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 $\leq 0.5\%$.

What is "human-level" error?

Error analysis example

- As your model starts performing well, as in case of (C) it becomes harder to decide where to focus on (bias or variance)
- In (B) and (A), it was easy to tell where to focus

Human (proxy for Bayes) 1% or .7% or .5%

	(A)	(B)	(C)
Training error	5%	1%	0.7%
Dev error	6%	5%	0.8%

In (C), Avoidable bias = 0 to 0.2%
Variance = 0.1%
Q Which bias to choose? should choose 0.2% as it shows there is Room for Improvement
Focus on bias (.2 > .1)

In (A), Variance = 1%
Avoidable bias is a Range 5-1 = 4% to 5-.5 = 4.5% (FOCUS HERE)

In (B), Avoidable bias doesn't matter can be b/w 0% & .5%
Variance = 4% (FOCUS HERE)

Andrew Ng

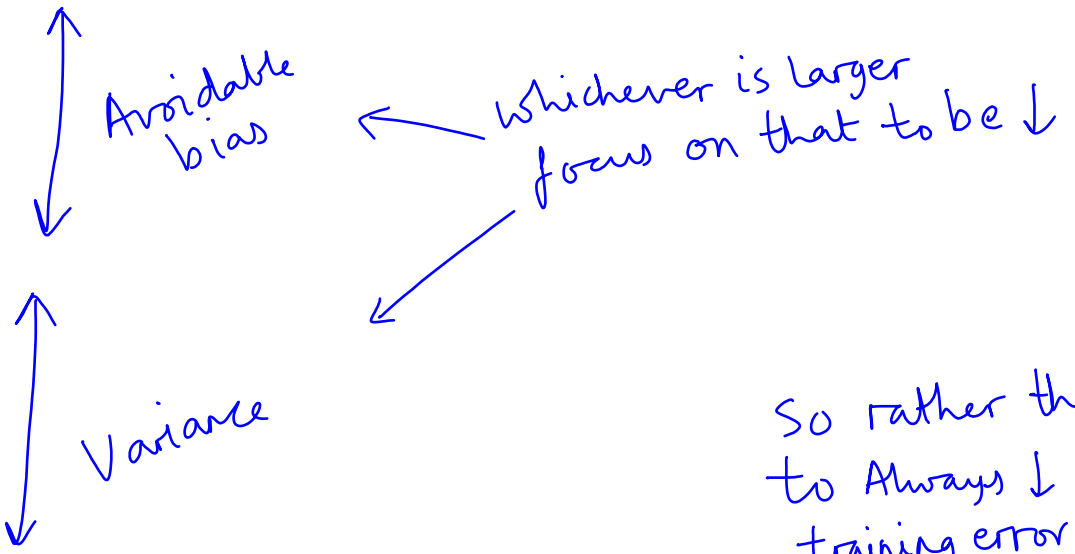
Summary of bias/variance with human-level performance

Human-level error

(proxy of Bayes)

Training error

Dev error



So rather than trying to always ↓ your training error to 0, try & see what the Bayes limit is - That will act as your lower limit, may not be always possible to go to 0
eg noisy Audio