



deeplearning.ai

Introduction to ML strategy

Why ML Strategy?

Motivating example



96%.

Ideas:

- Collect more data ←
- Collect more diverse training set
- Train algorithm longer with gradient descent
- Try Adam instead of gradient descent
- Try bigger network
- Try smaller network
- Try dropout
- Add L_2 regularization
- Network architecture
 - Activation functions
 - # hidden units
 - ...

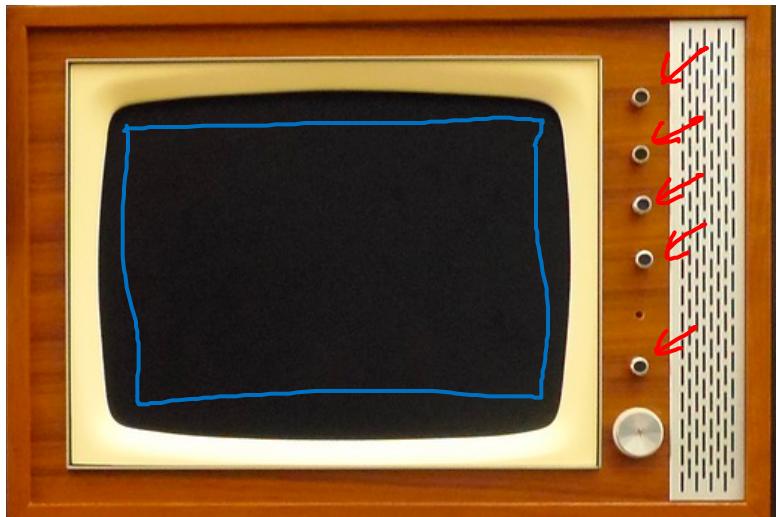


deeplearning.ai

Introduction to ML strategy

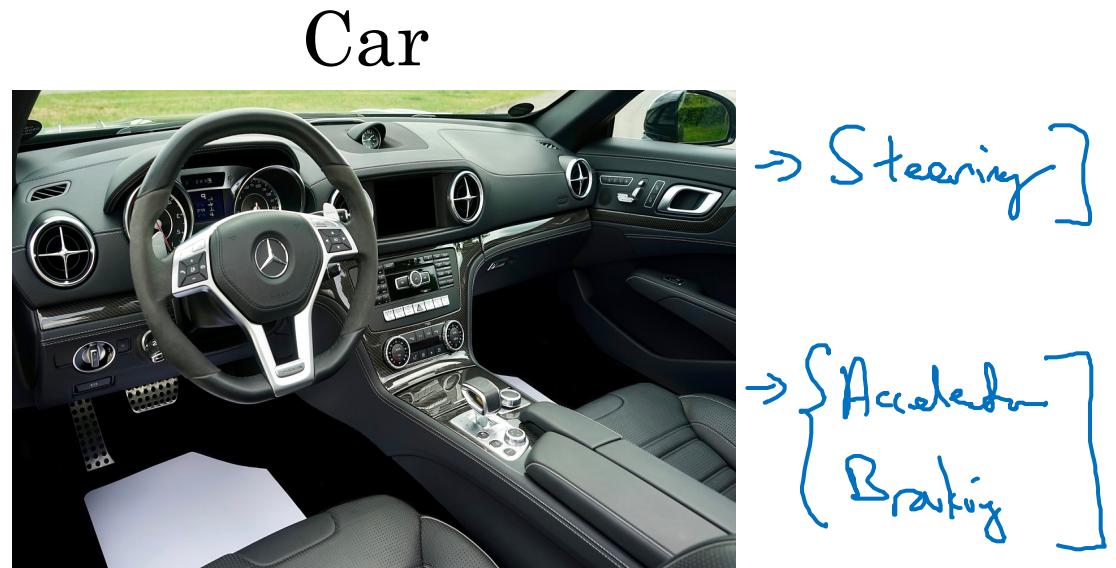
Orthogonalization

TV tuning example



Orthogonalization

$$\begin{aligned}
 & 0.1 \times \begin{array}{c} \uparrow \\ \downarrow \end{array} \\
 + & 0.3 \times \begin{array}{c} \leftarrow \\ \rightarrow \end{array} \\
 - & 1.7 \times \begin{array}{c} \diagdown \\ \diagup \end{array} \\
 + & 0.8 \times \begin{array}{c} \leftarrow \\ \rightarrow \end{array} \\
 + \dots & \begin{array}{c} \diagdown \\ \diagup \end{array}
 \end{aligned}$$



