Comparing to human-level performance

deeplearning.ai

Why human-level performance?

Comparing to human-level performance

$X \rightarrow y$

$\rightarrow$ audio $\rightarrow$ transcript
$\rightarrow$ image $\rightarrow$ cat (0/1)

Bayes error
Bayes optimal error

best possible error

human

accuracy

time

Andrew Ng
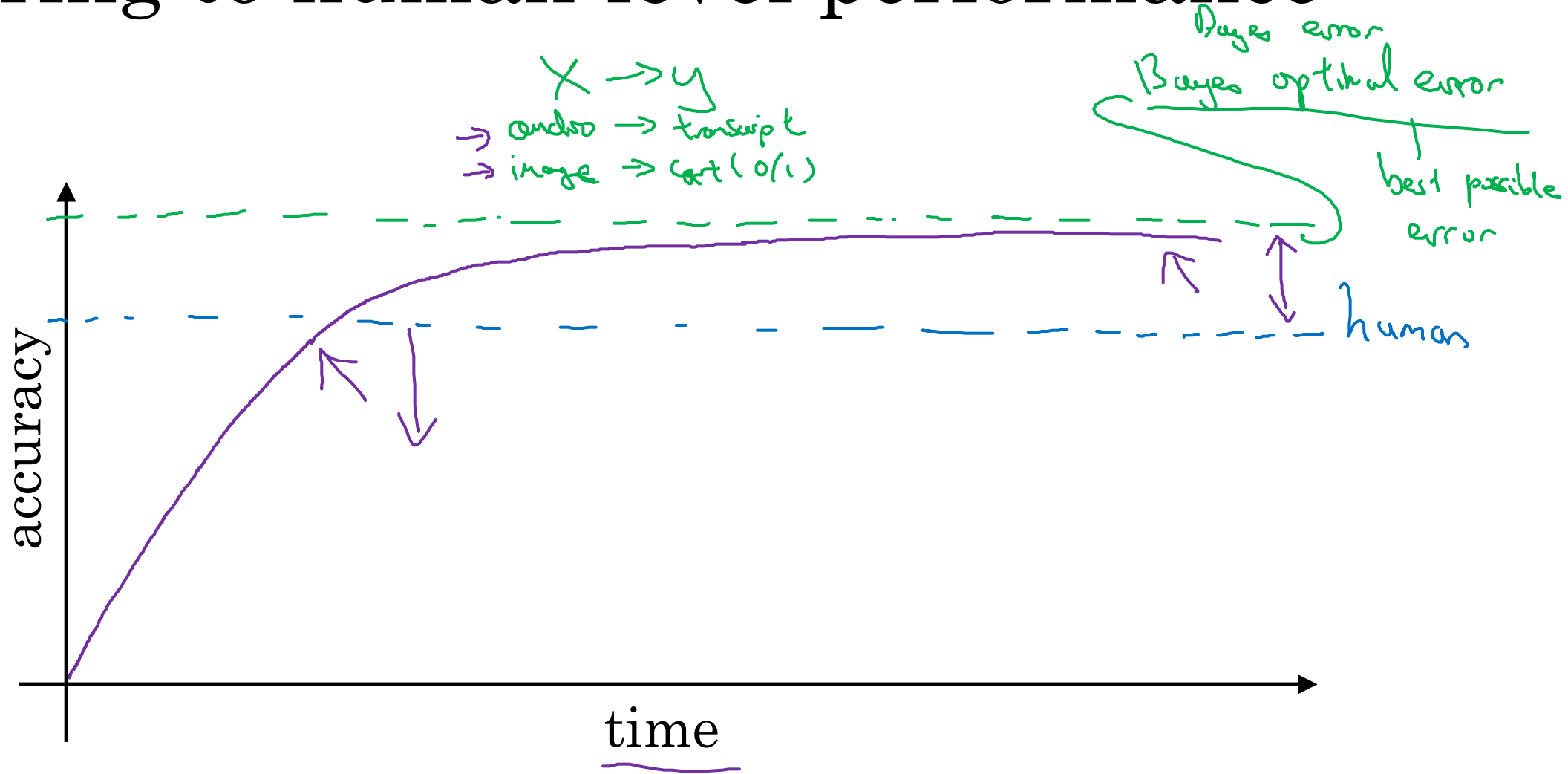
Why compare to human-level performance

