Mismatched training and dev/test data

Training and testing on different distribution

Cat app example

Data from webpages
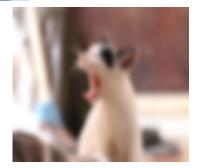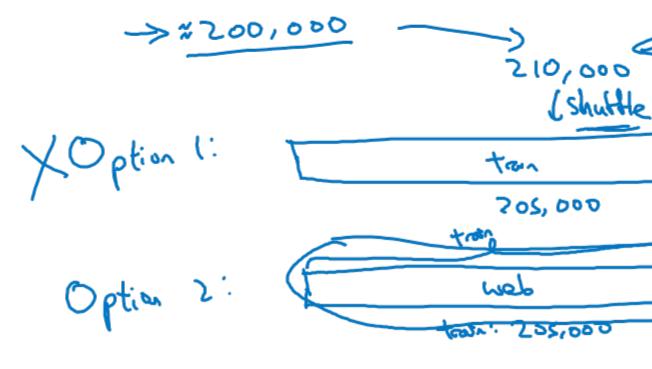
Data from mobile app

Care about this

≈200,000 → 210,000 ← ≈10,000

(shuttle

X O Option 1:

train ... dev | test

205,000

2,500
2,500

200K / 210K

2381 – web
119 – mobile app

Option 2:

web ... dev test

train: 205,000

app

app 2500 app 2800

Andrew Ng
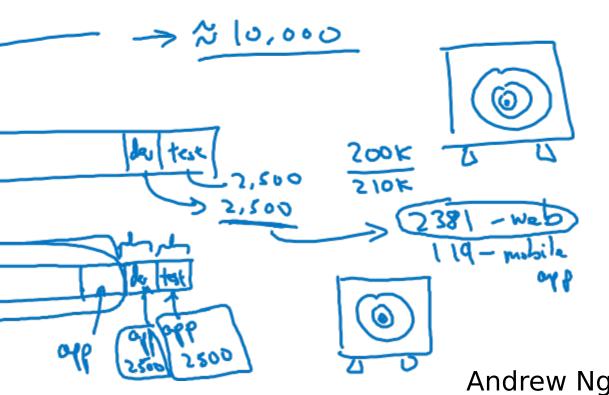
Speech recognition example

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

train — 10K 5K 5K D T

500K

510K

10K mirr

5K 5K