deeplearning.ai

Introduction to ML strategy

Why ML Strategy?

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

deeplearning.ai

Introduction to ML strategy

Orthogonalization

TV tuning example

Car

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

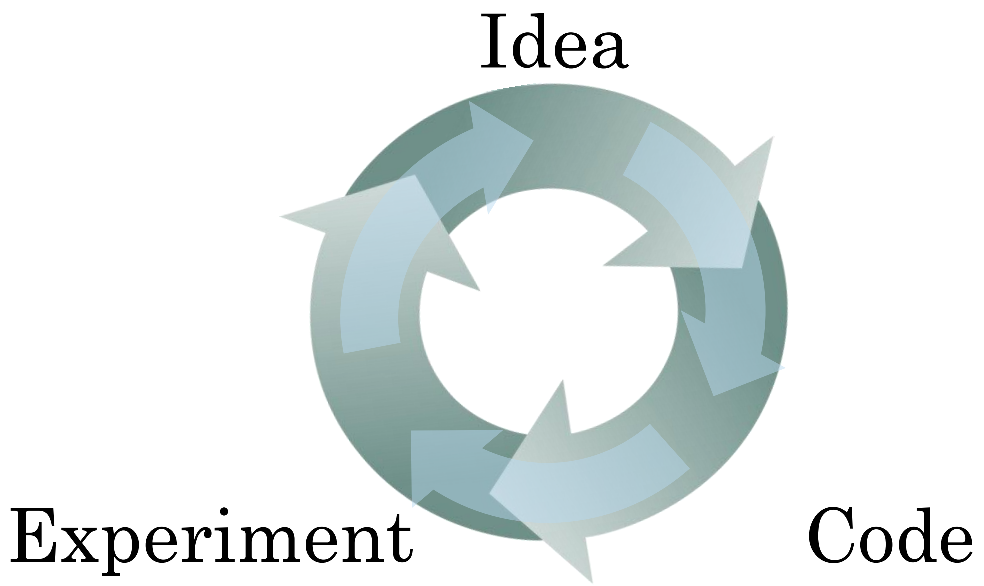
Performs well in real world

deeplearning.ai

Setting up your goal

Single number evaluation metric

Using a single number evaluation metric



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

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%

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

deeplearning.ai

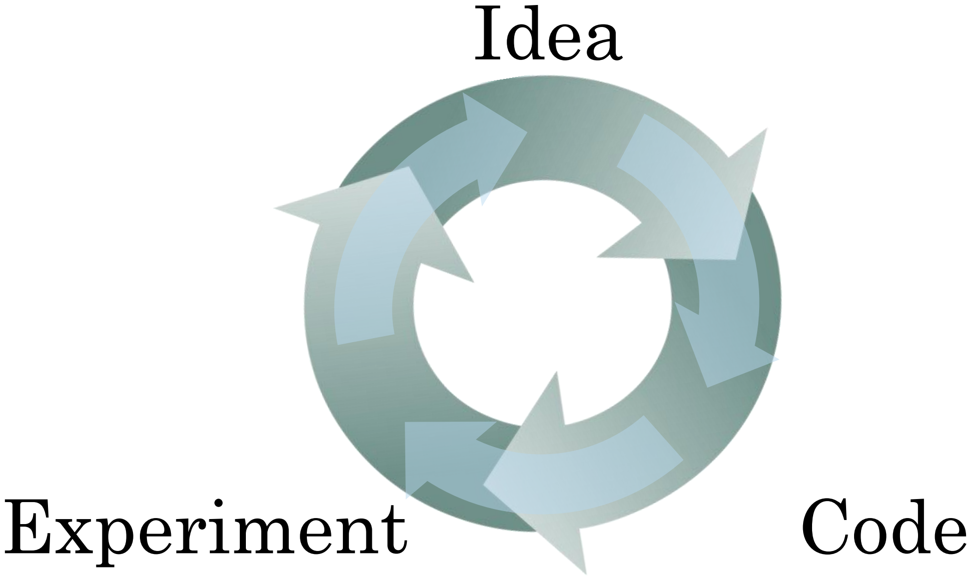
Setting up your goal

Train/dev/test distributions

Cat classification dev/test sets

Regions:

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



Andrew Ng

True story (details changed)

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

Tested on low income zip codes

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

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

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: classification error

Algorithm A: 3% error

Algorithm B: 5% error

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.

