



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?

그때그때 달라요.

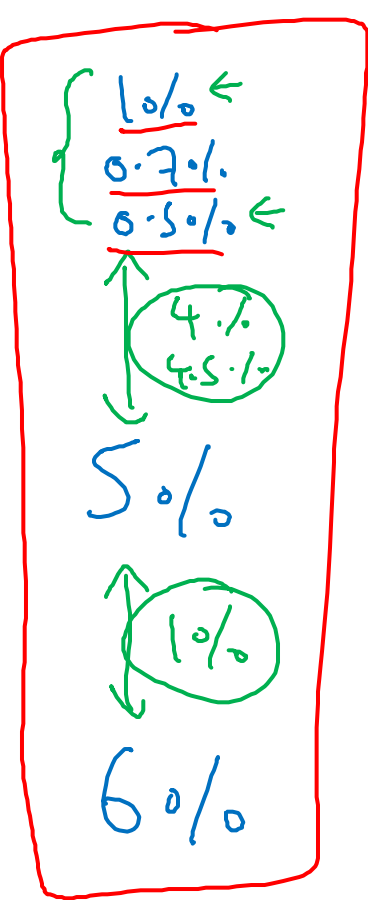
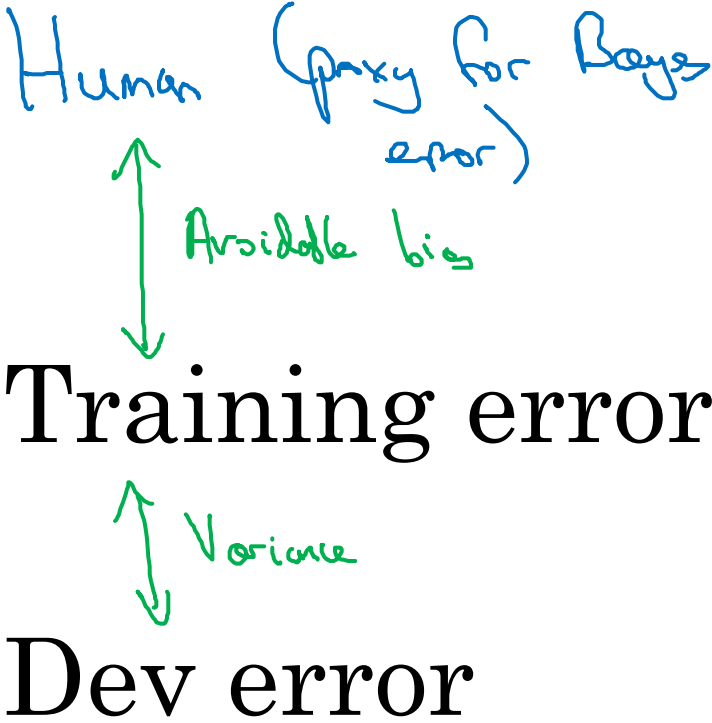
(b) : 연구 목적이 알고리즘이 의사를 이길수 있는가? 이면 이거

(d) : bayes optimal error의 추정값으로 쓸 목적이면 이거

Bayes error \leq 0.5%

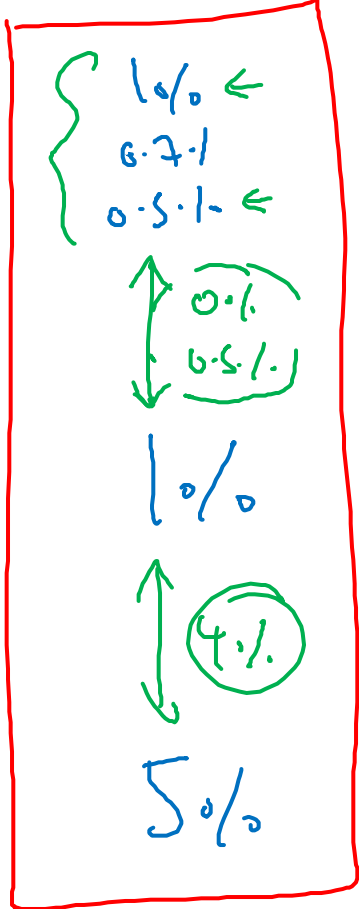
bayes error가 얼마나 더 낮아질 수 있는지는 알 수 없지만
분명한건 0.5% 보다는 클 수 없다는 것이다.

