Machine and Human Learning

Yixin Lin

Duke University

yixin.lin@duke.edu

September 22, 2017

Overview

Where does expert performance come from?

- What do I mean by expert performance?
  - Not money, intelligence, success, social acclaim, or happiness (though may correlate)
  - Demonstrable, reproducible ability to achieve desired outcomes in well-defined domain
  - Why?
- The typical dichotomy
  - Talent ("nature", innate ability, perhaps from genes; something inborn which defines your limits)
  - Hard work (10,000 hours $\Leftrightarrow$ world-class performance)

Could it be hard work?

- Billions of people spend $> 10,000$ hours on their work
  - Just 5 years of full-time work (40 hr/wk $\cdot$ 50 wk/yr $\cdot$ 5 yr $=$ 10,000hr)
- People don't get better with more work; some get *worse*
  - Doctors get worse at diagnosing X-rays, heart sounds, tests of medicine
- I'm not going to be Olympic gymnast, NBA player

Could it be talent?

- Types of conceptions of talent
  - Natural born skills (intelligence, memory)
  - Ability to grasp a field quickly/early
  - Larger limitations and potential
- Failure to generalize skills
  - Case study: chess memory, chunking[1,2], "numbers to leave numbers", "form to leave form"[3]
- If true, there should be individuals who achieve much faster than others: kind of, but not really (research which led to 10k hours)
- People given a new methodology suddenly overcome plateaus
- Progression of "expert" performance over time

[1] Chase, W. G., Simon, H. A. (1973a). Perception in chess. *Cognitive Psychology, 4*, 55-81.

[2] Gobet, Fernand, et al. "Chunking mechanisms in human learning." Trends in cognitive sciences 5.6 (2001): 236-243.

[3] *The Art of Learning*, Josh Waitzkin

Memory tasks

- *n*-back task
- Memory competitions (*Moonwalking with Einstein*)
- Not deliberate practice, but demonstrates power of repeated applied practice

Deliberate practice

- "Living in a cave does not make you a geologist..."[4]
- Theoretical framework developed by K. Anders Ericsson[5]
- Controversial whether it explains all expert performance
- "Gold standard" of improvement
- Key aspects
  - Designed, often by an expert teacher, to improve specific performance
  - Quick feedback loops
  - Requires intense concentration (therefore often only happens a few hours a day)
    - "If you practice with your fingers, no amount is enough. If you practice with your head, two hours is plenty." –Leopold Auer
  - Repeated a lot

[4] Ericsson, K. Anders, Michael J. Prietula, and Edward T. Cokely. "The making of an expert." Harvard business review 85.7/8 (2007): 114.

[5] Ericsson, K. Anders, Ralf T. Krampe, and Clemens Tesch-Rmer. "The role of deliberate practice in the acquisition of expert performance." Psychological review 100.3 (1993): 363.

Case study: Berlin violinist experiment

- Diaries of 3 groups of young violinists: best, better, teacher, as well as a professional group of top-tier orchestra
- Key factor: individual practice, versus all musical activities

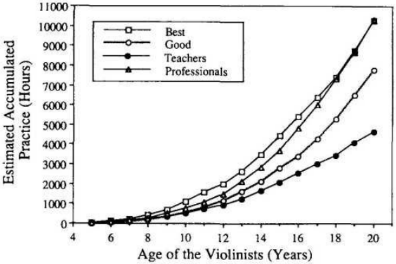


Figure 9. Accumulated amount of practice alone (on the basis of estimates of weekly practice) as a function of age for the middle-aged violinists (△), the best violinists (□), the good violinists (○), and the music teachers (●).