$$\rightarrow \underline{0.3 \times \text{angle}} - 0.8 \times \text{speed}$$

$$\rightarrow 2 \times \text{angle} + 0.9 \times \text{speed}.$$

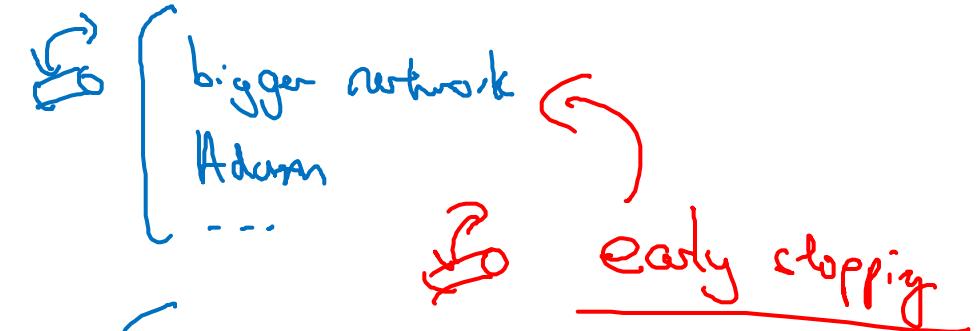


Chain of assumptions in ML

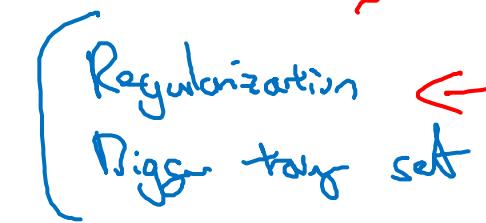
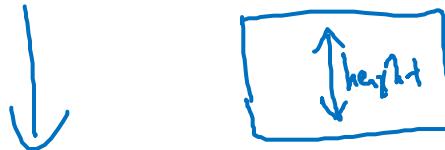
→ Fit training set well on cost function



(≈ human-level performance)



→ Fit dev set well on cost function



→ Fit test set well on cost function



Bigger dev set

→ Performs well in real world

(Happy cat pic off users.)

Change dev set or
cost function

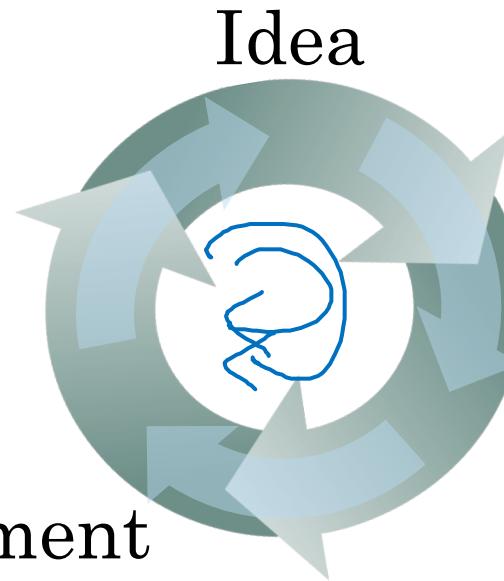


deeplearning.ai

Setting up
your goal

Single number
evaluation metric

Using a single number evaluation metric



Code

→ Of examples recognized as cont, what % actually are cont?

