



deeplearning.ai

# Introduction to ML strategy

## Why ML Strategy?

1

## 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

2

1



deeplearning.ai

## Introduction to ML strategy

---

## Orthogonalization

3

## TV tuning example



Car

Andrew Ng

4

2

## Chain of assumptions in ML

Fit training set well on cost function

Fit dev set well on cost function

Fit test set well on cost function

Performs well in real world

Andrew Ng

5



deeplearning.ai

Setting up  
your goal

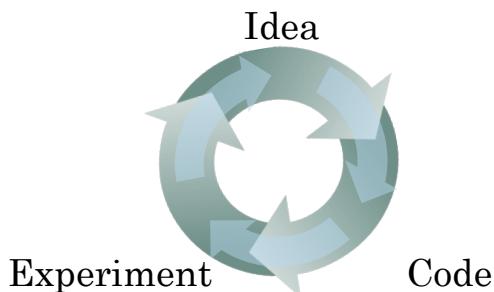
---

Single number  
evaluation metric

6

3

# Using a single number evaluation metric



Classifier	Precision	Recall
A	95%	90%
B	98%	85%

Andrew Ng

7

# Another example

Algorithm	US	China	India	Other
A	3%	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%

Andrew Ng

8



deeplearning.ai

Setting up  
your goal

---

Satisficing and  
optimizing metrics

9

## Another cat classification example

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

Andrew Ng

10



deeplearning.ai

Setting up  
your goal

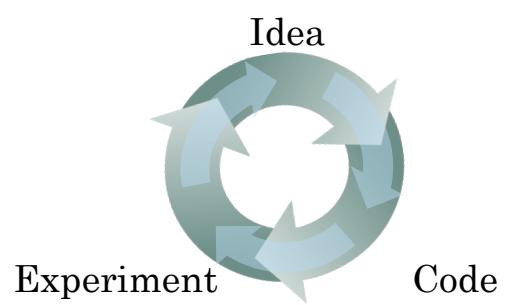
Train/dev/test  
distributions

11

## Cat classification dev/test sets

Regions:

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



Andrew Ng

12

## True story (details changed)

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

Tested on low income zip codes

Andrew Ng

13

## 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.

Andrew Ng

14



deeplearning.ai

Setting up  
your goal

---

Size of dev  
and test sets

15

Old way of splitting data

Andrew Ng

16

## Size of dev set

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

Andrew Ng

17

## Size of test set

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

Andrew Ng

18



deeplearning.ai

Setting up  
your goal

---

When to change  
dev/test sets and  
metrics

19

## Cat dataset examples

Metric: classification error

Algorithm A: 3% error

Algorithm B: 5% error

Andrew Ng

20

## Orthogonalization for cat pictures: anti-porn

1. So far we've only discussed how to define a metric to evaluate classifiers.
2. Worry separately about how to do well on this metric.



Andrew Ng

21

## 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.

Andrew Ng

22



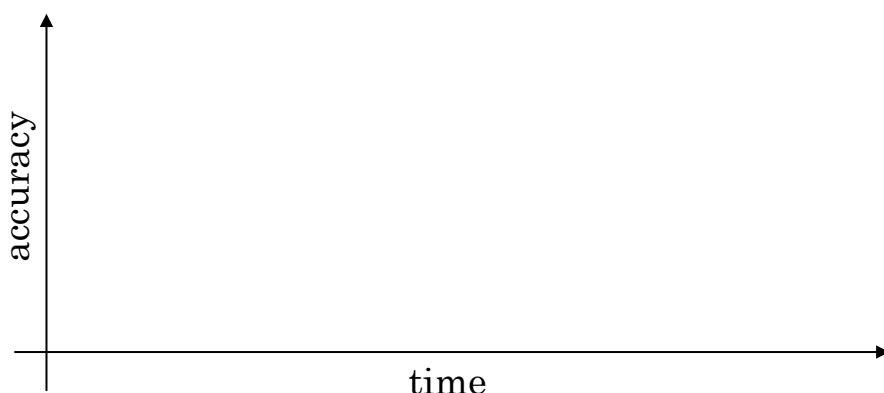
deeplearning.ai

Comparing to human-level performance

Why human-level performance?

23

Comparing to human-level performance



Andrew Ng

24

12

## Why compare to human-level performance

Humans are quite good at a lot of tasks. So long as ML is worse than humans, you can:

- Get labeled data from humans.
- Gain insight from manual error analysis:  
Why did a person get this right?
- Better analysis of bias/variance.

Andrew Ng

25



deeplearning.ai

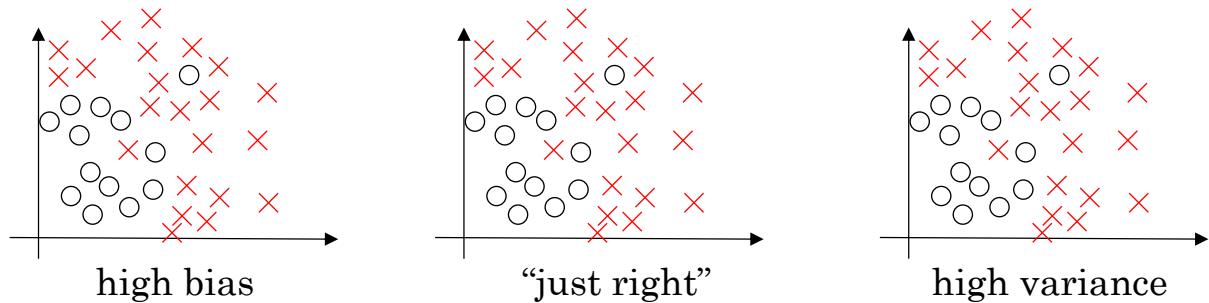
Comparing to human-level performance

---

Avoidable bias

26

## Bias and Variance



Andrew Ng

27

## Bias and Variance

Cat classification



Training set error:

Dev set error:

Andrew Ng

28

## Cat classification example

Training error	8%	8 %
Dev error	10%	10 %

Andrew Ng

29



deeplearning.ai

Comparing to human-level performance

Understanding  
human-level  
performance

30

## Human-level error as a proxy for Bayes error

Medical image classification example:



Suppose:

- (a) Typical human ..... 3 % error
- (b) Typical doctor ..... 1 % error
- (c) Experienced doctor ..... 0.7 % error
- (d) Team of experienced doctors .. 0.5 % error

What is “human-level” error?

Andrew Ng

31

## Error analysis example

Training error

Dev error

Andrew Ng

32

## Summary of bias/variance with human-level performance

Human-level error

Training error

Dev error

Andrew Ng

33



deeplearning.ai

Comparing to human-level performance

---

Surpassing human-level performance

34

## Surpassing human-level performance

Team of humans

One human

Training error

Dev error

Andrew Ng

35

## Problems where ML significantly surpasses human-level performance

- Online advertising
- Product recommendations
- Logistics (predicting transit time)
- Loan approvals

Andrew Ng

36



deeplearning.ai

Comparing to human-level performance

Improving your model performance

37

The two fundamental assumptions of supervised learning

1. You can fit the training set pretty well.
2. The training set performance generalizes pretty well to the dev/test set.

Andrew Ng

38

19

## Reducing (avoidable) bias and variance

Human-level

Train bigger model

Train longer/better optimization algorithms

Training error

NN architecture/hyperparameters search

Dev error

More data

Regularization

NN architecture/hyperparameters search

Andrew Ng