



deeplearning.ai

Mismatched training  
and dev/test data

---

Training and testing  
on different  
distributions

# Cat app example

Data from webpages



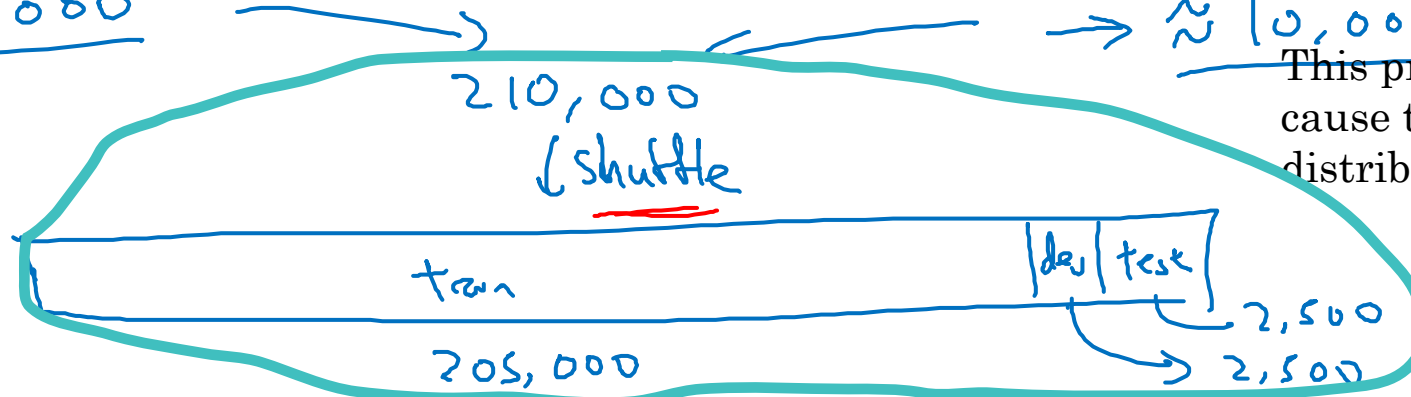
care about this  
Data from mobile app



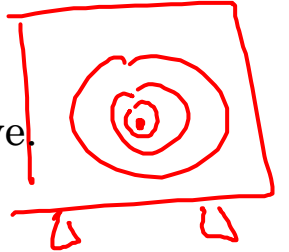
→  $\approx 200,000$

→  $\approx 10,000$

~~Option 1:~~

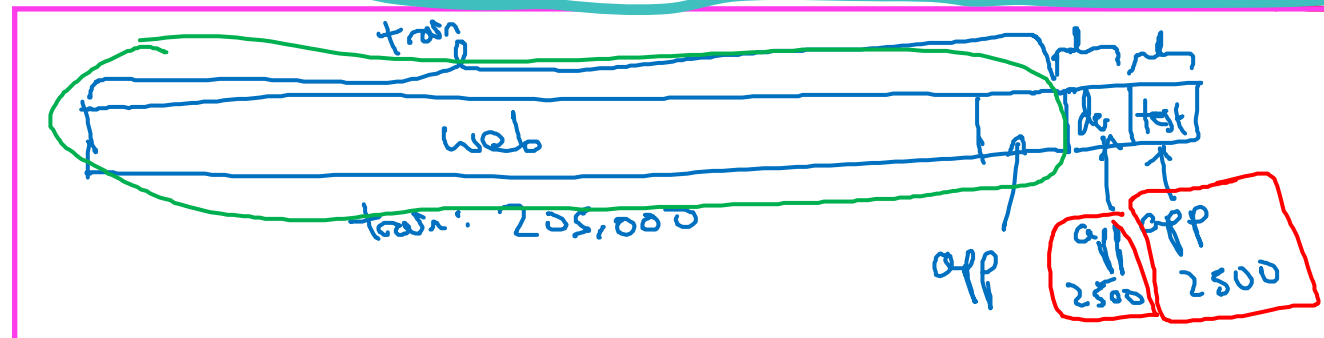


This process will cause the target/distribution to move.

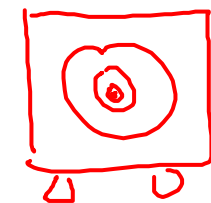


$\frac{200K}{210K}$

Option 2:



2381 - web  
119 - mobile app



# 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