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

- Get labeled data from humans. (x, y)

- Gain insight from manual error analysis:
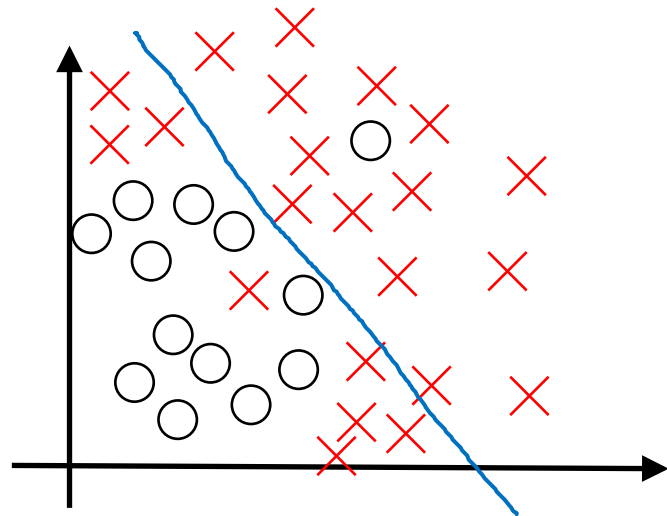 Why did a person get this right?

- Better analysis of bias/variance.

deeplearning.ai

Comparing to human-level performance
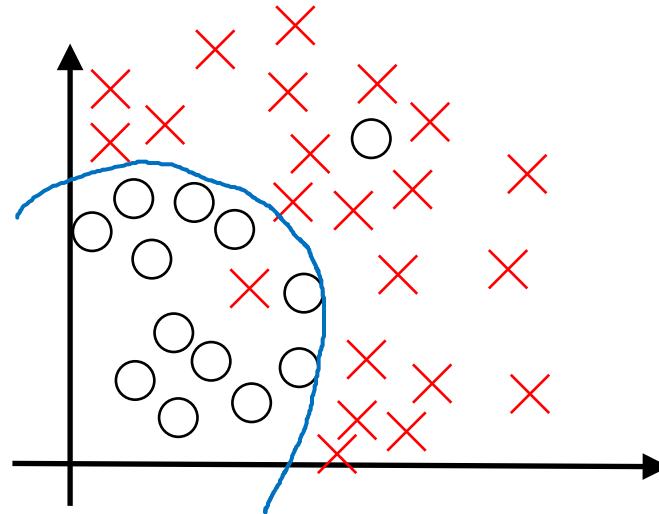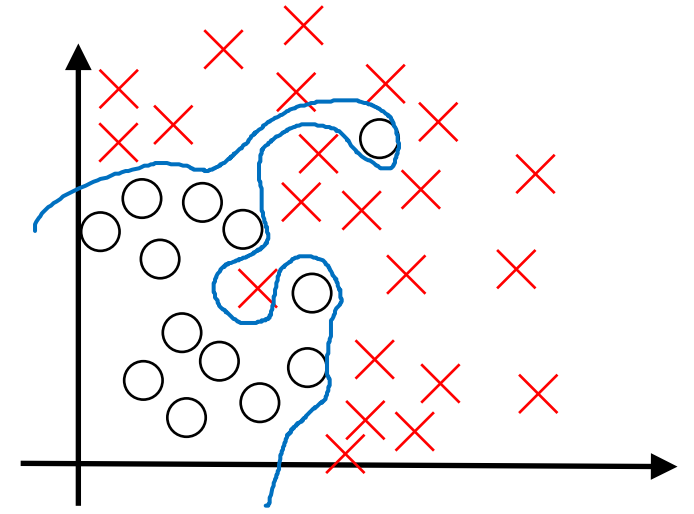
Avoidable bias

Bias and Variance

high bias
Underfitting

"just right"

high variance
Overfitting.

