

Week 1-1 ML Strategy: Orthogonalization, Evaluation Metric, Train/Dev/Test Distributions

笔记本: DL 3 - Structuring ML Projects

创建时间: 2021/1/11 11:38

更新时间: 2021/1/11 13:38

Motivating example



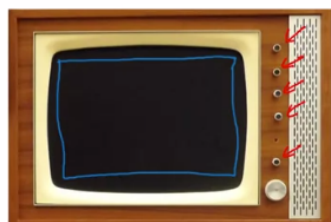
Ideas:

- 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
 - ...

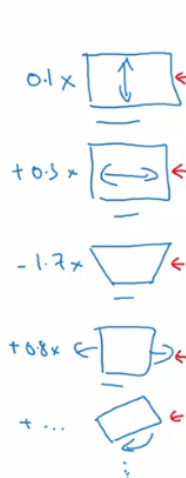
Andrew Ng

Orthogonalization

TV tuning example



Orthogonalization



Car



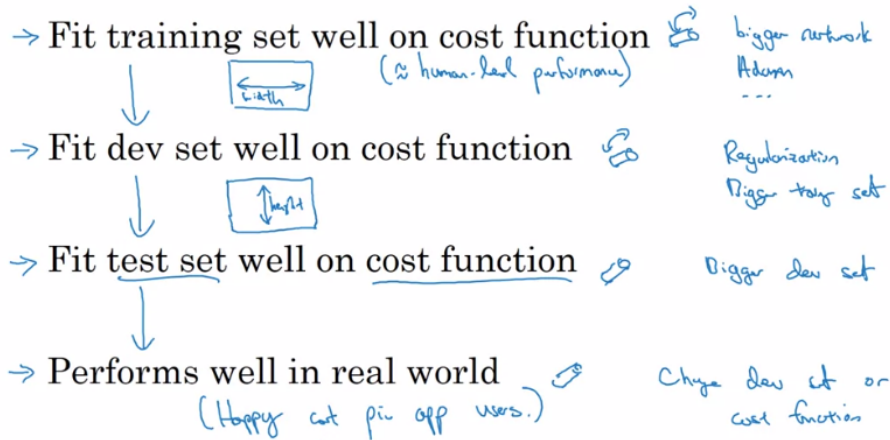
→ Steering]

→ Acceleration
Braking]

$$\begin{aligned} &\rightarrow 0.3 \times \text{angle} - 0.8 \times \text{speed} \\ &\rightarrow 2 \times \text{angle} + 0.9 \times \text{speed} \end{aligned}$$

Andrew Ng

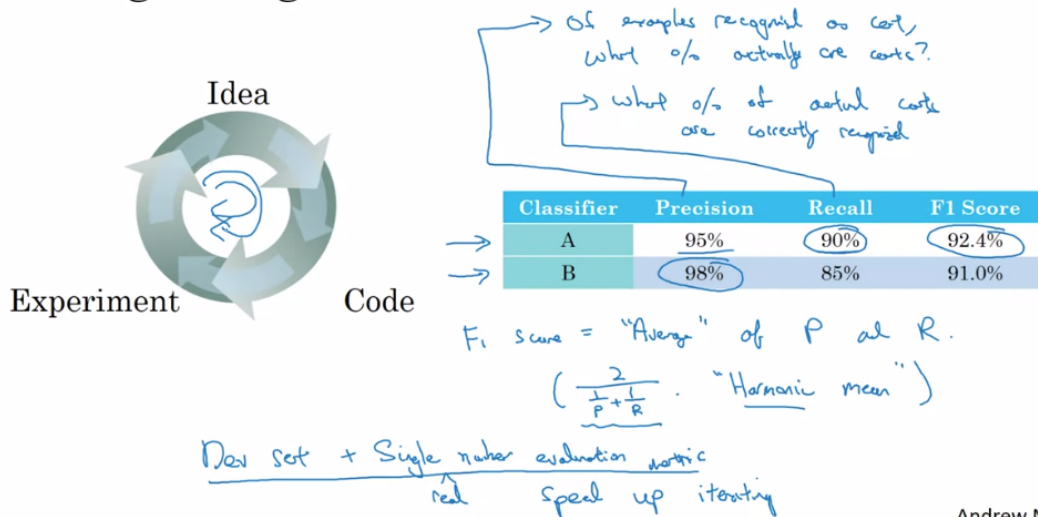
Chain of assumptions in ML



Andrew Ng

Single number evaluation metric

Using a single number evaluation metric



Andrew Ng

Satisfying and Optimizing Metric

Another cat classification example

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

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

Maximize Accuracy
Subject to Running Time $\leq 100 \text{ ms}$.

