deeplearning.ai

Mismatched training and dev/test data

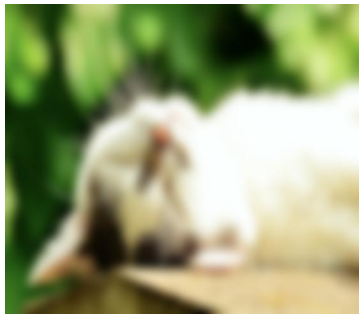
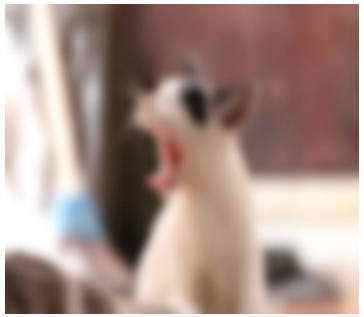
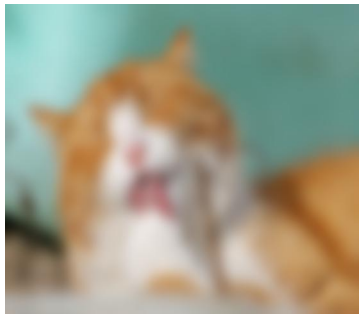
Training and testing on different distributions

Cat app example

Data from webpages

Data from mobile app

care about this

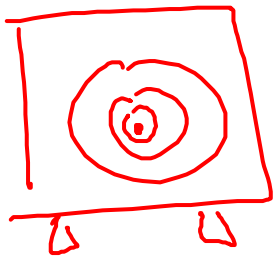




→ ≈200,000 → 210,000 ← → ≈10,000

(shuffle

Option 1:

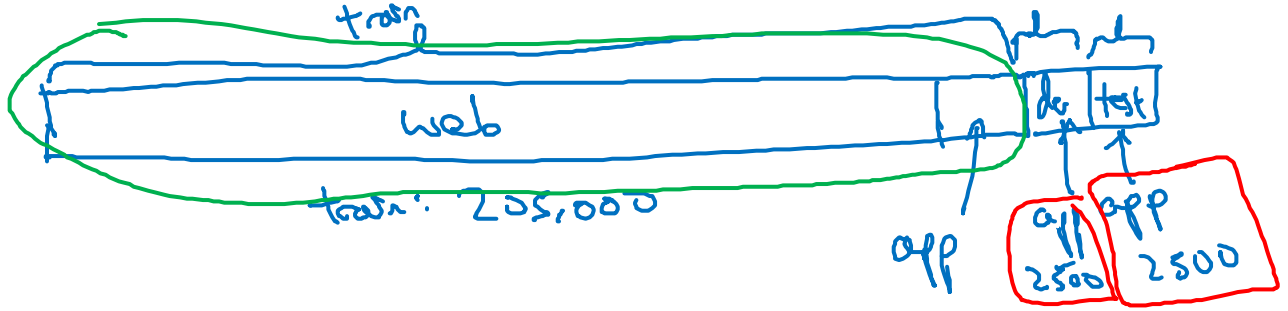
200K / 210K

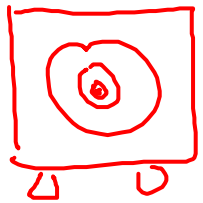


2381 - web

119 - mobile app

Option 2:





Speech recognition example

Speech activated rearview mirror

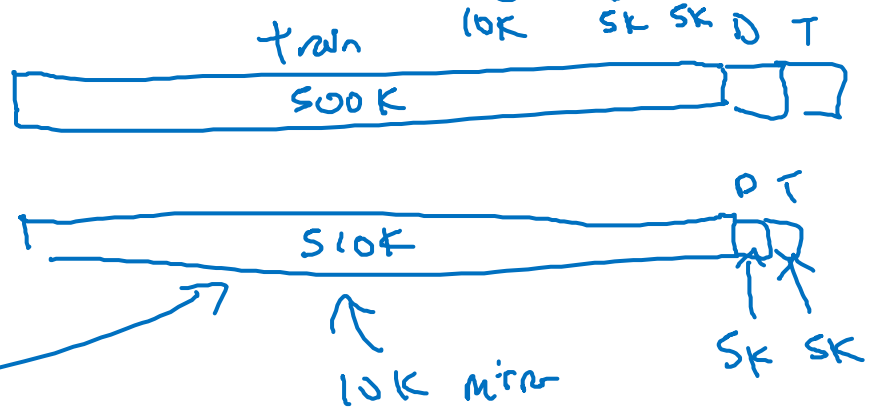
Training

- Purchased data x, y
- Smart speaker control
- Voice keyboard
- ...

500,000 utterances

Dev/test

Speech activated rearview mirror

→ 20,000

10K 5K 5K