



# Experimental Design for Machine Learning on Multimedia Data

## Lecture 4

**Dr. Gerald Friedland,**  
**[fractor@eecs.berkeley.edu](mailto:fractor@eecs.berkeley.edu)**

Website: <http://www.icsi.berkeley.edu/~fractor/fall2020/>

# Discuss Homework

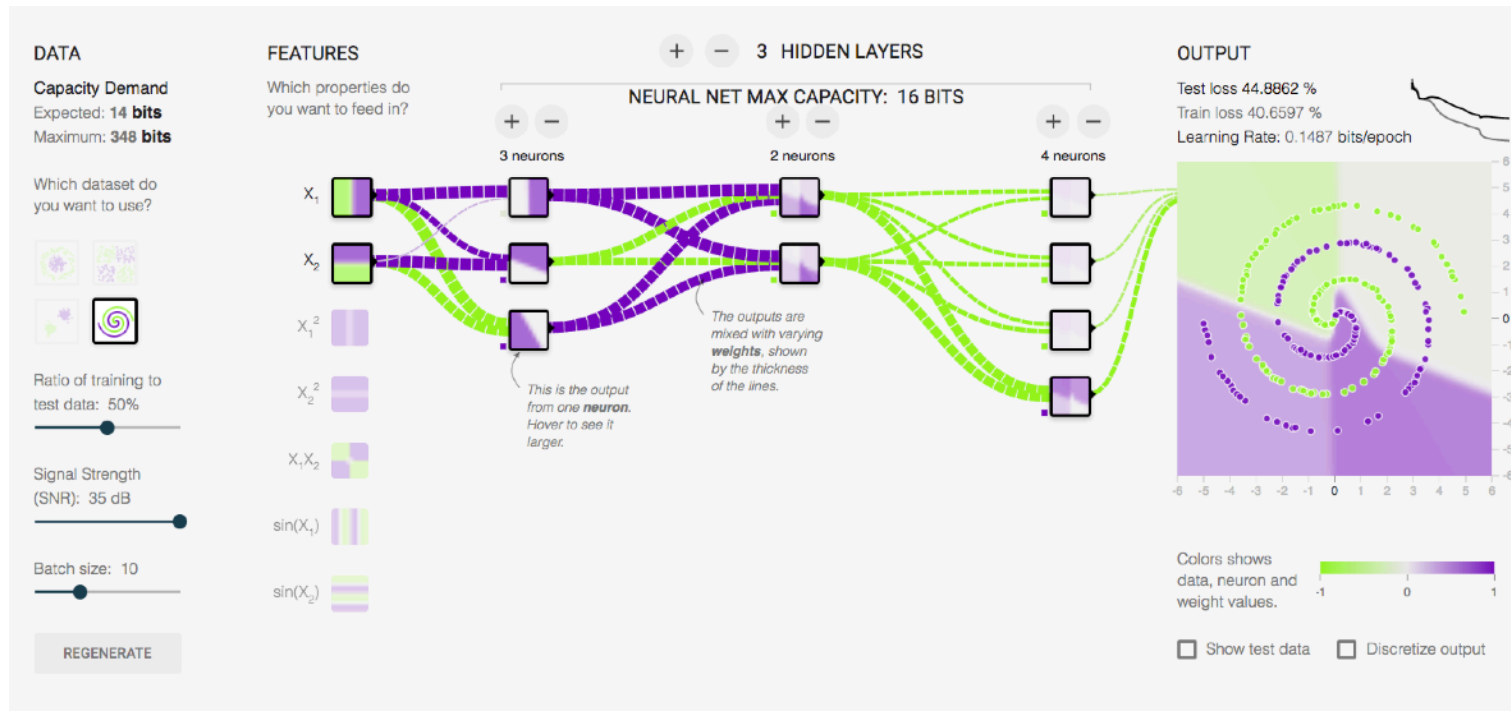
- Project:
- Please start forming teams.
- See Project description on Piazza

Homework:

