



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

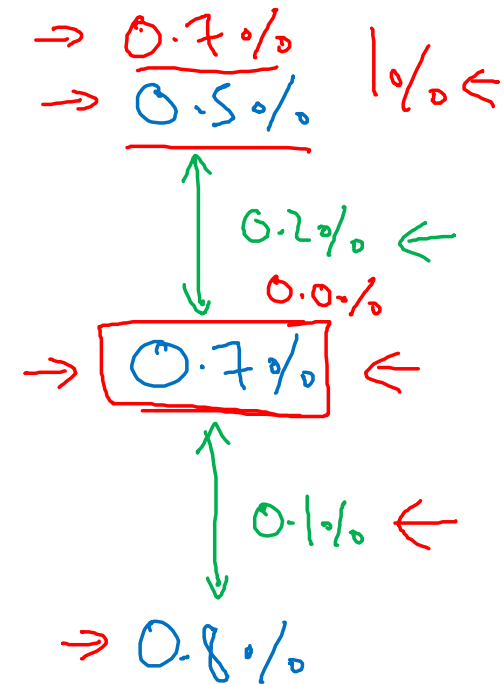
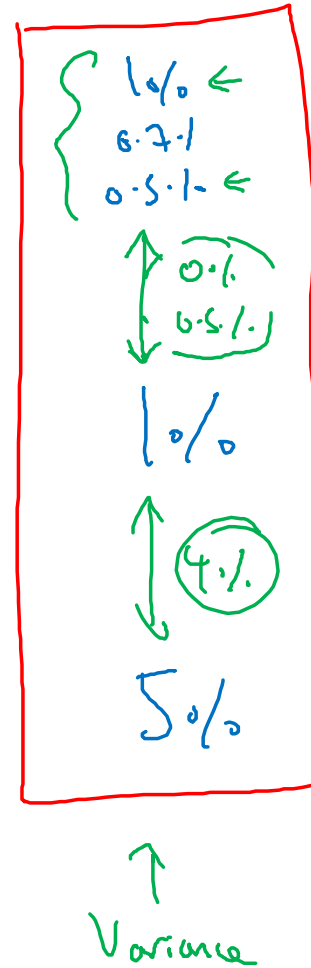
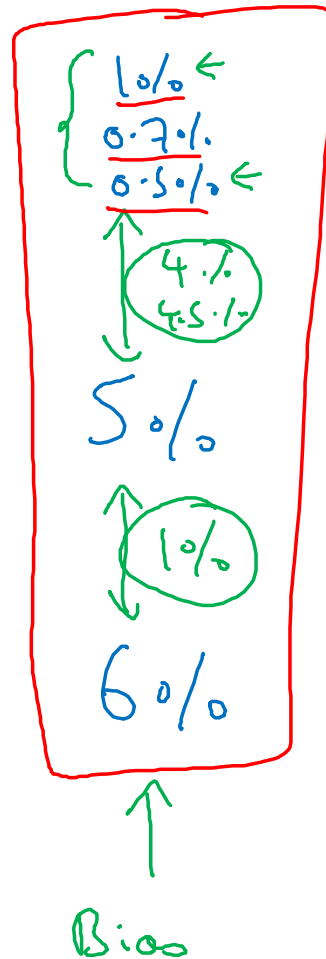
(c) Experienced doctor 0.7 % error

→ (d) Team of experienced doctors .. 0.5 % error ←

Bayes error \leq 0.5 %

What is “human-level” error?

Error analysis example



Summary of bias/variance with human-level performance