N metrics: 1 optimizing
N-1 satisfying

Wakewords / Trigger words
Alta, OK Google,
Hey Siri, nikosbaiden
你好 百度

accuracy.
#false positive

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

Andrew Ng

Train/Dev/Test Distributions

Wrong:

Cat classification dev/test sets

development set, hold out cross validation corp

Regions:

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

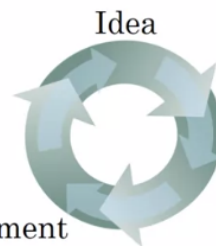
Dev

Test

Randomly shuffle into dev/test



dev set
+
Metric



Experiment

Code

Andrew Ng

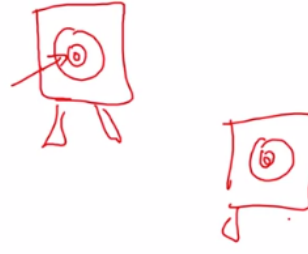
True story (details changed)

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

\uparrow $x \rightarrow y$ (repay loan?)

Tested on low income zip codes

~ 3 month



Guideline

Guideline

Some distribution

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

5. After setting up your train/dev/test sets, the City Council comes across another 1,000,000 images, called the "citizens' data". Apparently the citizens of Peacetopia are so scared of birds that they volunteered to take pictures of the sky and label them, thus contributing these additional 1,000,000 images. These images are different from the distribution of images the City Council had originally given you, but you think it could help your algorithm.

0 / 1 point

Notice that adding this additional data to the training set will make the distribution of the training set different from the distributions of the dev and test sets.

Is the following statement true or false?

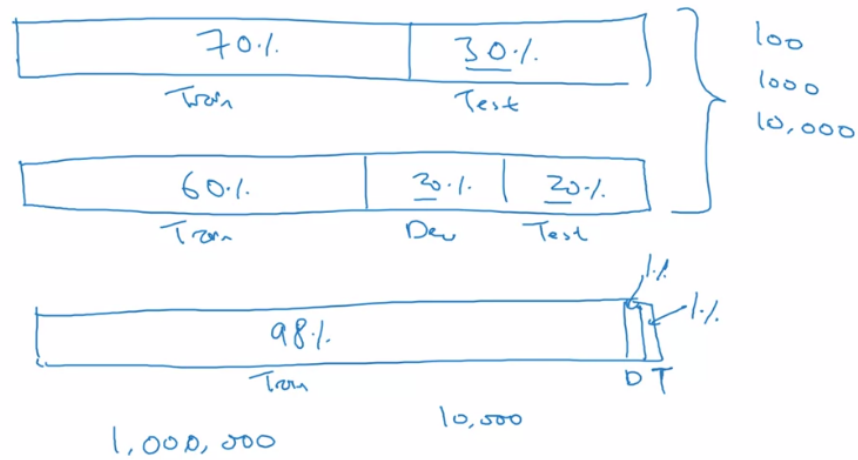
"You should not add the citizens' data to the training set, because if the training distribution is different from the dev and test sets, then this will not allow the model to perform well on the test set."

- ☒ True
- ☐ False

! Incorrect

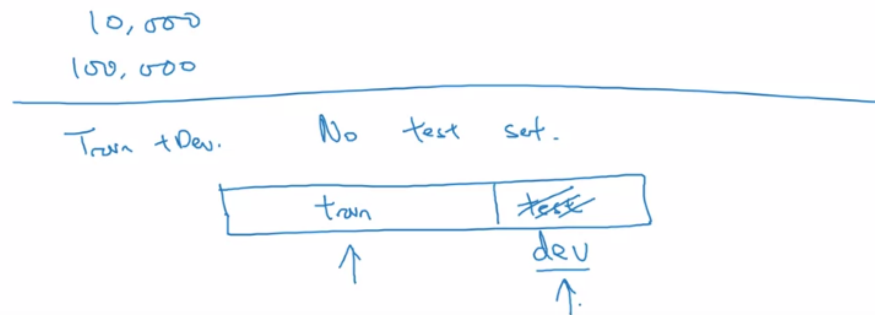
True is incorrect: Sometimes we'll need to train the model on the data that is available, and its distribution may not be the same as the data that will occur in production. Also, adding training data that differs from the dev set may still help the model improve performance on the dev set. What matters is that the dev and test set have the same distribution.

Old way of splitting data



Size of test set

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



When to change matrices and dev/test set

Cat dataset examples

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

→ Metric: classification error

Algorithm A: 3% error → pornographic

✓ Algorithm B: 5% error

$$\left\{ \begin{array}{l} \text{Error: } \frac{1}{\sum w^{(i)}} \times \frac{1}{m_{\text{dev}}} \sum_{i=1}^{m_{\text{dev}}} w^{(i)} \mathbb{I}\{y_{\text{pred}}^{(i)} \neq y^{(i)}\} \\ \rightarrow w^{(i)} = \begin{cases} 1 & \text{if } x^{(i)} \text{ is non-porn} \\ 10 & \text{if } x^{(i)} \text{ is porn} \end{cases} \end{array} \right.$$

$\mathbb{I}\{y_{\text{pred}}^{(i)} \neq y^{(i)}\}$ predicted value (0/1)

Another example

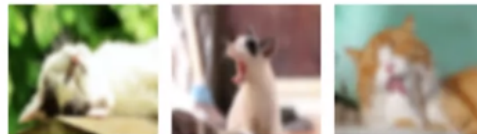
Algorithm A: 3% error

✓ Algorithm B: 5% error ←

→ Dev/test



→ User images



If doing well on your metric + dev/test set does not correspond to doing well on your application, change your metric and/or dev/test set.