→ what % of actual cont are correctly recognized

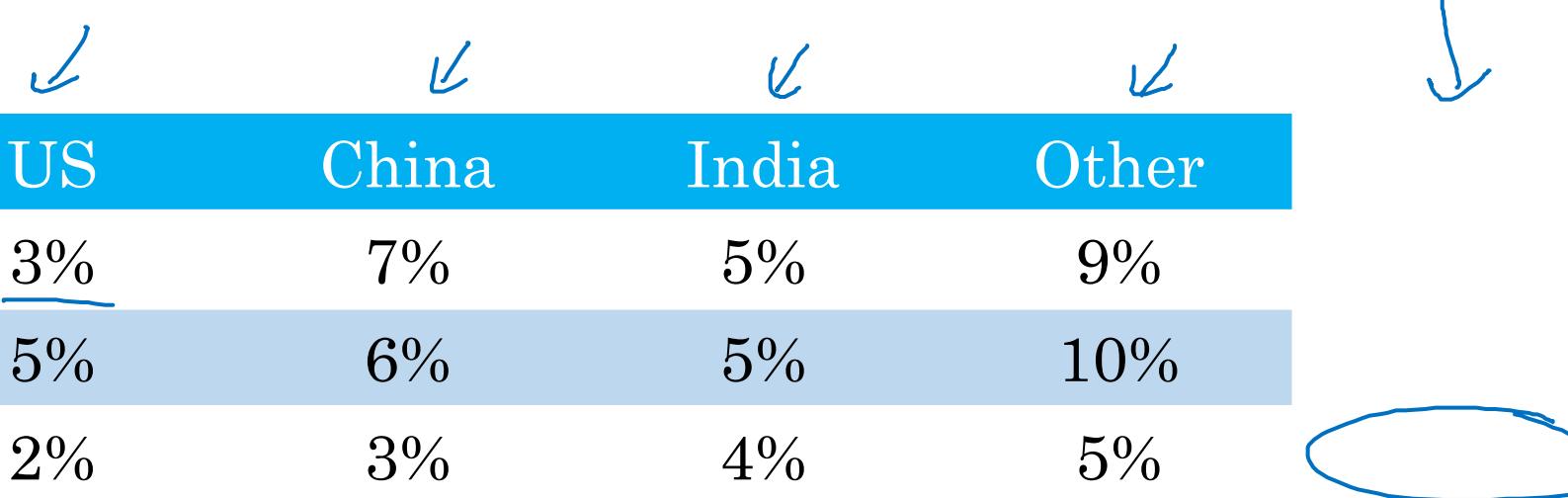
Classifier	Precision	Recall
A	<u>95%</u>	90%
B	98%	85%

F₁ Score = "Average" of P and R.

$$\left(\underbrace{\frac{2}{\frac{1}{P} + \frac{1}{R}}}_{\text{Harmonic mean}} \right)$$

Dev set + Single number evaluation metric
real Speel up iterating

Another example



Algorithm	US	China	India	Other
A	<u>3%</u>	7%	5%	9%
B	5%	6%	5%	10%
C	2%	3%	4%	5%
D	5%	8%	7%	2%
E	4%	5%	2%	4%
F	7%	11%	8%	12%



deeplearning.ai

Setting up
your goal

Satisficing and
optimizing metrics

Another cat classification example

Classifier	Accuracy	Running time
A	90%	80ms
B	92%	95ms
C	95%	1,500ms

$$\text{Cost} = \underline{\text{accuracy}} - 0.5 \times \underline{\text{running Time}}$$

Maximize accuracy

Subject to running Time \leq 100 ms.

N metrics : 1 optimizing
N-1 satisfying

optimizing



satisficing

Wakewords / Trigger words

Alexa, OK Google,

Hey Siri, nihao baidu

你好 百度

accuracy.