Another example

Algorithm A: 3% error

Algorithm B: 5% error

Dev/test

User images

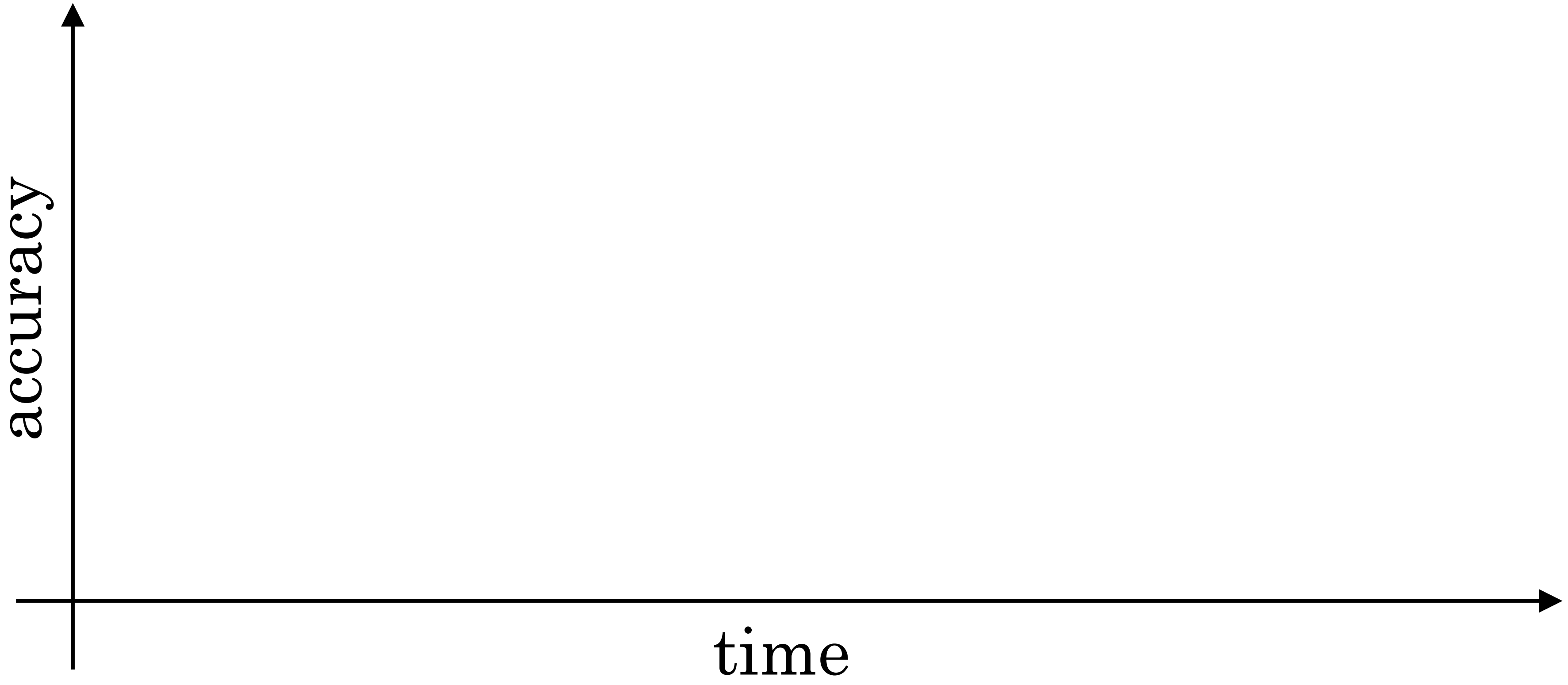
If doing well on your metric + dev/test set does not correspond to doing well on your application, change your metric and/or dev/test set.

deeplearning.ai

Comparing to human-level performance

Why human-level performance?

Comparing to human-level performance



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?

