



deeplearning.ai

Mismatched training
and dev/test data

Training and testing
on different
distributions

Cat app example

Data from webpages



care about this
Data from mobile app

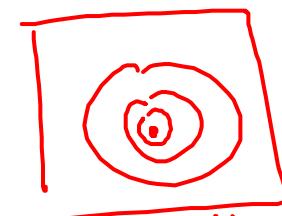


$\rightarrow \approx 200,000$

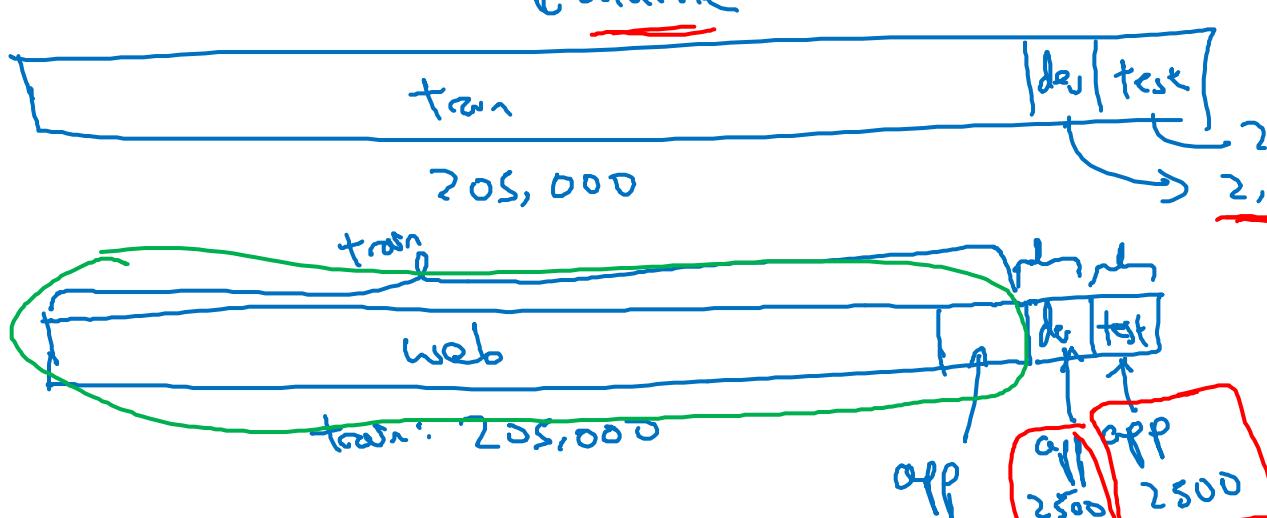
$210,000$
shuffle

$\rightarrow \approx 10,000$

X Option 1:



Option 2:



Speech recognition example

Speech activated rearview mirror



Training

Purchased data $\downarrow \downarrow$
 x, y

Smart speaker control

Voice keyboard

...

500,000 utterances

Dev/test

Speech activated
rearview mirror

$\Rightarrow 20,000$

10K SK SK D T
500 K

train
500 K

510K

10K mirr

OT

SK SK