

Carry out error analysis:

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 ←

feel free to add checking item if you need

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

Perform Error Analysis on dev set

Andrew Ng

Clean up wrong labeled data:

Incorrectly labeled examples

x

y

1 0 1 1 0 1 1

Training set.

The sixth image (white dog) and its label '1' are circled in blue, with an arrow pointing to the label.

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

it's ok

Systematic errors

if a white is always labeled as cat, then it has problem, it is systematic error.

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.
3 things need to look at	% of total	8%	43%	61%	6%	
Overall	dev set error			10%		2%
Overall	Errors due incorrect labels			0.6%	0.6%	0.6%
Overall	Errors due to other causes			9.4%		1.4%

just small percent don't waste time
 large percent, need to fix incorrect labels

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

Andrew

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. 2%
 - Train and dev/test data may now come from slightly different distributions. ! but train does not have to from exactly same distribution as dev/test
- dev / test must be same distribution

Build a system first, and then iterate it:

Speech recognition example

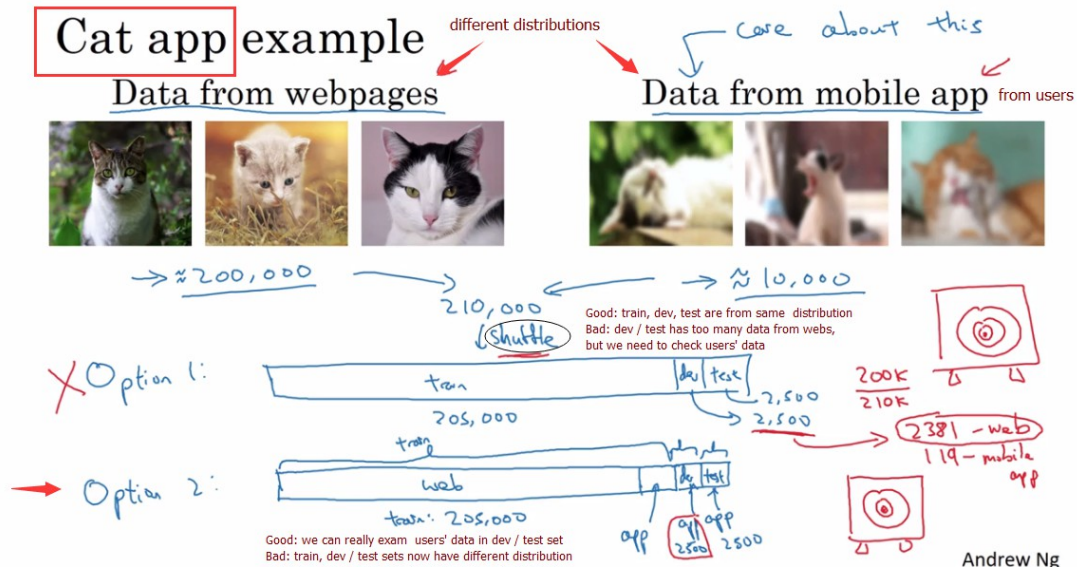


- • Noisy background
 - • Café noise
 - • Car noise
- • Accented speech
- • Far from microphone
- • Young children's speech
- • Stuttering *uh, ah, um, ...*
- • ...

- • Set up dev/test set and metric
- • Build initial system quickly don't make it too complex at first
- • Use Bias/Variance analysis & Error analysis to prioritize next steps.

Andrew Ng

Mismatched Train data and dev / test data:



Speech recognition example

Speech activated rearview mirror

speech activated rearview mirror

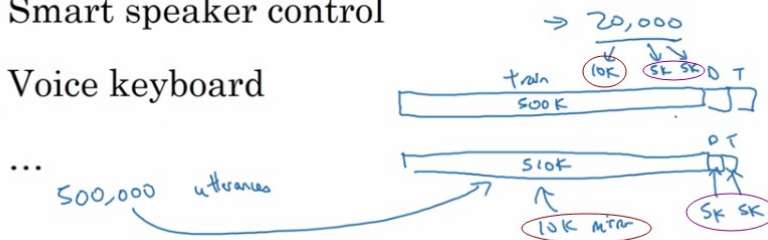


Training

Purchased data x, y
Smart speaker control
Voice keyboard

Dev/test

Speech activated rearview mirror



Andrew Ng

Bias and Variance with Mismatched Data:

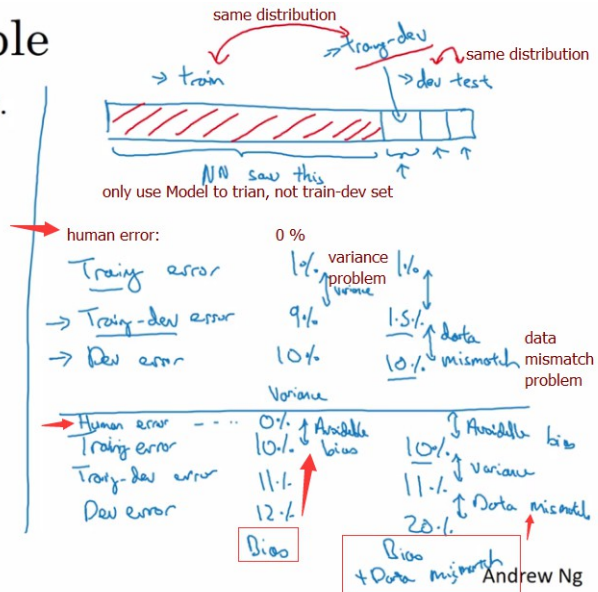
Since now train set has different distribution, we can't just simply use bias and variance to evaluate. We also need to identify the data mismatch problem.

Cat classifier example

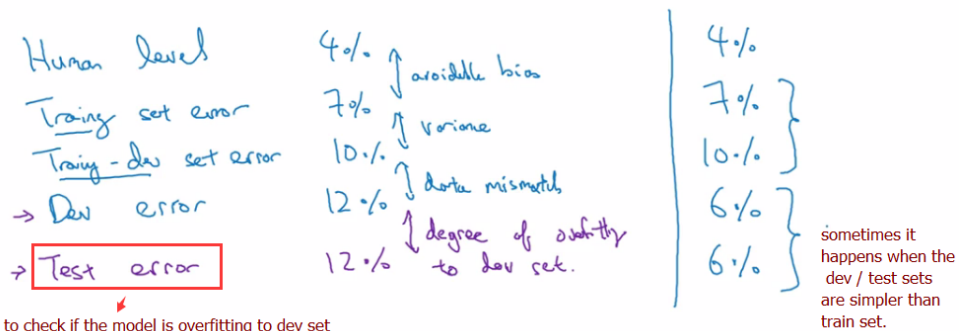
Assume humans get $\approx 0\%$ error.

Training error 1%
 Dev error 10%

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

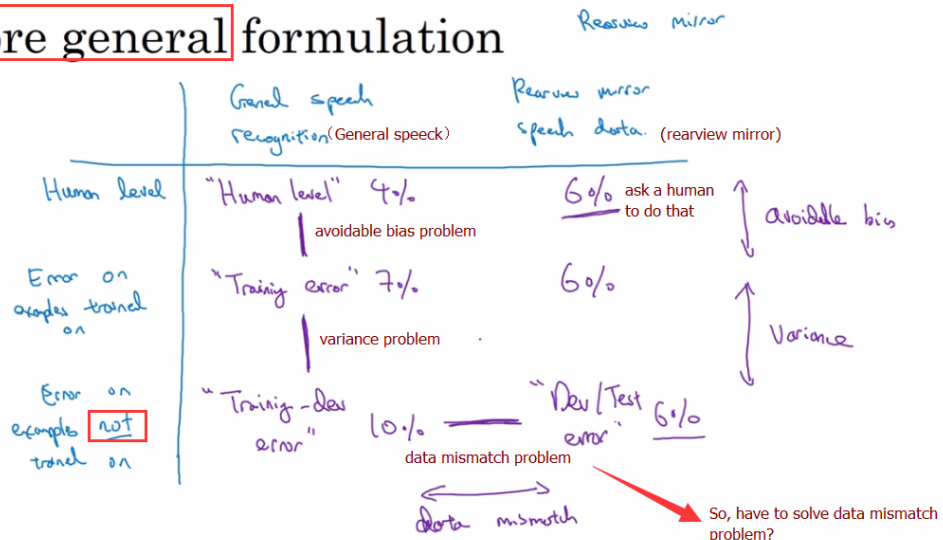


Bias/variance on mismatched training and dev/test sets



use test set to check if the model is overfitting to dev set

More general formulation



How to solve data mismatch problem:

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