#false positive

Maximize accuracy.

s.t. ≤ 1 false positive
every 24 hours.



deeplearning.ai

Setting up
your goal

Train/dev/test
distributions

Cat classification dev/test sets

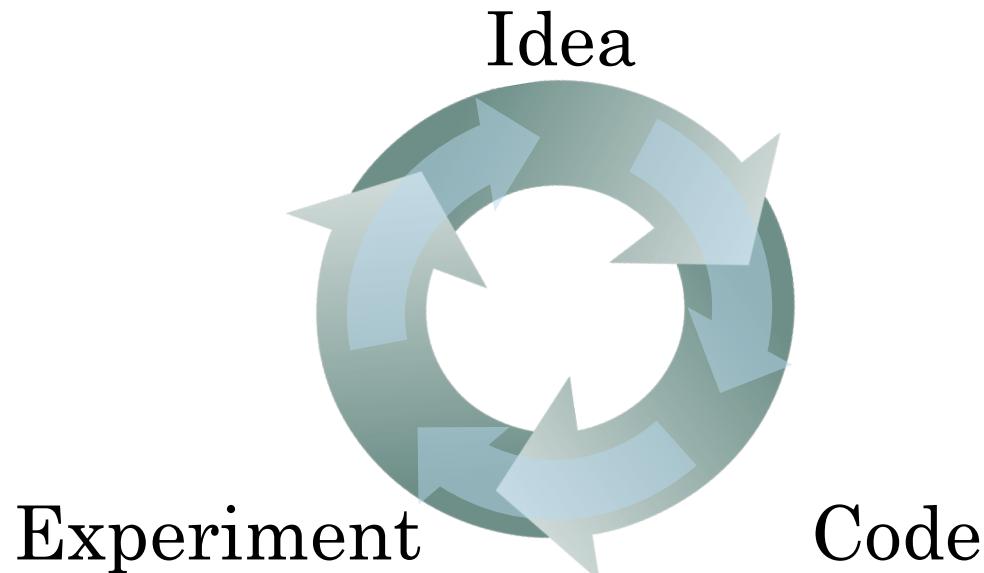
↳ development set, hold out cross validation set

Regions:

- US
- UK
- Other Europe
- South America
- India
- China
- Other Asia
- Australia



Randomly shuffle into dev/test



True story (details changed)

[Optimizing on dev set on loan approvals for
medium income zip codes

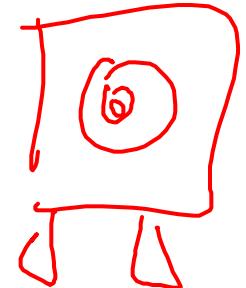


$$x \rightarrow y \text{ (repay loan?)}$$



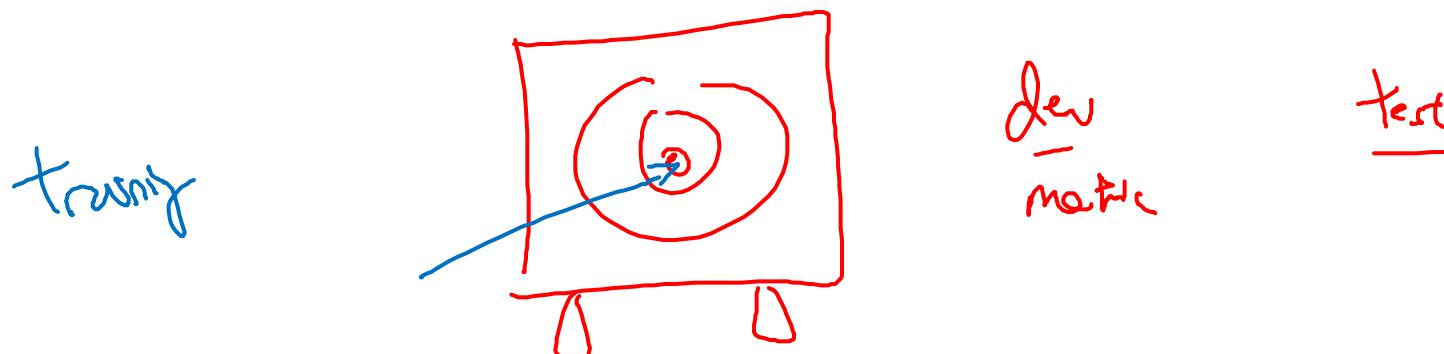
[Tested on low income zip codes

~ 3 month



Guideline

Choose a dev set and test set to reflect data you expect to get in the future and consider important to do well on.