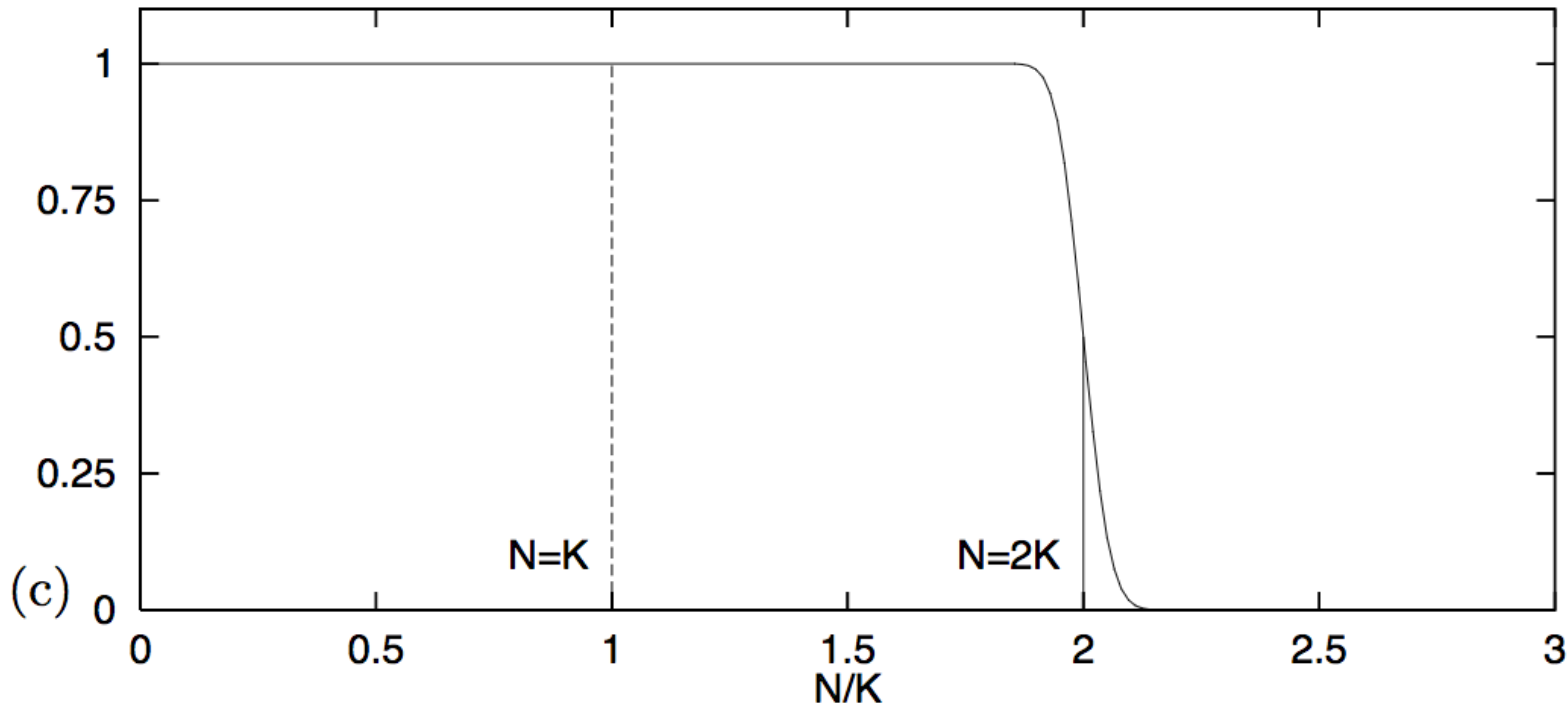
- Questions?

# Demo Time!

- Capacity for Neural Networks explained: See also cheat sheet.
- Practical applications
- Demo



# Critical Points: Perceptron (Cover, MacKay)



$N=K$ : VC Dimension (for points in random position)

