deeplearning.ai

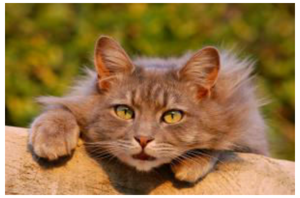
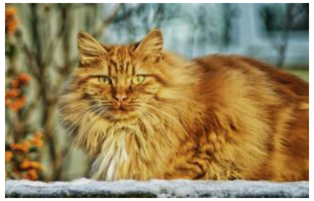
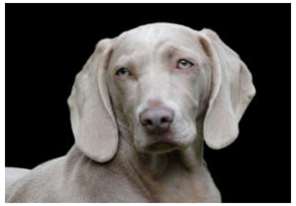
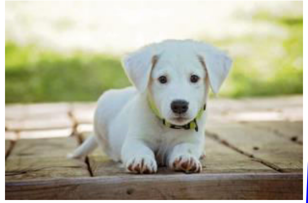
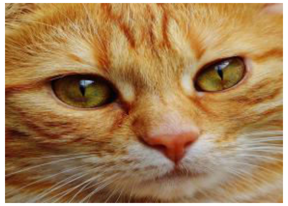
Error Analysis

Cleaning up Incorrectly labeled data

Incorrectly labeled examples

Mislabelled example

data could be in train, dev, Test set

x							
y	1	0	1	1	0	1	1

$y = 1$ = cat

$y = 0$ = Non cat

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

↳ As long as these errors are small in # (less # of misclassified examples) & They occur randomly we are fine

DL Algos are less robust to systematic errors

eg All white dog Images are classified as cats, that is a problem!

Andrew Ng

Error analysis

If you're worried, incorrectly labelled examples may occur in dev/test set, Add a column

Error Analysis on dev/test set is a good exercise as it helps you decide where to go next

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%~~ 51	6%	

Comments regarding misclassification or incorrect labelling can help you identify trends

→ Not that big a deal, no need to Invest time going through dev set & fixing mislabels

Is it worth going back to fix mislabels

Find these

Overall dev set error → 10%

Errors due incorrect labels → 6% of 10% = 0.6%

Errors due to other causes → 10 − 0.6 = 9.4%

Focus on these

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

Say overall dev set error = 2.1% for A, 1.9% for B

& Error (Incorrect label) = 0.6%

Now, $0.6 = x \cdot 2.1\%$

$\Rightarrow x$ = # Incorrect labels = 30%

∴ Needs correct labelling to decide b/w 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 (If you clean up dev set, must clean up test set)

- Consider examining examples your algorithm got right as well as ones it got wrong. (All white dogs labelled as cat, can give some Insight)

- Train and dev/test data may now come from slightly different distributions. (You only cleaned dev, test, Not train set)

Why? large data set! ← Its actually okay for train dist to be diff ← Not feasible to clean Training set (too large)