Memorization vs. generalization

Sebastian Seung

How many boolean functions of N variables?

2^N $\quad N2^N$ $\quad 2^{2^N}$ $\quad 2^{N^2}$

A B C D

Every boolean function corresponds to a truth table

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
1	1	0	1
1	1	1	1

How many boolean functions are realized by an LT neuron?

2^N $\quad$ $N2^N$ $\quad$ 2^{2^N} $\quad$ 2^{N^2}

A $\quad$ B $\quad$ C $\quad$ D

Few boolean functions can be computed by an LT neuron

(To be proved later)

Number of dichotomies

- Given m input vectors in d dimensions
- Dichotomy = mapping of the vectors to binary values {0,1}

- How many dichotomies can be realized by a homogeneous LT neuron?
  - This number will be called $C(m,d)$

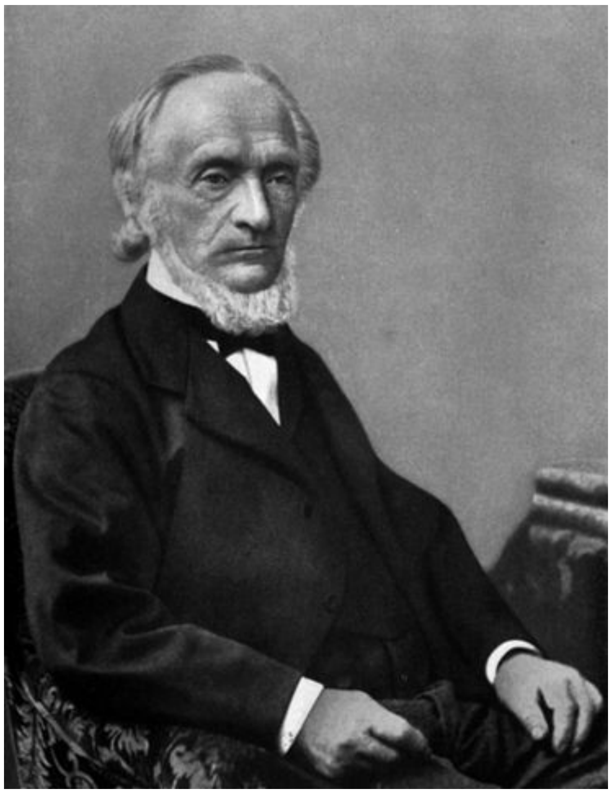
View from input space (d=2)

- Decision boundary is line through origin.
- Rotate line to generate all possible dichotomies.
- m=1
- m=2
- Exclude linearly dependent sets of input vectors

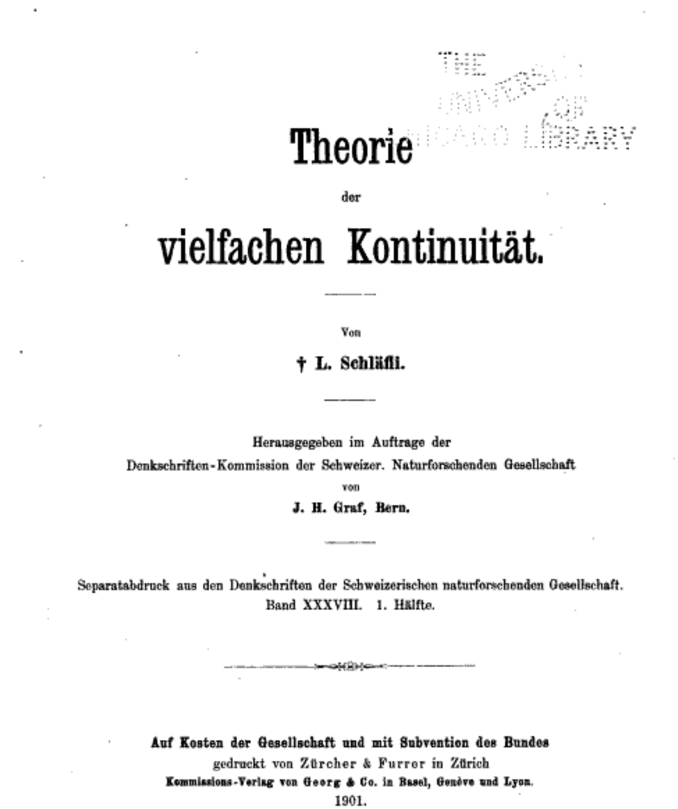
View from weight space (d=2)

- An input vector divides the space of weight vectors.
- How many regions are created by m lines through the origin?
- This yields $C(m,2)$

Ludwig Schläfli



1814-95



Theorie

der

vielfachen Kontinuität.

Von

† L. Schläfli.

Herausgegeben im Auftrage der
Denkschriften-Kommission der Schweizer. Naturforschenden Gesellschaft
von
J. H. Graf, Bern.

Separatabdruck aus den Denkschriften der Schweizerischen naturforschenden Gesellschaft.
Band XXXVIII. 1. Hälfte.

Auf Kosten der Gesellschaft und mit Subvention des Bundes
gedruckt von Zürcher & Furrer in Zürich
Kommissions-Verlag von Georg & Co. in Basel, Genève und Lyon.
1901.

Schläfli's formula

- Draw *m* hyperplanes through the origin of *d* dimensional space
- How many regions are created?

$$C(m,d) = 2\sum_{i=0}^{d-1}\binom{m-1}{i}$$

General position

- Set of *m* points in *d* dimensions
- All subsets of *d* or fewer points are linearly independent

- Schläfli's formula does not depend on properties of the hyperplanes, provided they are in general position.