Andrew Ng

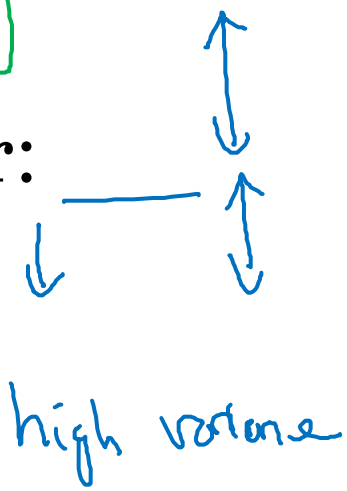# Bias and Variance

Cat classification

Human-level $\approx$ 0%
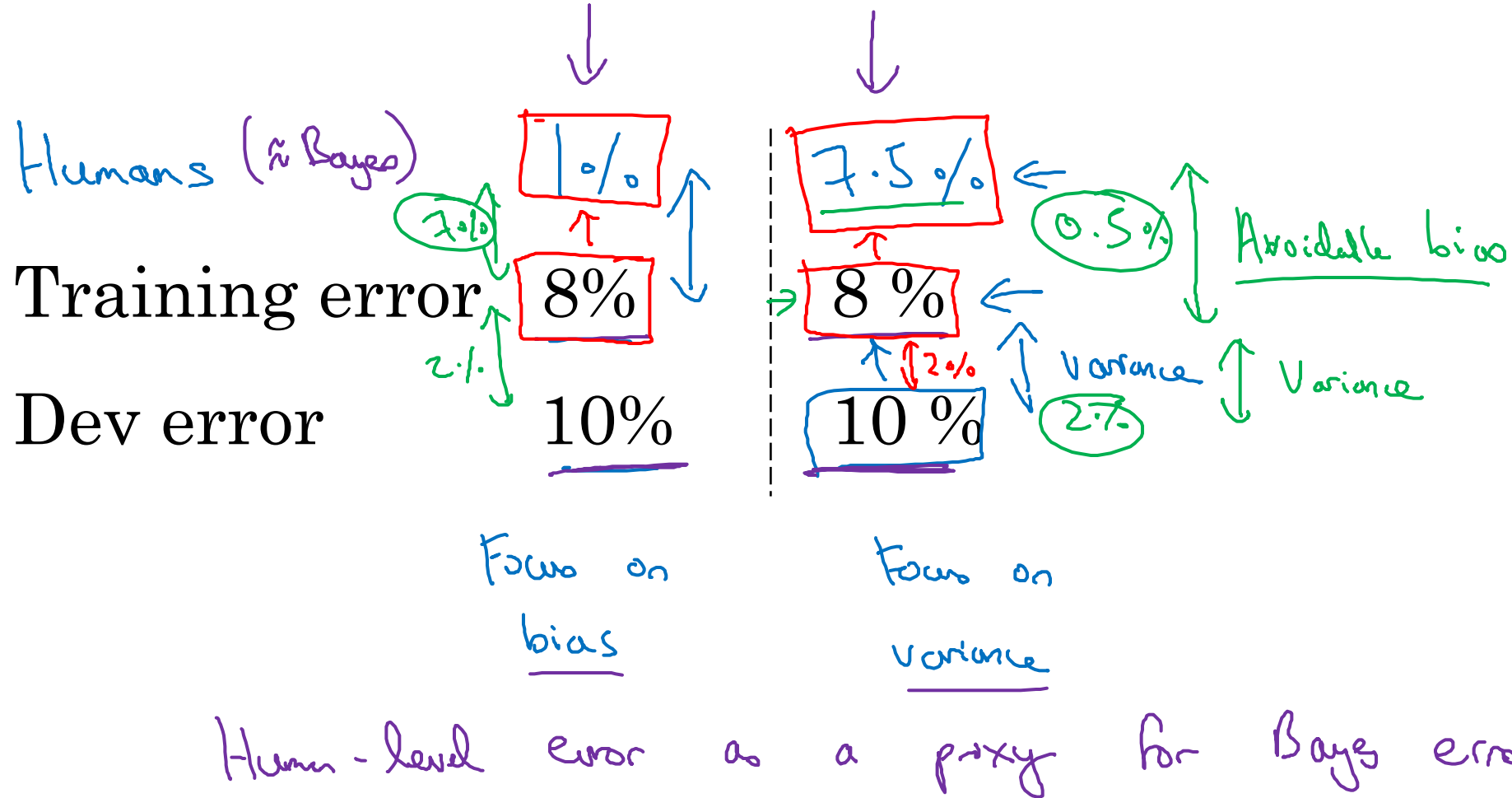
Training set error:

Dev set error:

high variance   high bias   high bias   low bias
                                           high variance   low variance

Cat classification example

Humans ($\tilde{\approx}$ Bayes)

Training error

Dev error

	Box 1	Box 2
Humans ($\tilde{\approx}$ Bayes)	1%	7.5%
Training error	8%	8%
Dev error	10%	10%

7% · 2%

0.5% — Avoidable bias

2% — Variance

2%

Focus on bias

Focus on variance

Human-level error as a proxy for Bayes error.

Andrew Ng

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 ←

$$\text{Bayes error} \leq 0.5\%$$

What is "human-level" error?

Error analysis example
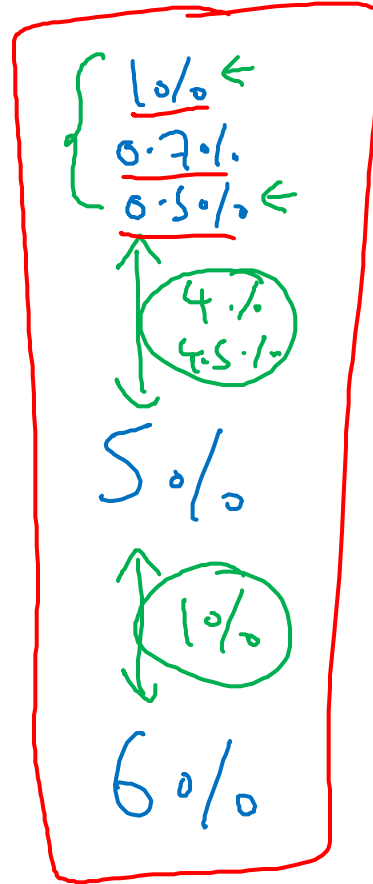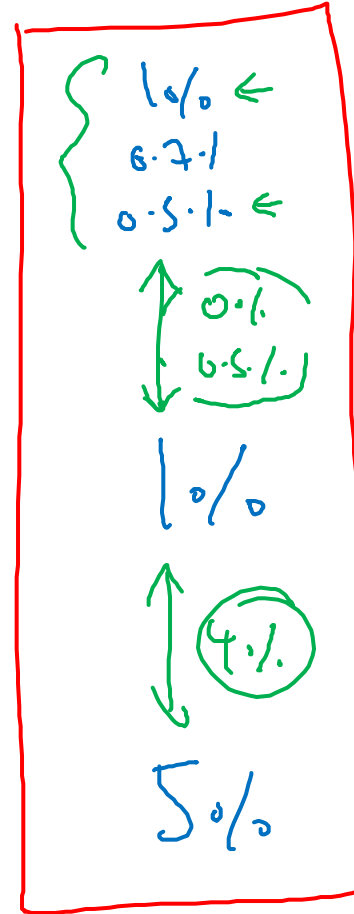

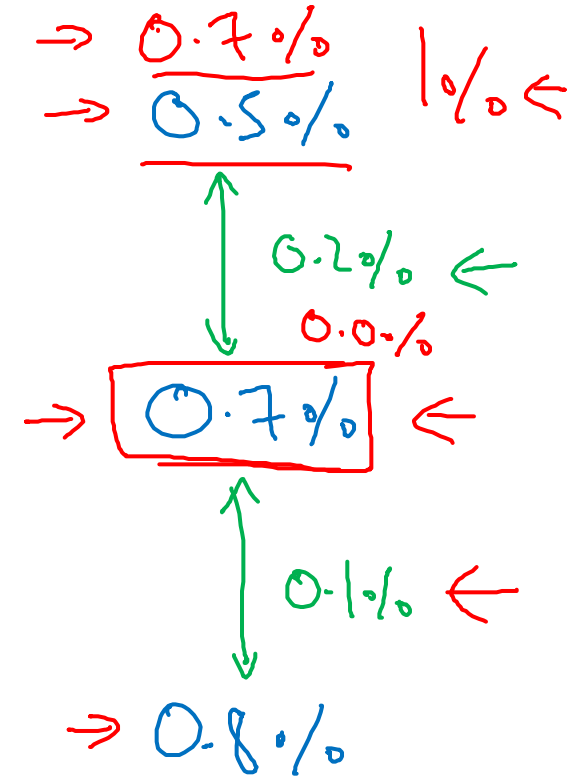Human (proxy for Bayes error)