$N=2K$ : Cover/MacKay Capacity

# Generalizing from Perceptron to Perceptron Networks

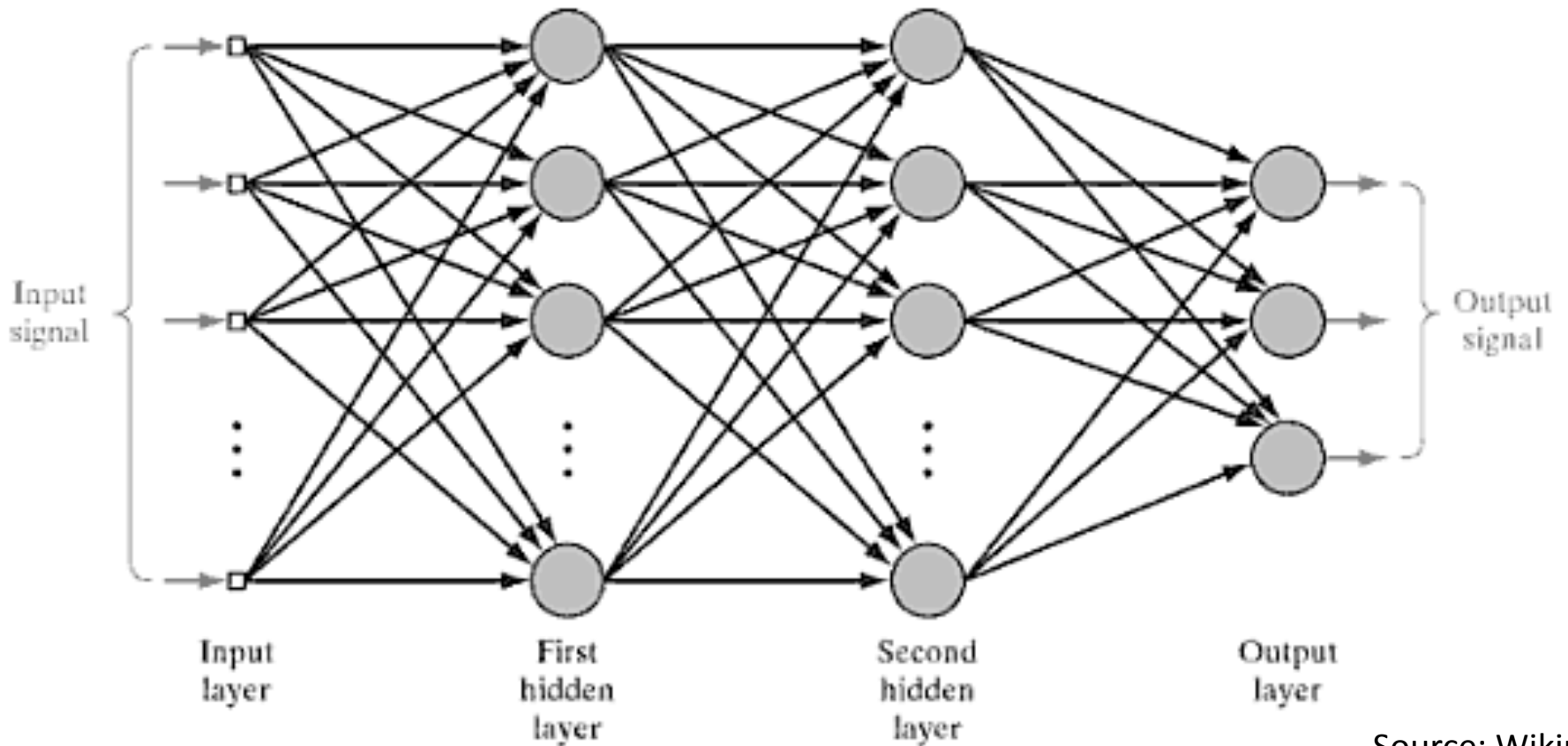
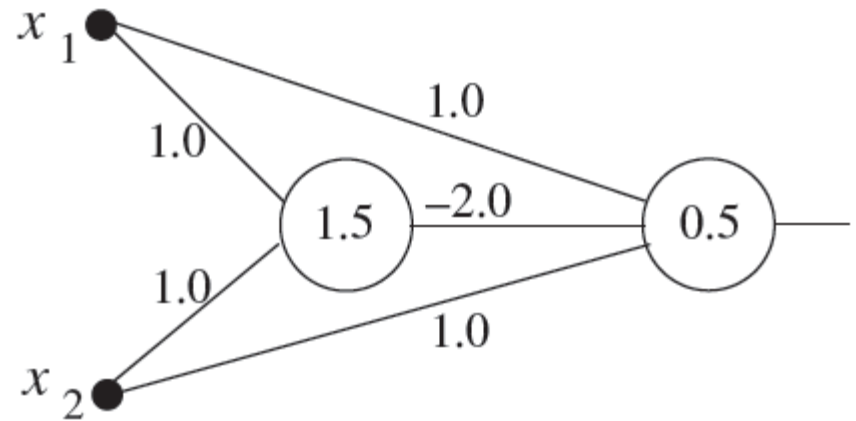
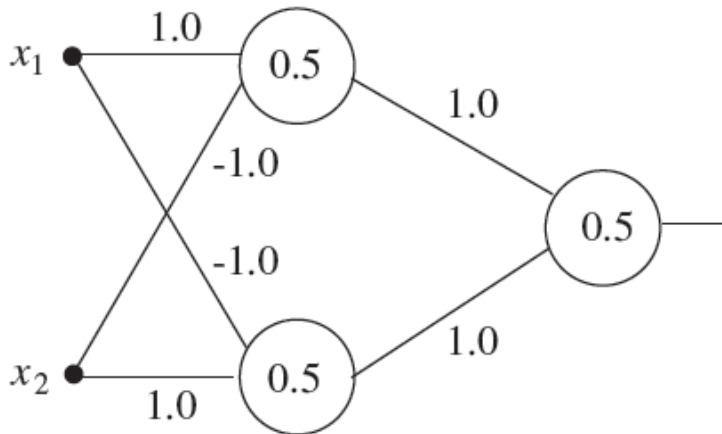


