deeplearning.ai

Mismatched training and dev/test data

Training and testing on different distributions

Cat app example

Data from webpages

Data from mobile app

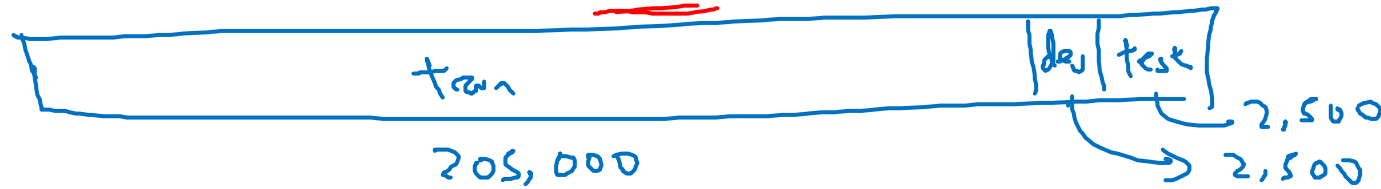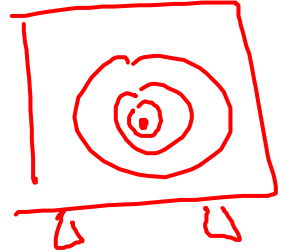Care about this

≈ 200,000 → ← ≈ 10,000

210,000

(Shuffle

Option 1:

| train | dev | test |

205,000

2,500
2,500

200K / 210K

2381 — web
119 — mobile app

Option 2:

| train |
| web |

train: 205,000

app

app 2500   app 2500

Andrew Ng
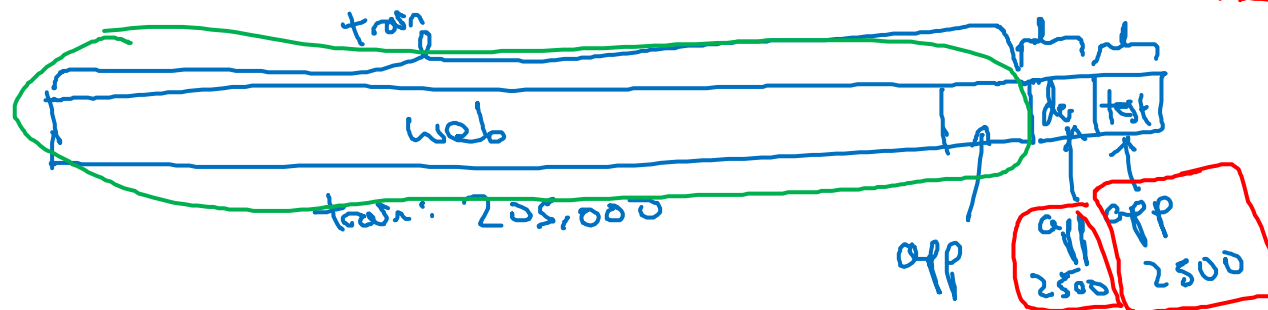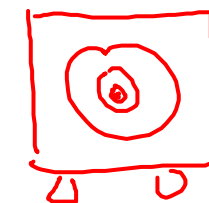
Speech recognition example

Speech *activated* rearview mirror
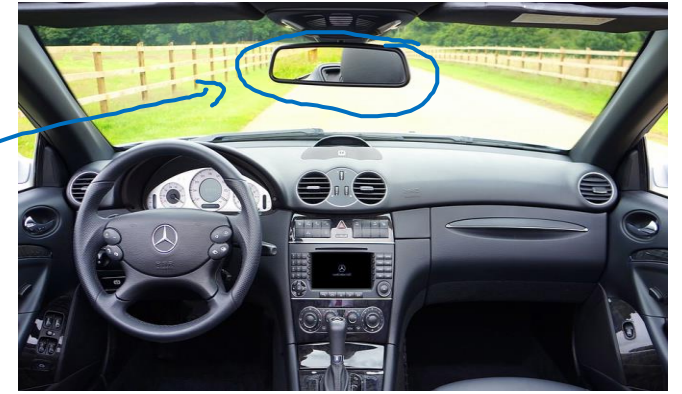
Training

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

510K   D T

10K mirror   5K 5K

Andrew Ng