Avoidable bias

Training error

Variance

Dev error

| 1% |
| 0.7% |
| 0.5% |

4% / 4.5%

5%

1%

6%

Bias

| 1% |
| 0.7% |
| 0.5% |

0% / 0.5%

1%

4%

5%

Variance

0.7%
0.5%   1%

0.2%
0.0%

0.7%

0.1%

0.8%

Andrew Ng

Summary of bias/variance with human-level performance

$O\%$

"Bias"

Human-level error

(proxy for Bayes error)
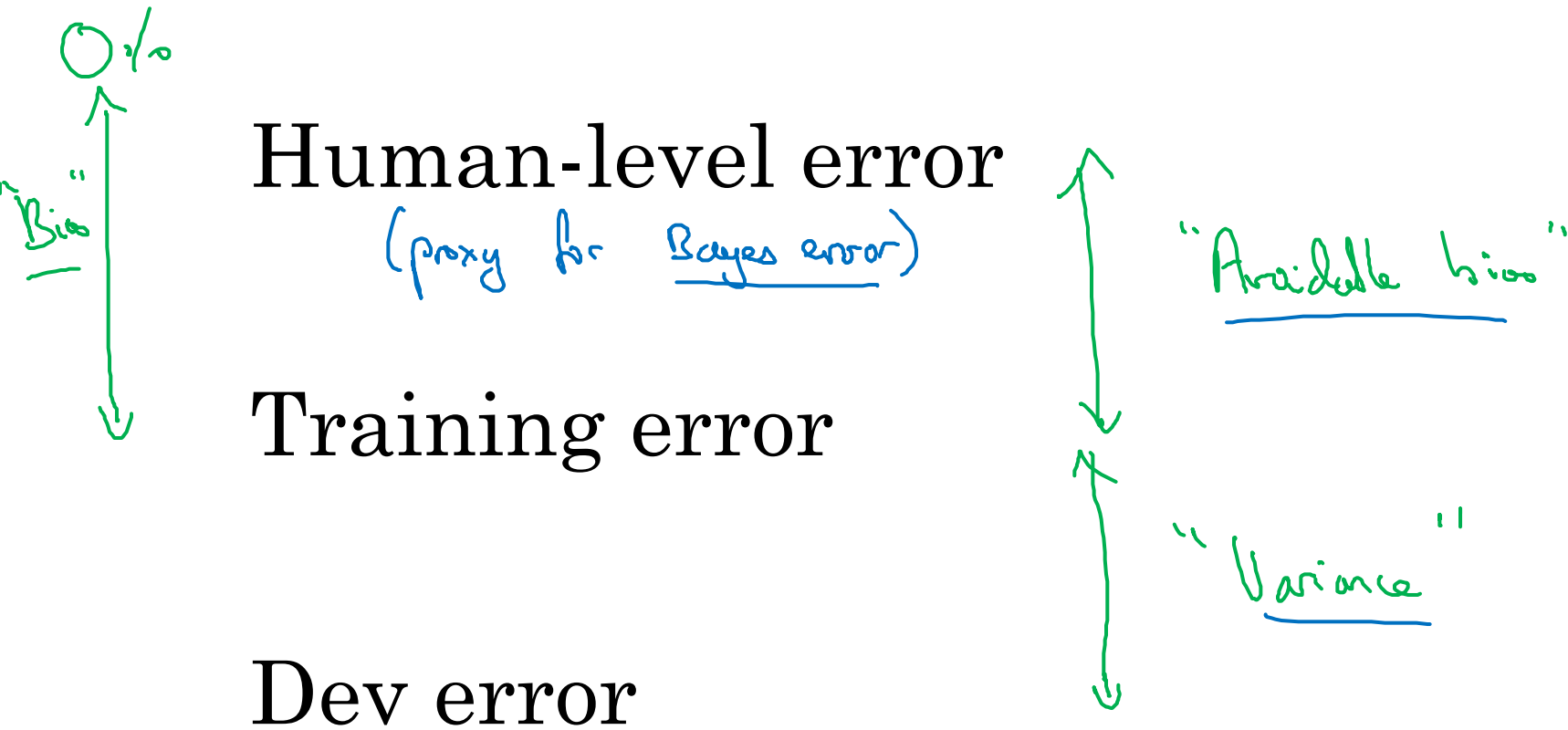
"Avoidable bias"