What's wrong with 10,000 hours

- Ericsson's thesis: Monotonic improvement with deliberate practice
- Gladwell's thesis: 10,000 hours $\Leftrightarrow$ mastery
- Problems with Gladwell's thesis, according to Ericsson:
  - 10k hours arbitrary number: they weren't masters yet; e.g. 20-25k for international piano competition winners
  - Only 10k by age 20 *on average*
  - Not necessarily deliberate practice
  - Not "if and only if"

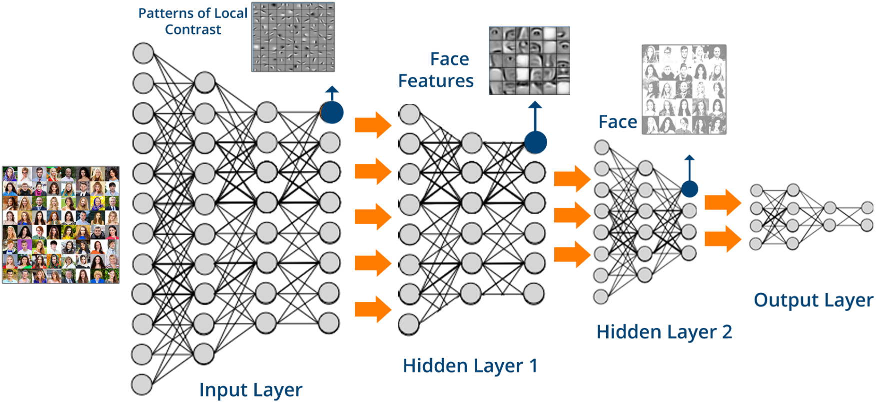
The takeaway

- Not attempting to settle "hard work vs. talent", IQ, etc. As usual, more subtle
- What it means to systematically get better
- Surprising aspects
    - Tangible reasons why people are "good"
    - Development of expertise: minimums
    - Enormous capacity for improvement
- 10k hours not quite correct

Machine learning basics

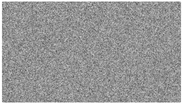
- Three components to any ML problem: the **task**, the **performance measure** and the **data**
- Essential definitions
  - **Features**
  - **Model**
  - **Parameters**
  - **Loss**

Deep learning



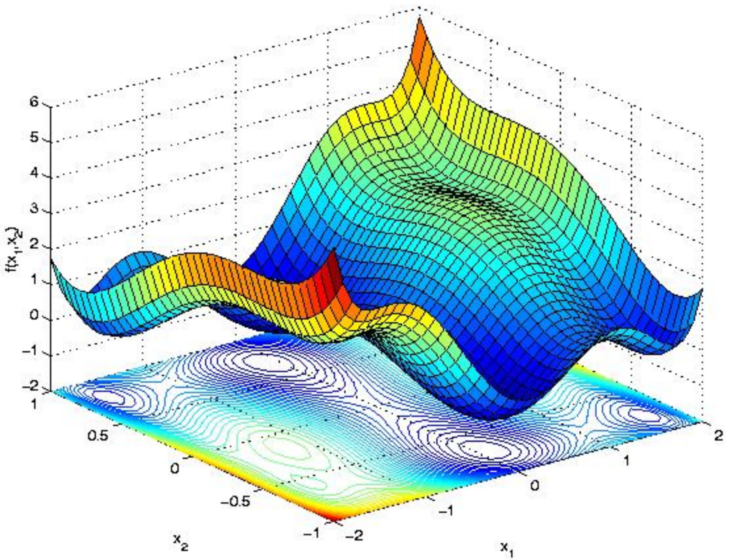
Why deep learning?

- Probably not
  - Universal approximation theorem
  - No free lunch theorem



- Probably
  - Feature learning
  - Hierarchical representations

Optimization of a nonconvex loss function



Connections to human learning

- Same sensory inputs ⇒ different tasks ⇒ different representations[6]
  - Chunking, "numbers to leave numbers": extracting relevant information, process nonlinearly, arrive at simple chunks
- Feature extraction is hierarchical
- Deliberate practice as nonconvex optimization
  - Seeking out strong gradients
  - "Pretrain", understand optimization landscape through experience
  - Local minima
  - Surprising solutions and flexibility with consistent optimization: OpenAI blog post

[6] yixinlin.net/writings//2016/10/18/representations.html

Takeaways

- Humans learn expertise from deliberate practice
- Basics of machine learning
- Shared insights: learning is *one concept* carried out by biology or silicon