deeplearning.ai

Setting up your ML application

Train/dev/test sets

Applied ML is a highly iterative process

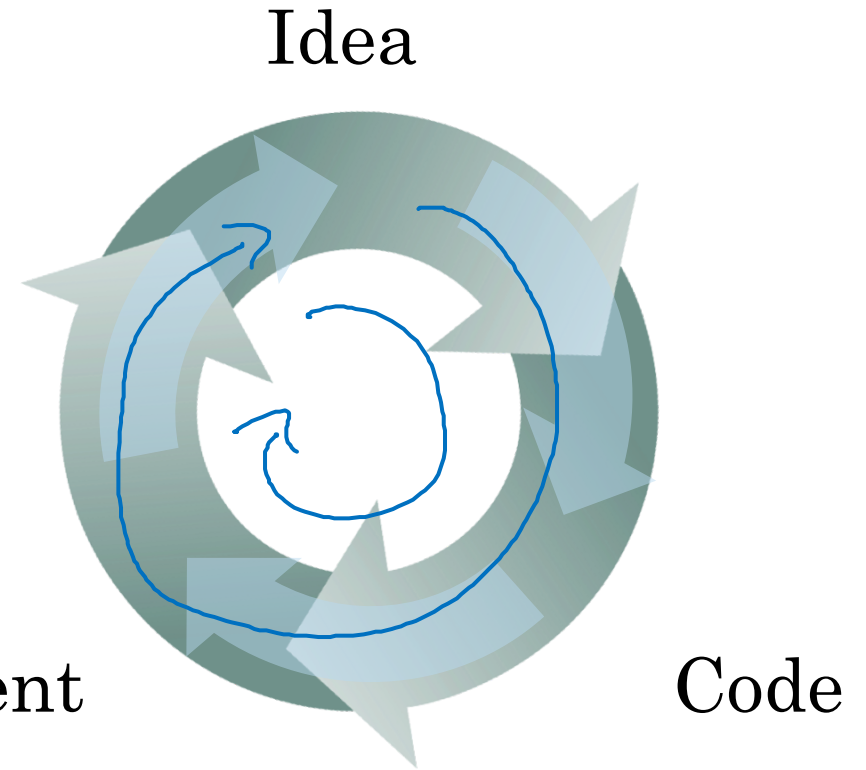# layers

hidden units

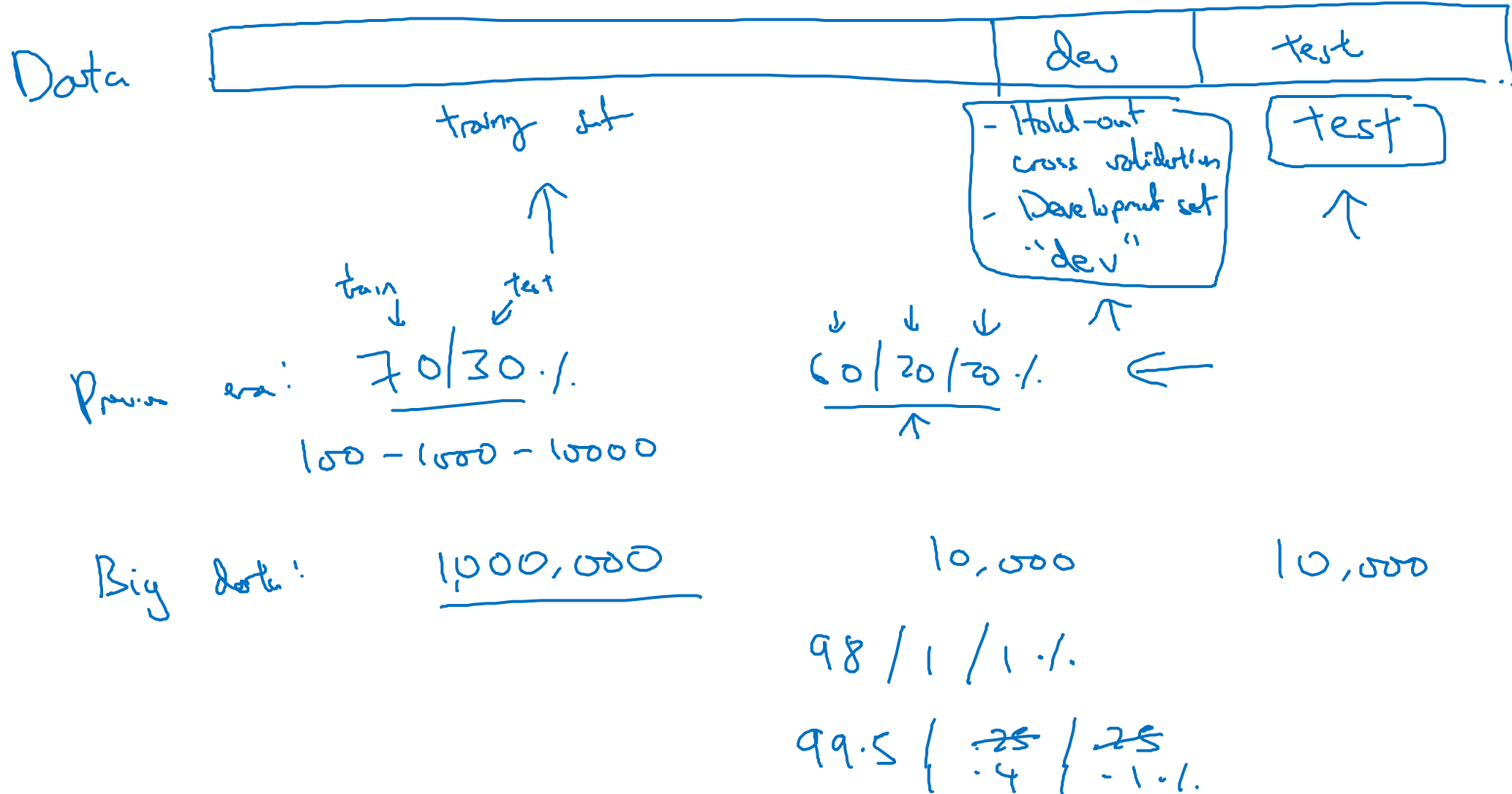learning rates

activation functions

...

Idea

Experiment

Code

NLP, Vision, Speech, Structured data

Ads   Search   Security   logistic ....

Andrew Ng

Train/dev/test sets

Data

training set

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

test

train ↓   test ↓

Previous era: 70/30 %

100 - 1000 - 10000

60/20/20 %

Big data: 1,000,000   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.)