Comparing to human-level performance

Understanding human-level performance

deeplearning.ai

Human-level error as a proxy for Bayes error

Medical image classification example:

Suppose:

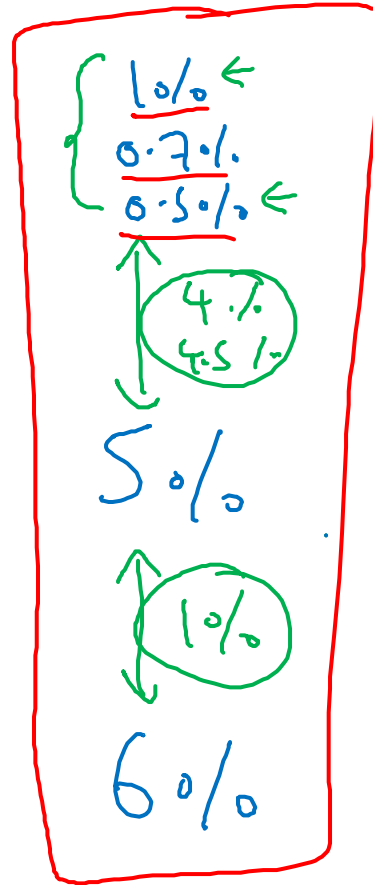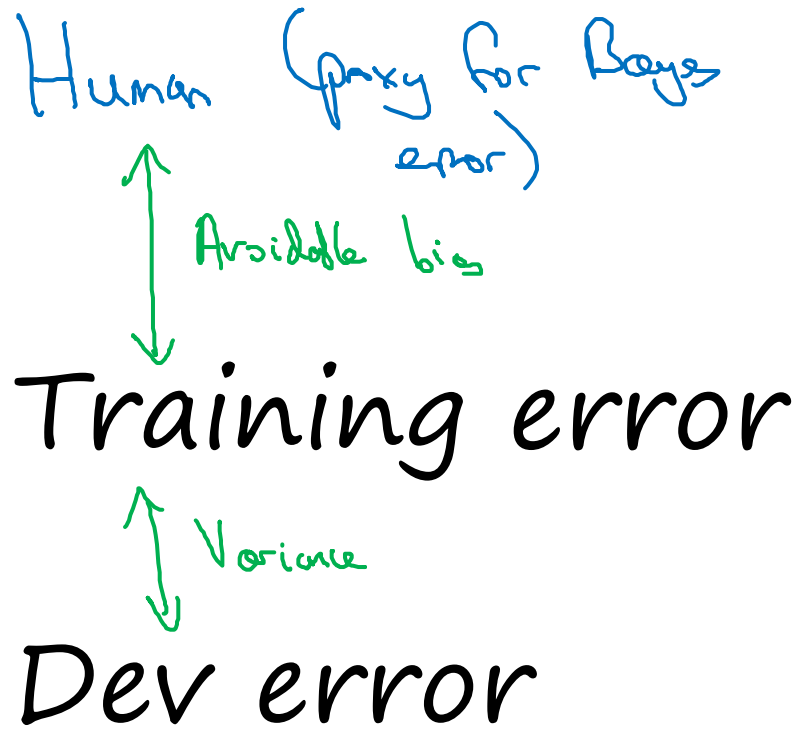   (a) Typical human .... 3 % error

    (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%

↕ (0.1% / 0.5%)

1%

↕ (4%)

5%

↑ Variance

⇒ 0.7%
⇒ 0.5%   1% ←

↕ 0.2% ←
  0.0% →

⇒ [0.7%] ←

↕ 0.1% ←

⇒ 0.8%

Andrew Ng
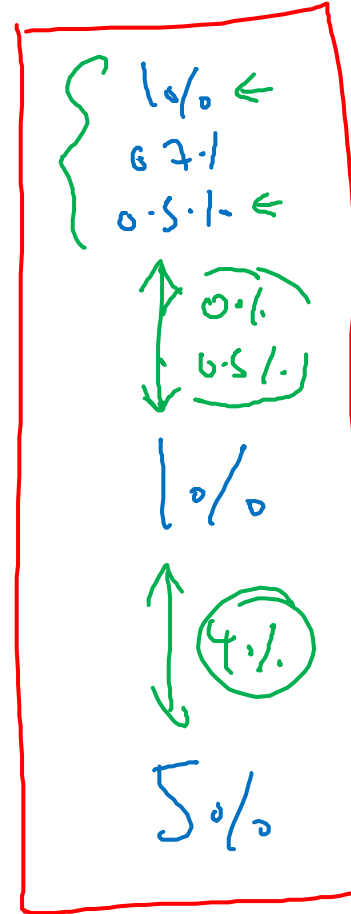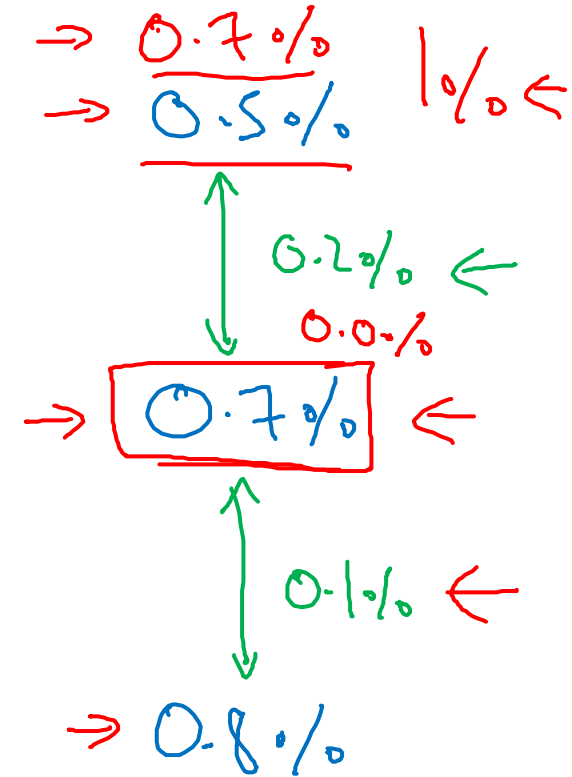
Summary of bias/variance with human-level performance

0%

"Bias"

Human-level error

(proxy for Bayes error)

"Avoidable bias"

Training error

"Variance"

Dev error