# Lecture 21 Lowerbounds for Neural Networks

Sunday, November 13, 2016 3:34 PM

- Recall: neural net conjecture

  All functions we want to learn in practice can be represented by a small neural network.

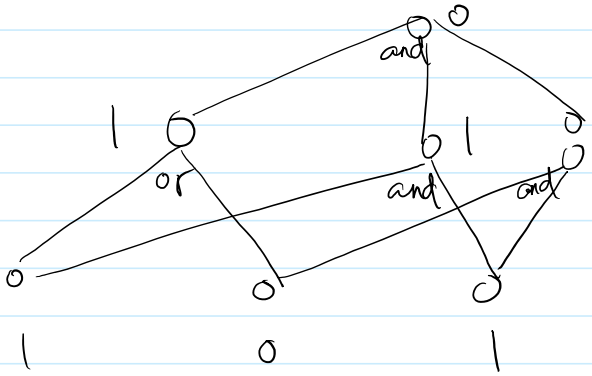
- last lecture: two-layer neural net already sufficient for many functions.

- Q: Are there natural functions that cannot be represented / approximated by a neural net?

  (also, what is "natural"?)

- Old results: circuit lowerbounds.

  - circuit: a model of computation. maps boolean inputs to boolean outputs via logical gates




  - circuit lowerbound: show a function $f$ cannot be computed by a certain class of circuits.

  - Typical circuit classes:

    $AC^0$: and / or / not gates, unbounded fan-in (# input), constant layer

    $TC^0$: threshold gates ($y = \mathrm{sgn}(w^T x + b)$), unbounded fan-in, constant layer.

$TC^0$: threshold gates ($y = \text{sgn}(w^T x + b)$), unbounded fan-in, constant layer.

sign function $= \begin{cases} 1 & w^T x + b > 0 \\ 0 & \text{o.w.} \end{cases}$

$TC^0$ is more similar to neural nets (still not the same)

- What's known

[Hastad] The XOR function $f(x) = X_1 \oplus X_2 + \cdots + X_d$ cannot be represented by an $AC^0$ circuit of polynomial size.

Problem: 1. XOR is trivially in $TC^0$/neural net
2. Is XOR natural?

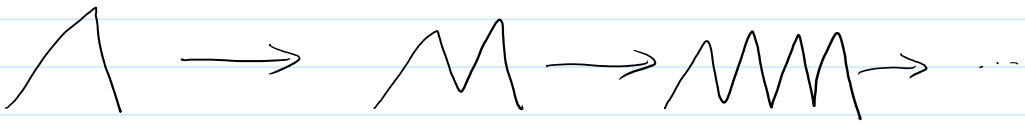
- Why is circuit lowerbound so hard?
  - in circuit lowerbound, we need an "explicit" function
  - "explicit" means deterministic, computable somewhat efficiently.
  - If we don't care, can use simple counting argument.

# function from $d$ inputs to 0/1: $2^{2^d}$ (huge!)

# possible $TC^0$ circuits of size $s$: $2^{\text{poly}(d,s)}$

  - so a random function cannot be represented.
  - but a random function is also not "natural".
- Current status of circuit lowerbound
  - Have some lowerbound for $ACC^0$ (2014) (and/or/not/mod 2/mod 3 gates)
  - no idea on $TC^0$.
- New attempts: use property of real-valued functions
  - 3 recent attempts from COLT this year.
  - [Telgarski] There is a neural network with $k^3$ layers, $\Theta(1)$ nodes per layer that cannot be approximated with a small, $k$-layer neural network.

Idea: more layers $\Rightarrow$ more alternations in 1 direction



cannot hope to create more alternations using a "shallow" neural net.

Q: Is a function with $2^{k^3}$ alternations natural?

- [Eldan, Shamir] There are simple functions representable by small 3 layer neural network, but cannot be approximated by any small 2-layer network.

Idea: In three layers can implement $f(x) = g(\|x\|)$ for any nonlinear function $g$.

Using tools from Fourier Analysis, can show this kind of symmetry is impossible for small 2 layer network.

- [Cohen, Sharir, Shashua] characterize the size of neural network by tensor rank.

Show a random deep network not representable by shallower networks

random (which is simpler), but random over a restricted set, so not implied by counting argument.