Training error

"Variance"

Dev error

deeplearning.ai

Comparing to human-level performance

Surpassing human-level performance

Surpassing human-level performance
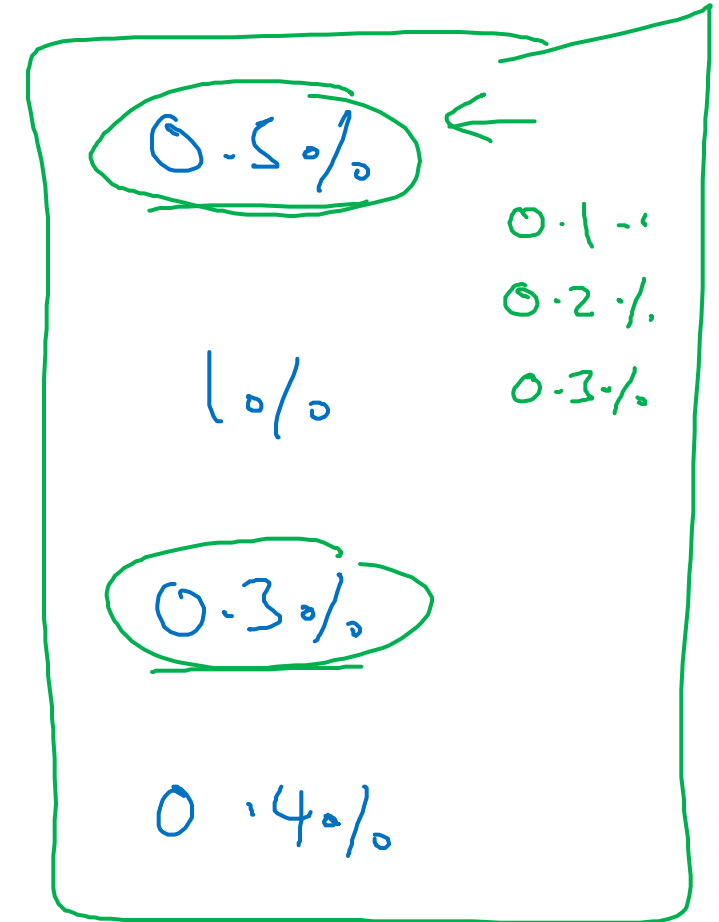
Team of humans   0.5%

One human   ~~1%~~

Training error   0.6%

Dev error   0.8%

0.1

0.2

What is avoidable bias?

0.5%

0.1%
0.2%
0.3%

1%

0.3%

0.4%

Problems where ML significantly surpasses human-level performance

$\rightarrow$ - Online advertising

$\rightarrow$ - Product recommendations

$\rightarrow$ - Logistics (predicting transit time)

$\rightarrow$ - Loan approvals

Structured data

Not natural perception

Lots of data

- Speech recognition
- Some image recognition
- Medical
    - ECG, Skin cancer, ...

Andrew Ng

Comparing to human-level performance

Improving your model performance

deeplearning.ai

The two fundamental assumptions of supervised learning

1. You can fit the training set pretty well.

~ Avoidable bias

2. The training set performance generalizes pretty well to the dev/test set.

~ Variance

Andrew Ng

Reducing (avoidable) bias and variance

Human-level

Avoidable bias

Training error

Variance

Dev error

Train bigger model

Train longer/better optimization algorithms
- Momentum, RMSprop, Adam

NN architecture/hyperparameters search

RNN
CNN

More data

Regularization
- L2, dropout, Data augmentation

NN architecture/hyperparameters search

Andrew Ng