FIGURE 4.1 Architectural graph of a multilayer perceptron with two hidden layers.

Source: Wikipedia

# Careful: Other Architectures



## Example Solutions to XOR

Source: R. Rojas, Intro to Neural Networks

# Best Case Scenario?

**Just measure in bits!**

The *memory* capacity of **any binary classifier** cannot be better than the number of relevant bits in the model (pigeon hole principle, no universal lossless compression).

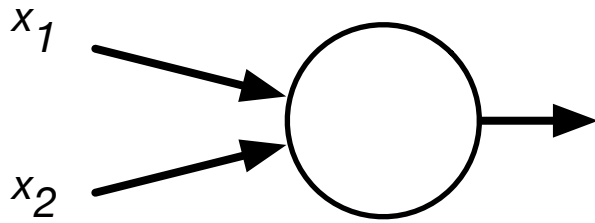
This is:  $n$  bits in the model can *maximally* model  $n$  bits of data.

# Engineering Principles for Neural Networks

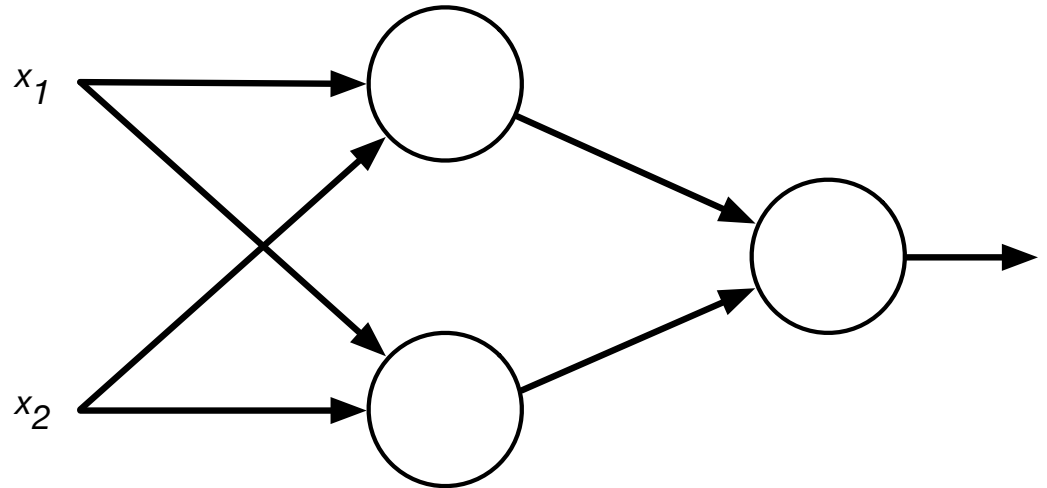
- 1) The output of a perceptron is maximally 1 bit.
- 2) The maximum memory capacity of a perceptron is the number of parameters (including bias) in bits.  
*(MacKay 2003)*
- 3) The maximum memory capacity of perceptrons in parallel is additive.  
*(MacKay 2003 speculative, Friedland and Krell 2017)*
- 4) The maximum memory capacity of a layer of perceptrons depending on a previous layer of perceptrons is limited by the maximum output (in bits) of the previous layer.  
*(Data Processing Inequality, Tishby 2012)*