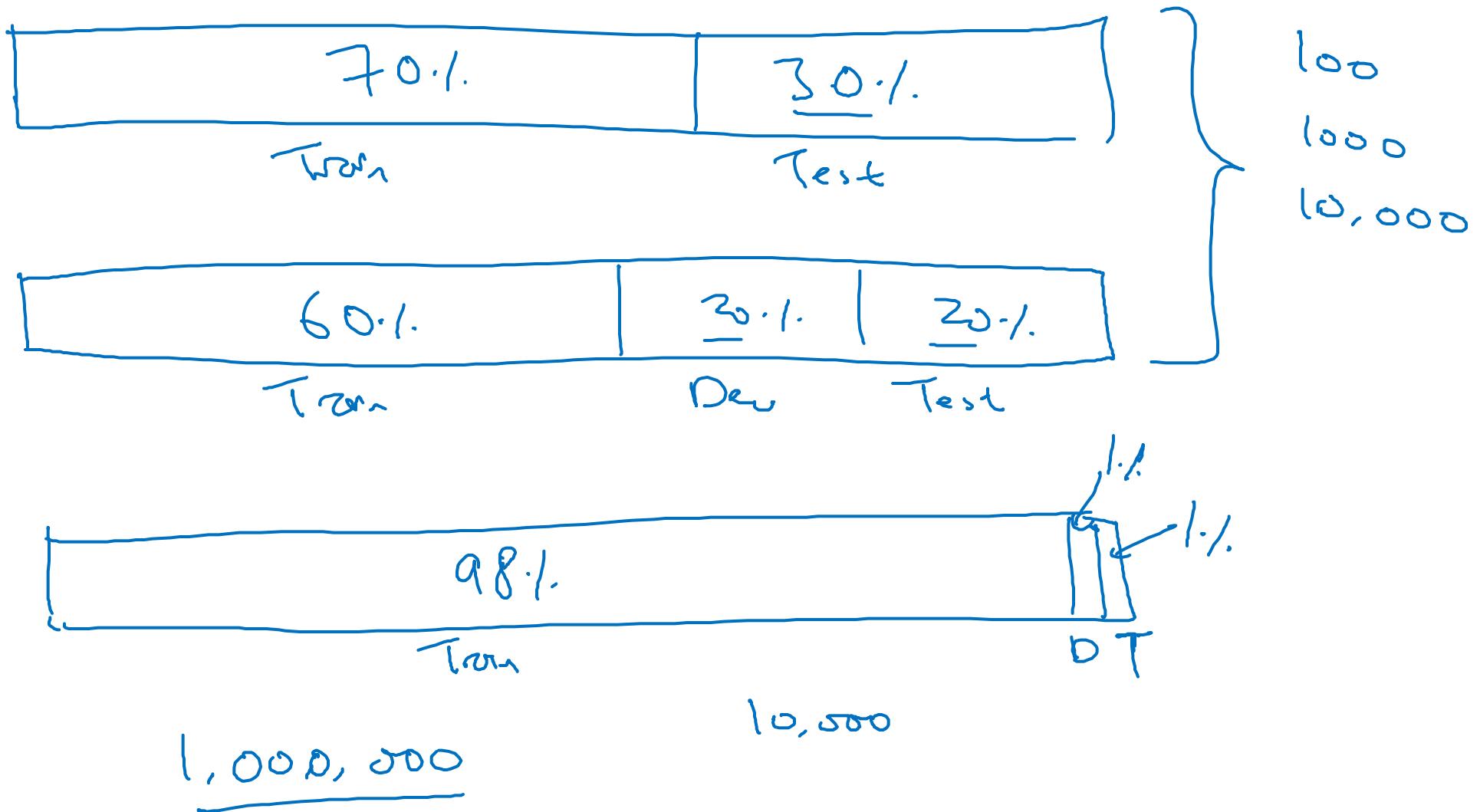


deeplearning.ai

Setting up
your goal

Size of dev
and test sets

Old way of splitting data



Size of dev set

A B

Set your dev set to be big enough to detect differences in
algorithm/models you're trying out.

