Multilayer perceptrons

Sebastian Seung
Layered networks

- Two layers of LT neurons
  - (three layers if input neurons are included)
- Two layers of synapses.
- No loops
Two questions about representational power

• What can an MLP compute?
  – any boolean function

• What can an MLP compute efficiently?
  – a vanishingly small fraction of boolean functions
Claim:
Any Boolean function can be computed by a perceptron with two layers of synapses.
Disjunctive normal form (DNF)

• Any boolean function can be written in disjunctive normal form.
• Disjunction of conjunctions
DNF construction

\[
\begin{array}{cccc}
  x_1 & x_2 & x_3 & f(x_1, x_2, x_3) \\
  0 & 0 & 0 & 0 \\
  0 & 0 & 1 & 0 \\
  0 & 1 & 0 & 0 \\
  0 & 1 & 1 & 0 \\
  1 & 0 & 0 & 0 \\
  1 & 0 & 1 & 1 \quad \rightarrow \quad x_1 \land \bar{x}_2 \land x_3 \\
  1 & 1 & 0 & 1 \quad \rightarrow \quad x_1 \land x_2 \land \bar{x}_3 \\
  1 & 1 & 1 & 1 \quad \rightarrow \quad x_1 \land x_2 \land x_3
\end{array}
\]

\[
f = (x_1 \land \bar{x}_2 \land x_3) \lor (x_1 \land x_2 \land \bar{x}_3) \lor (x_1 \land x_2 \land x_3)
\]

\[
= x_1 \land (x_2 \lor x_3)
\]
Any DNF can be written as a 2-layer perceptron

• AND of variables or their negations
  – LT neuron with excitatory and inhibitory synapses.

• OR
  – LT neuron with excitatory synapses and low threshold.
What’s the catch?

- The number of conjunctions required may be exponentially large.
- I.e., there is no guarantee that the perceptron representation is efficient.
Efficiency

• Number of synapses
• Serial computer
  – Time
• Parallel computer
  – Space
  – Energy
Claim:
Most boolean functions require a perceptron of exponential size.
How many boolean functions can be realized by an MLP with $S$ synapses?

- How many MLPs are there?

- How many functions can be realized by each MLP?
Parity function

• generalization of XOR to \( N \) variables
• addition modulo 2

\[
\Pi(x_1, \ldots, x_N) = \begin{cases} 
1, & \text{if } \sum_{i=1}^{N} x_i \text{ is odd} \\
0, & \text{otherwise}
\end{cases}
\]
Which 2nd layer weights implement parity?

\[ \Pi(x_1, x_2, x_3, x_4) \]

- A: \(-1, 1, -1, 1\)
- B: \(1, -1, 1, -1\)
- C: \(1, 2, 3, 4\)
- D: \(1, 2, 4, 8\)
Parity as an MLP with 2 layers

- $O(N)$ neurons
- $O(N^2)$ synapses
Parity as an MLP with log N layers

- $O(N)$ neurons
- $O(N)$ synapses
Depth-size tradeoff

• A deeper network may be able to realize a computation with fewer neurons or synapses.
Circuit complexity theory

• What can circuits of logic gates compute?
• How large a circuit is required?
• Branch of computational complexity theory.
Nonlinearity is essential for computational power

- Increasing the depth of a perceptron makes no difference if the activation function is linear.
Perceptrons can also be used to compute analog functions. For example, use a sigmoidal activation function.
Any smooth function can be approximated by an MLP with two layers of synapses and a sigmoidal activation function.
Connectedness
Minsky-Papert definition

- Requires only that the final step in the computation is an LT neuron.

\[ H \left( \sum_{a} w_a \varphi_a (x) \right) \]
Diameter-limited perceptrons

- Assume that the input vector is organized as a 2d image.
- Each $\varphi$ depends on a set of pixels with a limited diameter.
Theorem

No diameter-limited perceptron can compute connectedness.
Proof

• Reduce to XOR