Setting up your
ML application

Train/dev/test
sets

deeplearning.ai

Applied ML is a highly iterative process

layers
hidden units
learning rates
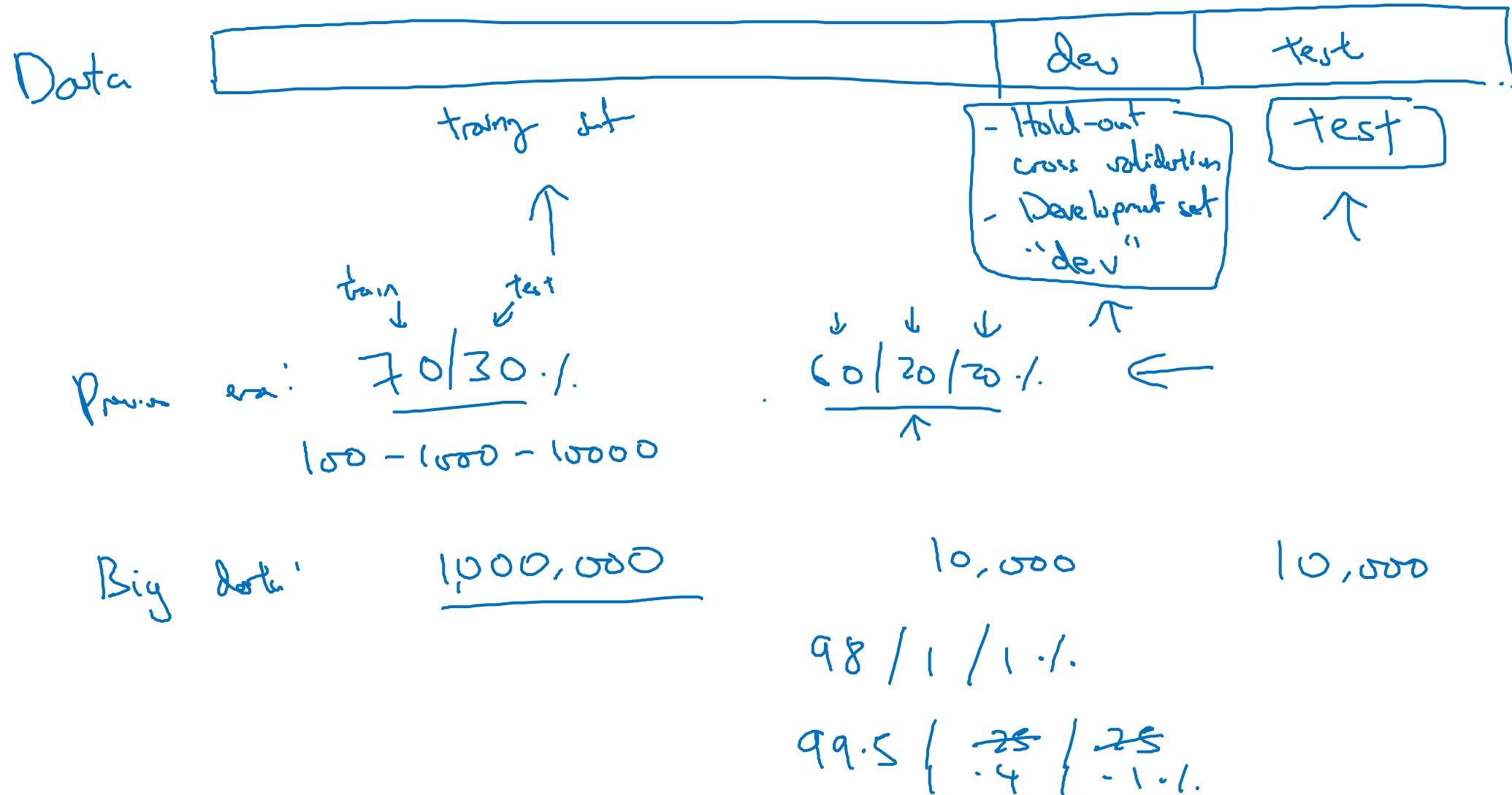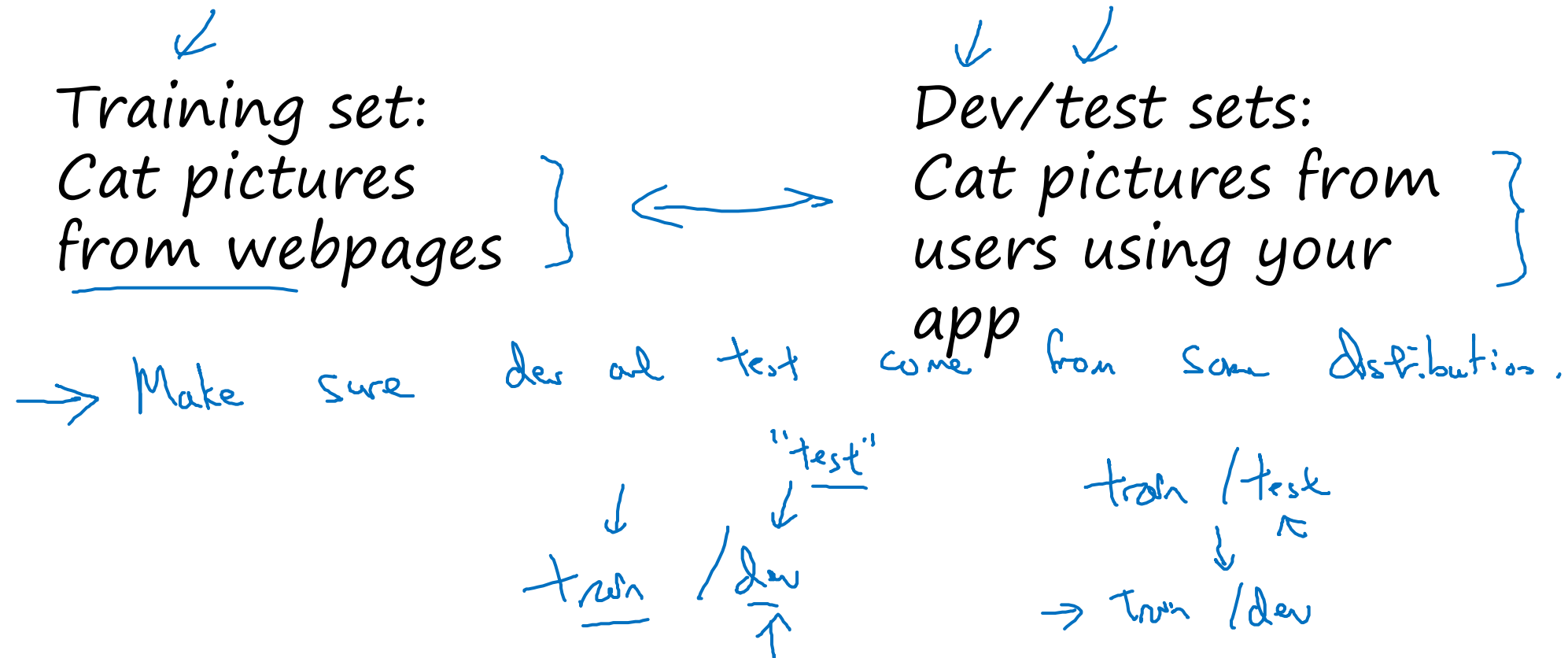
activation functions
...