100: small
 $\frac{1}{100} \approx 1\%$

A B
97% \rightarrow 97.1%
 $\frac{0.1\%}{1}$

1,000

10,000

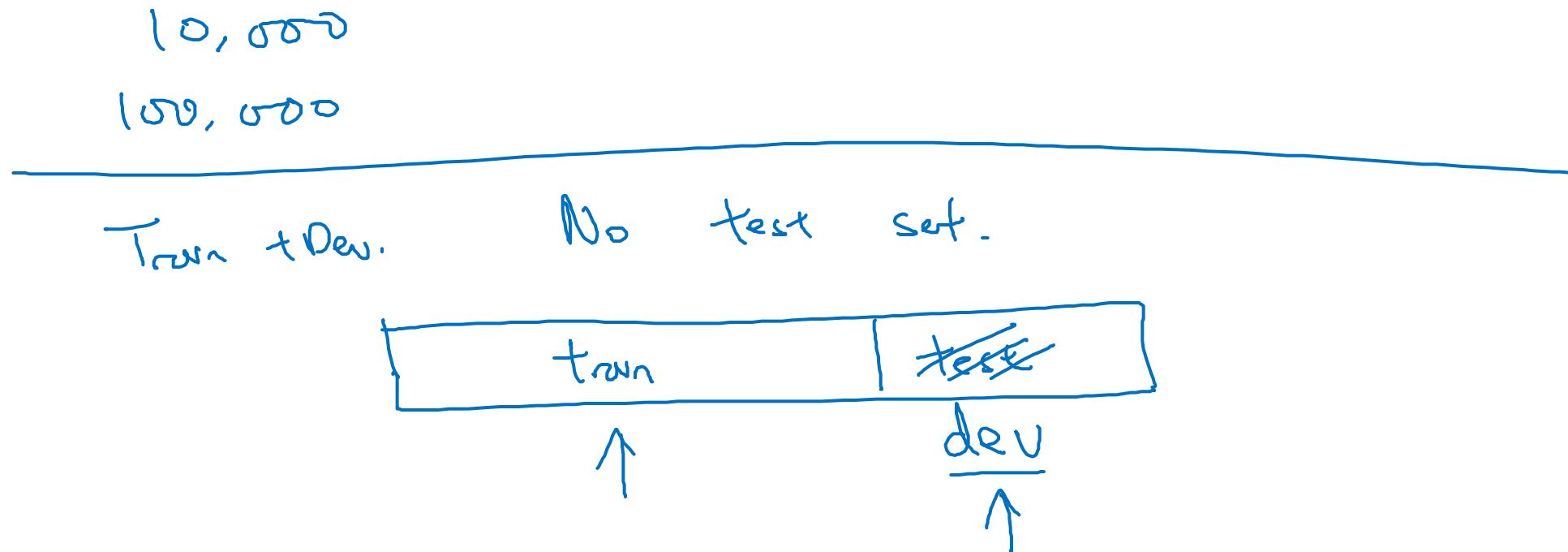
100,000

$\frac{0.01\%}{1}$
 $\frac{0.001\%}{1}$

Online advertising

Size of test set

→ Set your test set to be big enough to give high confidence in the overall performance of your system.





deeplearning.ai

Setting up
your goal

When to change
dev/test sets and
metrics

Cat dataset examples

Metric + Dev : Prefer A
You/Users : Prefer B.

→ Metric: classification error

Algorithm A: 3% error → Pornographic

✓ Algorithm B: 5% error

$$\left\{
 \begin{array}{l}
 \text{Error: } \frac{1}{m_{\text{dev}}} \sum_{i=1}^{m_{\text{dev}}} \omega^{(i)} \downarrow \left\{ \frac{y_{\text{pred}}^{(i)} + y^{(i)}}{\text{predicted value (0/1)}} \right\} \\
 \rightarrow \omega^{(i)} = \begin{cases} 1 & \text{if } x^{(i)} \text{ is non-zero} \\ 10 & \text{if } x^{(i)} \text{ is zero} \end{cases}
 \end{array}
 \right.$$