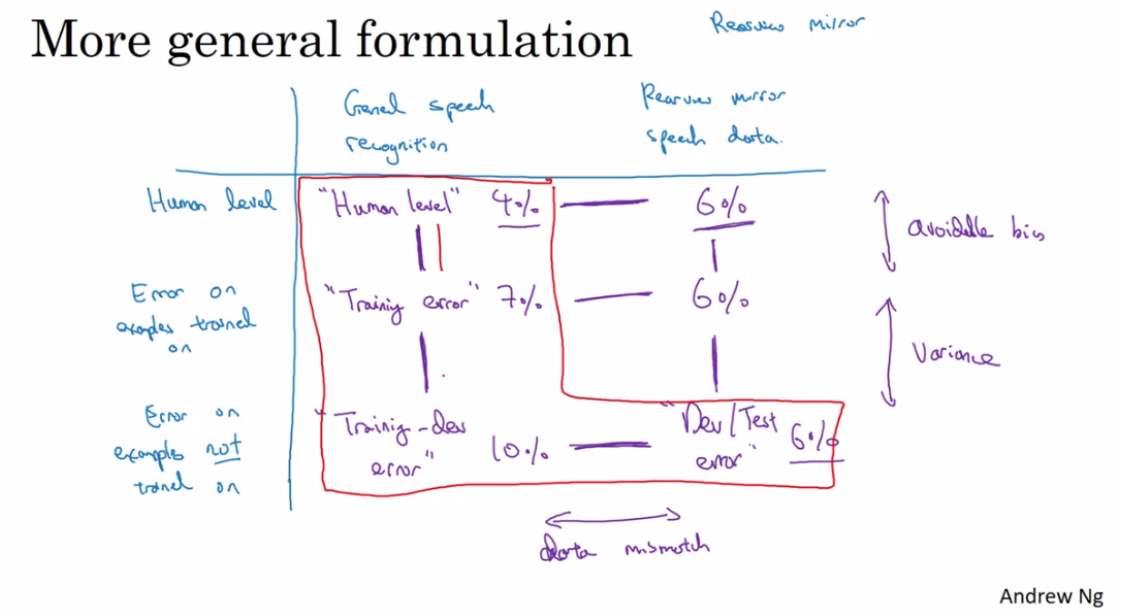
Andrew Ng

Bias/variance on mismatched training and dev/test sets



But what-if the trainning set is more difficult than the dev/test set?





Andrew Ng

Compare human performance to get a perception is how hard are the 2 sets

How to address data mismatch?

Addressing data mismatch

- Carry out manual error analysis to try to understand difference between training and dev/test sets

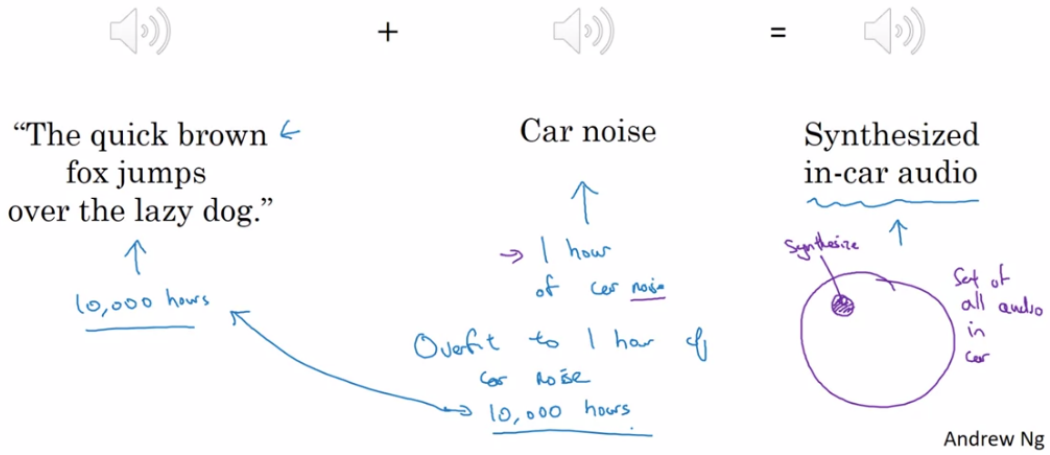
  E.g. noisy – car noise street numbers

- Make training data more similar; or collect more data similar to dev/test sets

  E.g. Simulate noisy in-car data

(Artificial Data Synthesis)

Artificial data synthesis



risk: overfit to the 1 hour car noise (synthesized data)