Idea

Code

Experiment
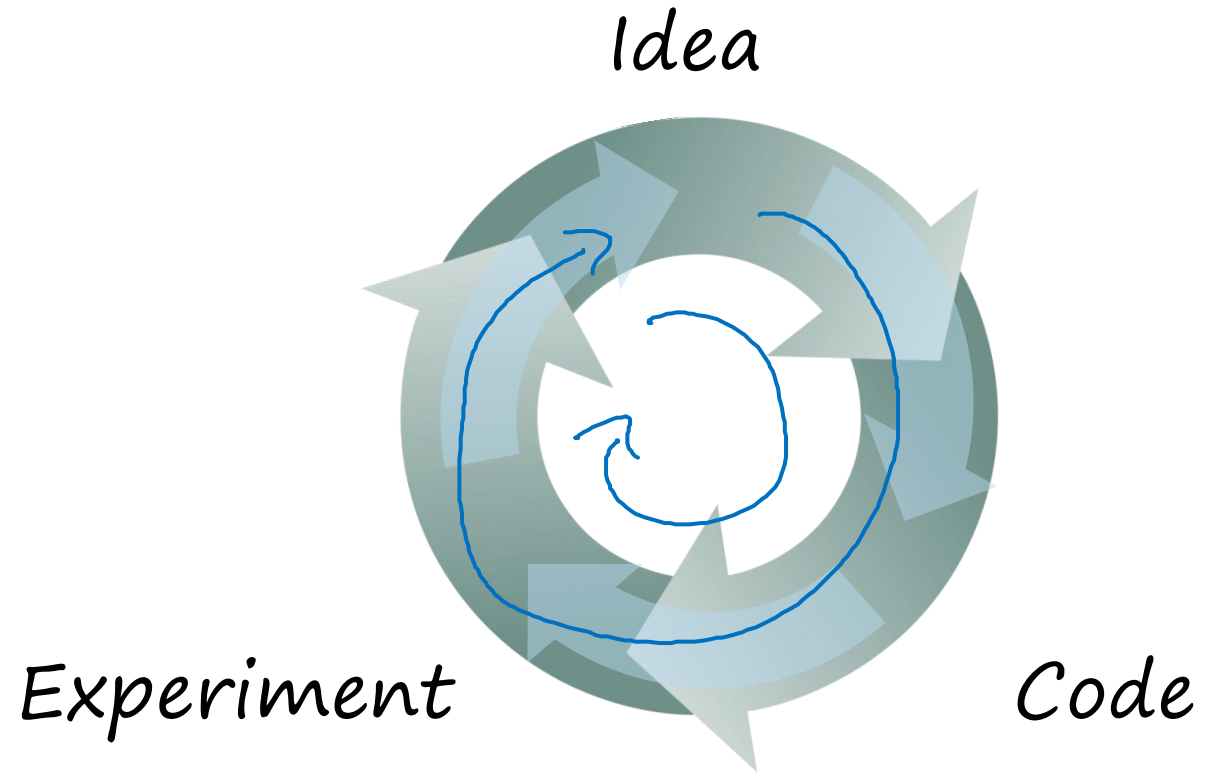
NLP, Vision, Speech, Structured data

Ads   Search   Security   logistic ....

Andrew Ng

Train/dev/test sets

Data

training set

dev

test

- Hold-out cross validation
- Development set "dev"

test

train ← → test

Previous era: 70/30%

60/20/20%

100 - 1000 - 10000

Big data: 1000,000   10,000   10,000

98 / 1 / 1%

99.5 / .25 / .25
       .4 / .1%

Andrew Ng

Mismatched train/test distribution _Cats_

↓

Training set:
Cat pictures
from webpages }

↔

↓ ↓

Dev/test sets:
Cat pictures from
users using your
app }

→ Make sure dev and test come from same distribution.

↓   "test"
         ↓ ─

train / dev
        ↑

train / test
      ↓ ↖
   → Train / dev

Not having a test set might be okay. (Only dev set.)

Andrew Ng