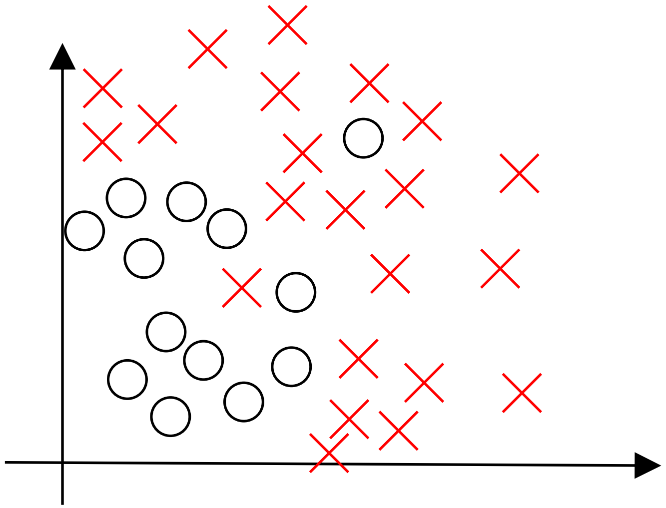
- Better analysis of bias/variance.

deeplearning.ai

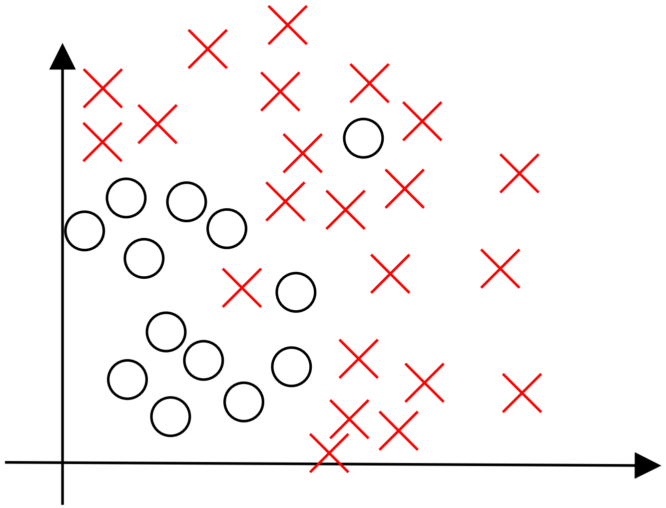
Comparing to human-level performance

Avoidable bias

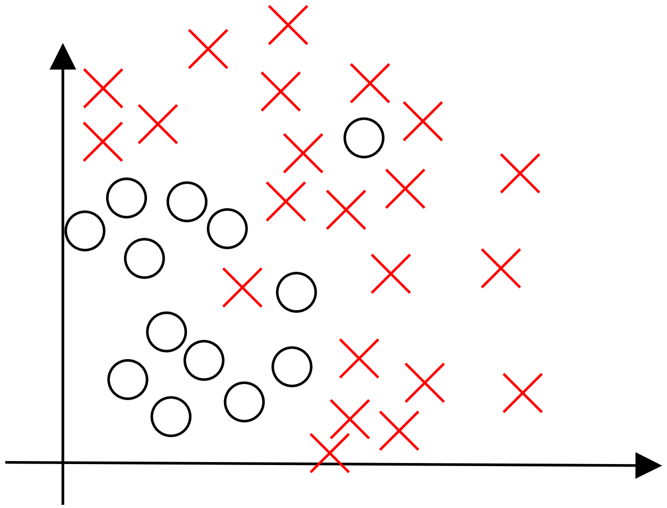
Bias and Variance



high bias



"just right"



high variance

Andrew Ng

Bias and Variance

Cat classification

Training set error:

Dev set error:

Cat classification example

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

deeplearning.ai

Comparing to human-level performance

Understanding human-level performance

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?

Error analysis example

Training error

Dev error

Summary of bias/variance with human-level performance

Human-level error

Training error

Dev error

deeplearning.ai

Comparing to human-level performance

Surpassing human-level performance

Surpassing human-level performance

Team of humans

One human

Training error

Dev error

Problems where ML significantly surpasses human-level performance

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

deeplearning.ai

Comparing to human-level performance

Improving your model performance

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.

Reducing (avoidable) bias and variance

Human-level

Train bigger model

Train longer/better optimization algorithms

NN architecture/hyperparameters search

Training error

More data

Regularization

Dev error

NN architecture/hyperparameters search