Proof sketch

- Consider all dichotomies on first *m*–1 examples
- Some extend uniquely to *m*th input.
  - If this were true for all dichotomies, we would have $C(m, d) = C(m-1, d)$
- But some dichotomies are ambiguous
  - How many?
  - This undercounting must be corrected.

Number of ambiguities

- Ambiguity if there exists **w** consistent with first m–1 examples s.t. $\mathbf{w} \cdot \mathbf{x}_m = 0$.
  - Number of ambiguous dichotomies given by the number of dichotomies on the d–1 dimensional space orthogonal to $\mathbf{x}_\mathrm{m}$, or $C(m-1,d-1)$
  - Correcting for undercounting leads to recursion relation

$$C(m,d) = C(m-1,d) + C(m-1,d-1)$$

Exponential vs. polynomial

$$C(m,d) \begin{cases} = 2^m, & m \leq d \\ \leq m^d, & m \geq d \end{cases}$$

Sharp transition

- In the limit of infinite *m* and *d*, with the ratio *m*/*d* held fixed

$$\frac{C(m,d)}{2^m} \to \begin{cases} 1, & m/d < 2, \\ 0, & m/d > 2 \end{cases}$$

- An LT neuron can store 2*d* random examples

Storage capacity

- The LT neuron can memorize $m=2d$ patterns
  - input vectors in general position
  - desired outputs chosen randomly

Number of boolean functions

- Vertices of the boolean cube $\{0,1\}^d$ are not in general position
- Schläfli's formula becomes an upper bound, and in turn is bounded by

$$C(2^d, d) \leq 2^{d^2}$$

- Lower bound turns out to be similar.

Memorization and generalization

- Suppose an LT neuron has successfully memorized *m* examples
- What is the probability that it will correctly predict the label of a novel input?

Prediction game

- Given examples $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)$ and new input $\mathbf{x}_{m+1}$
- Your opponent
  - Generates candidate y_{m+1} by fair coin toss
  - Accepts if there is an LT neuron consistent with all examples
- You predict by the same procedure

Probability of ambiguity

$$A(m,d) = \frac{C(m,d-1)}{C(m,d)}$$

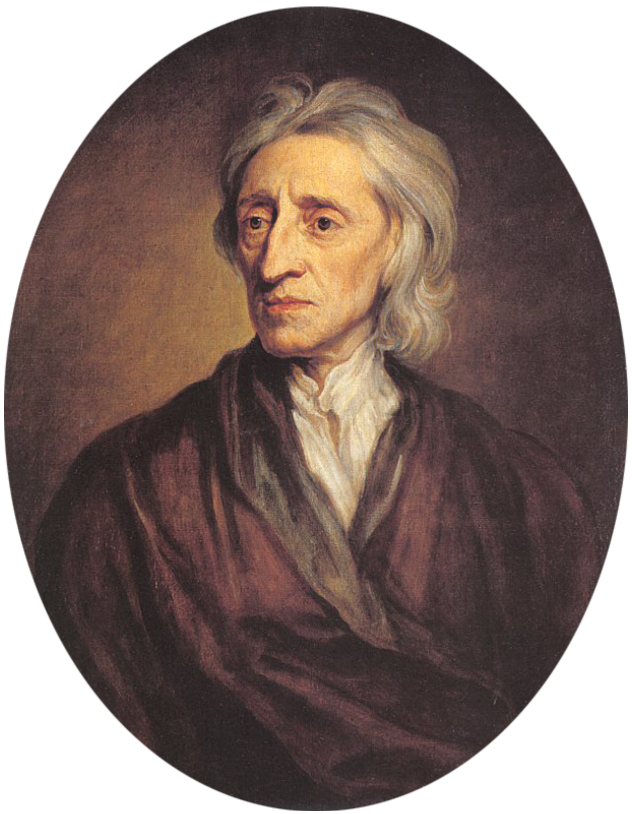
- In the limit of infinite *m* and *d*, with the ratio *m*/*d* held fixed

$$A(m,d) \to \begin{cases} 1, & m/d \le 2, \\ \left(\frac{m}{d} - 1\right)^{-1}, & m/d \ge 2 \end{cases}$$

Limited capacity: strength and weakness

- Generalization starts when perfect memorization ends.
- An LT neuron has limited power to represent functions
- This limitation is essential for generalization to novel inputs

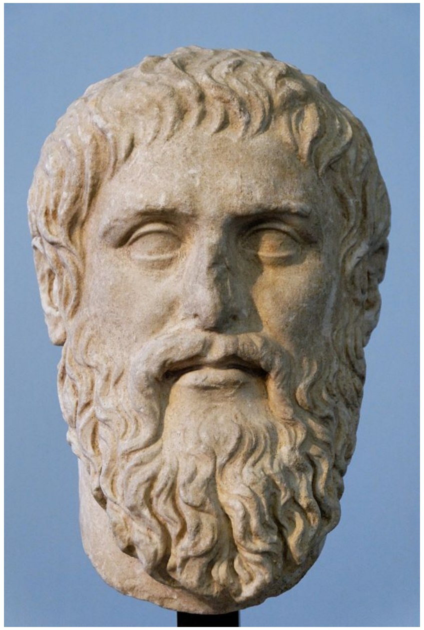
John Locke



1632-1704

Let us then suppose the mind to be, as we say, white paper, void of all characters, without any ideas: — How comes it to be furnished? … To this I answer, in one word, from EXPERIENCE.

Plato



428-348 BC

the soul, since it is immortal and has been born many times, and has seen all things both here and in the other world, has learned everything that is…seeking and learning are in fact nothing but recollection.