Error analysis example



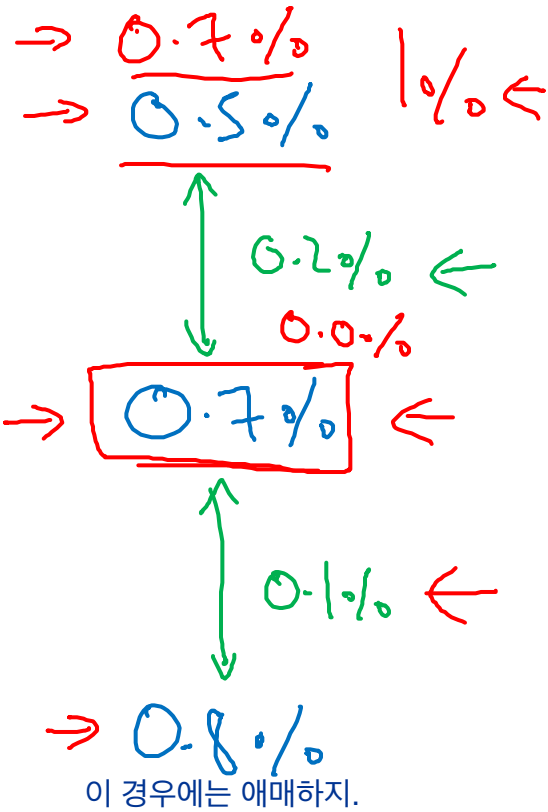
Bias

이 경우에는 휴먼레벨을 밑로 보든간에 bias 줄이는 쪽으로 가야되고



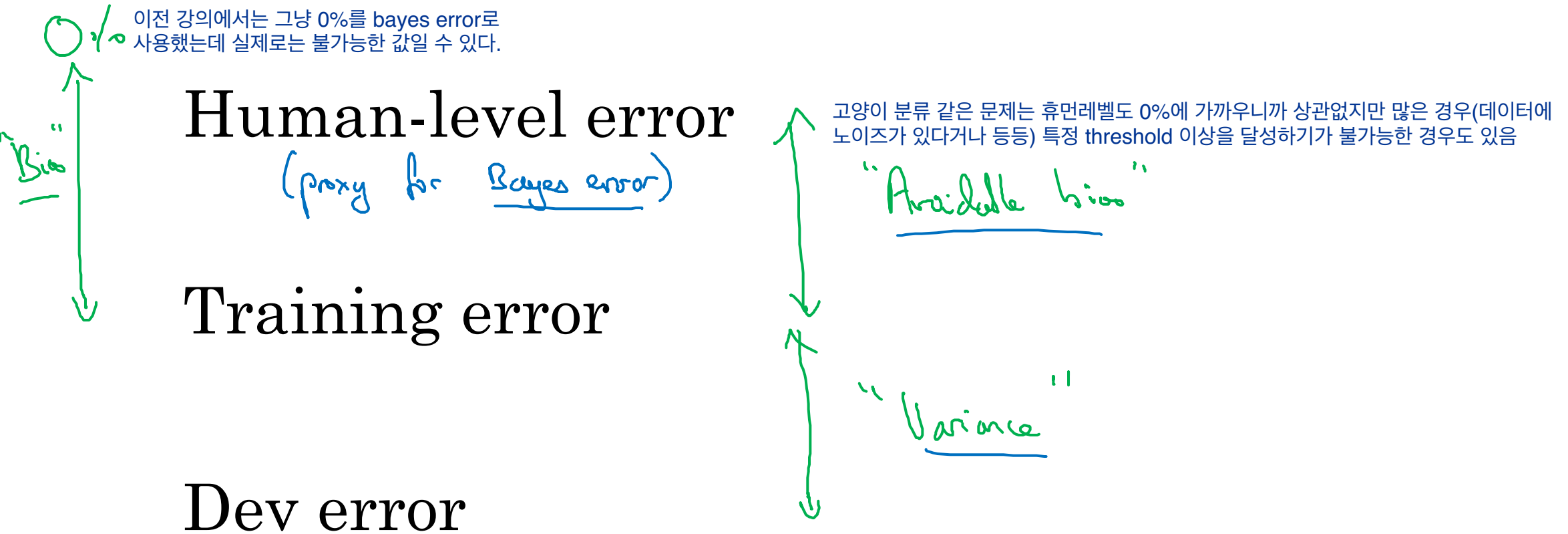
Variance

이 경우에는 휴먼레벨을 밑로 보든간에 variance를 줄이는 쪽으로 가야되고



이것만 봐도 휴먼레벨에 접근할 수록 왜 급격한 성능향상이 어려울지를 알 수 있다. 어떤 정책을 써야할지도 애매하고 어떤걸 휴먼레벨로 정해야 할지도 애매하고 bayes error가 어디까지 좋아지는지도 알 수 없으므로. 반면 왼쪽 두 경우는 그냥

Summary of bias/variance with human-level performance



휴먼레벨 퍼포먼스의 추정값은 bayes error에 대한 추정을 해주며 어떤 정책을 취해야하는지 빨리 알 수 있다.
근데 이런 테크닉은 휴먼레벨을 넘기 전까지만 잘 동작한다. 휴먼레벨 넘어가면 어떻게 해야하는지 다음 강의에서 알아볼까?