Mismatched training and dev/test data

deeplearning.ai

Training and testing on different distributions

Cat app example

## Data from webpages


Data from mobile app

Care about this

$\approx 200,000 \rightarrow$

$210,000$

$\approx 10,000 \leftarrow$

shuffle

Option 1:

| train | dev | test |

$205,000$

$2,500$
$2,500$

$\frac{200K}{210K}$

$2381 - web$
$119 - mobile\ app$

Option 2:

train

web

train: $205,000$

dev test

app
app 2500
app 2500

all mobile app images

Andrew Ng
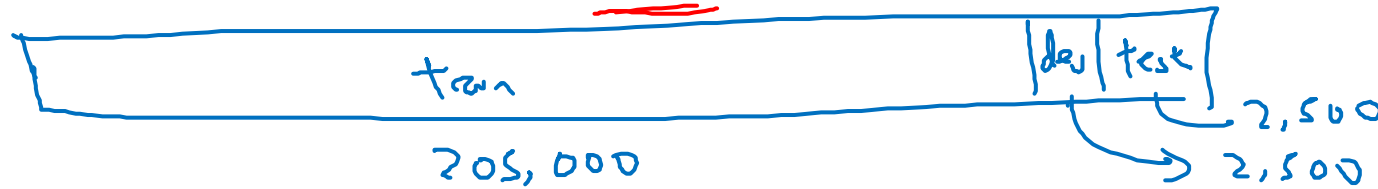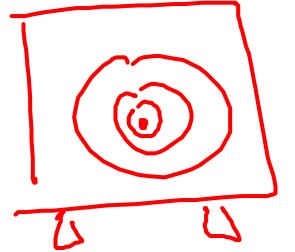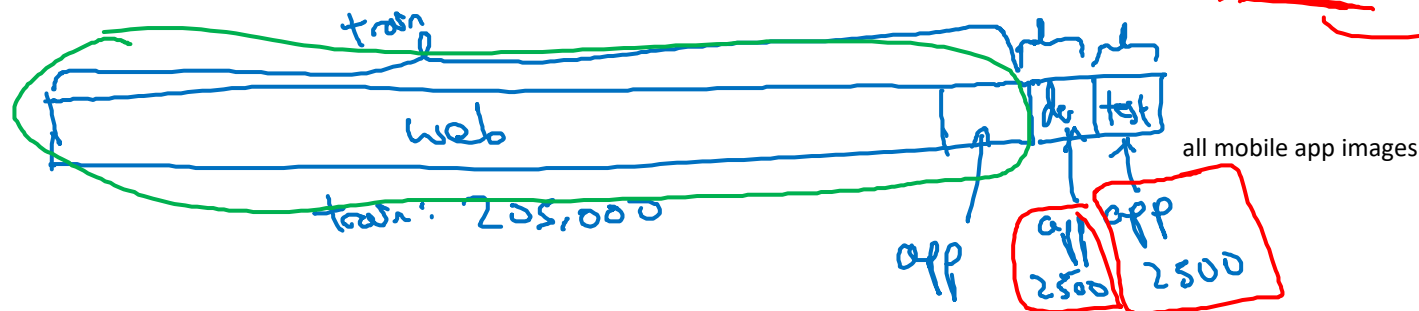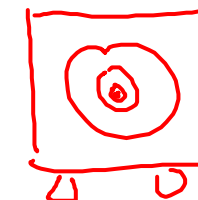
Speech recognition example

Speech activated rearview mirror

Training

Purchased data x, y

Smart speaker control

Voice keyboard

...

500,000 utterances

Dev/test

Speech activated rearview mirror

20,000

10k 5k 5k D T

train
500 k

510 k   D T

10k mirror   5k 5k

Andrew Ng