# Examples: How many bits of maximal capacity?

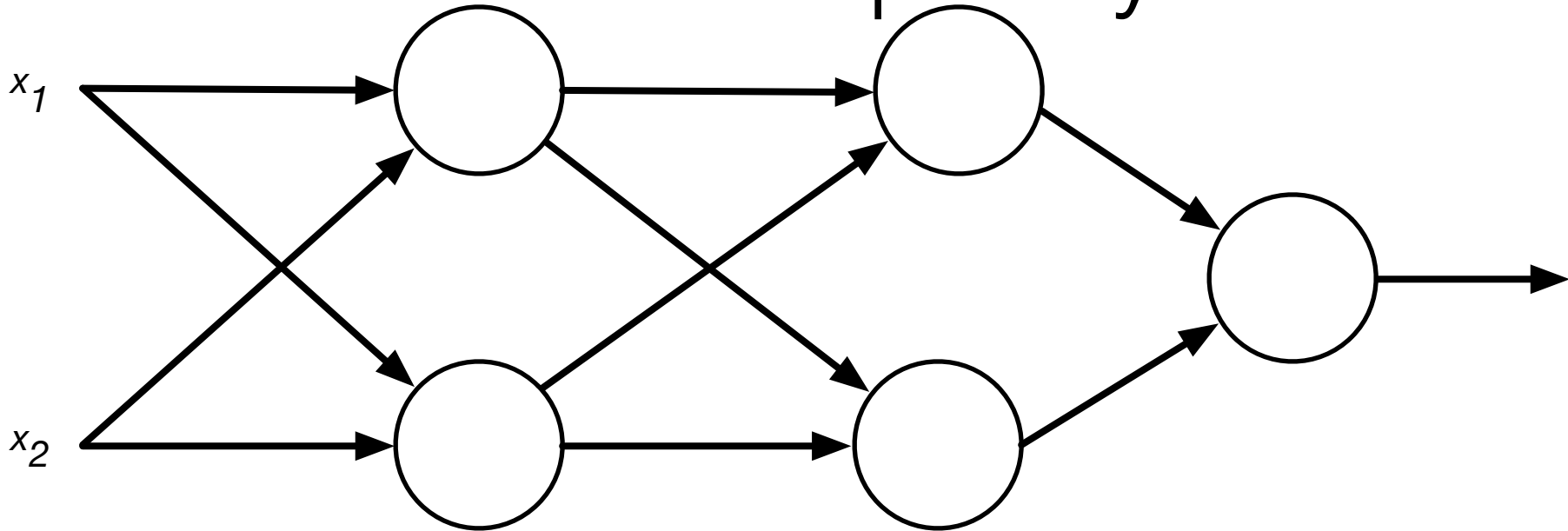


3 bits



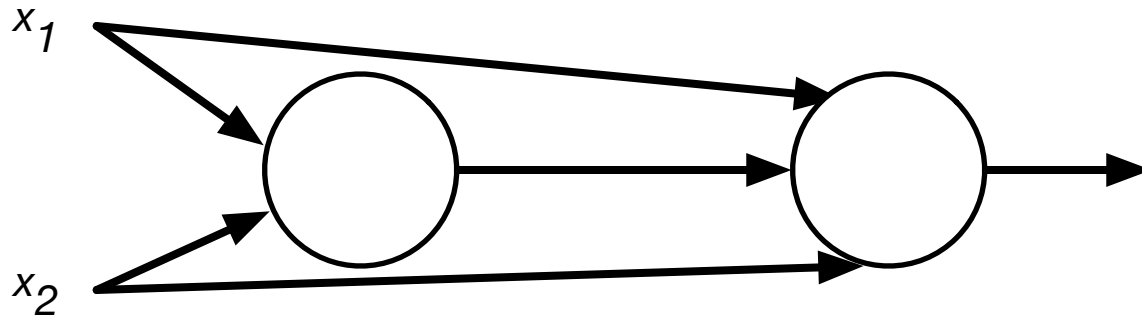
$2 * 3$  bits +  $\min(2, 3)$  bits = 8 bits

Examples: How many bits of maximal capacity?



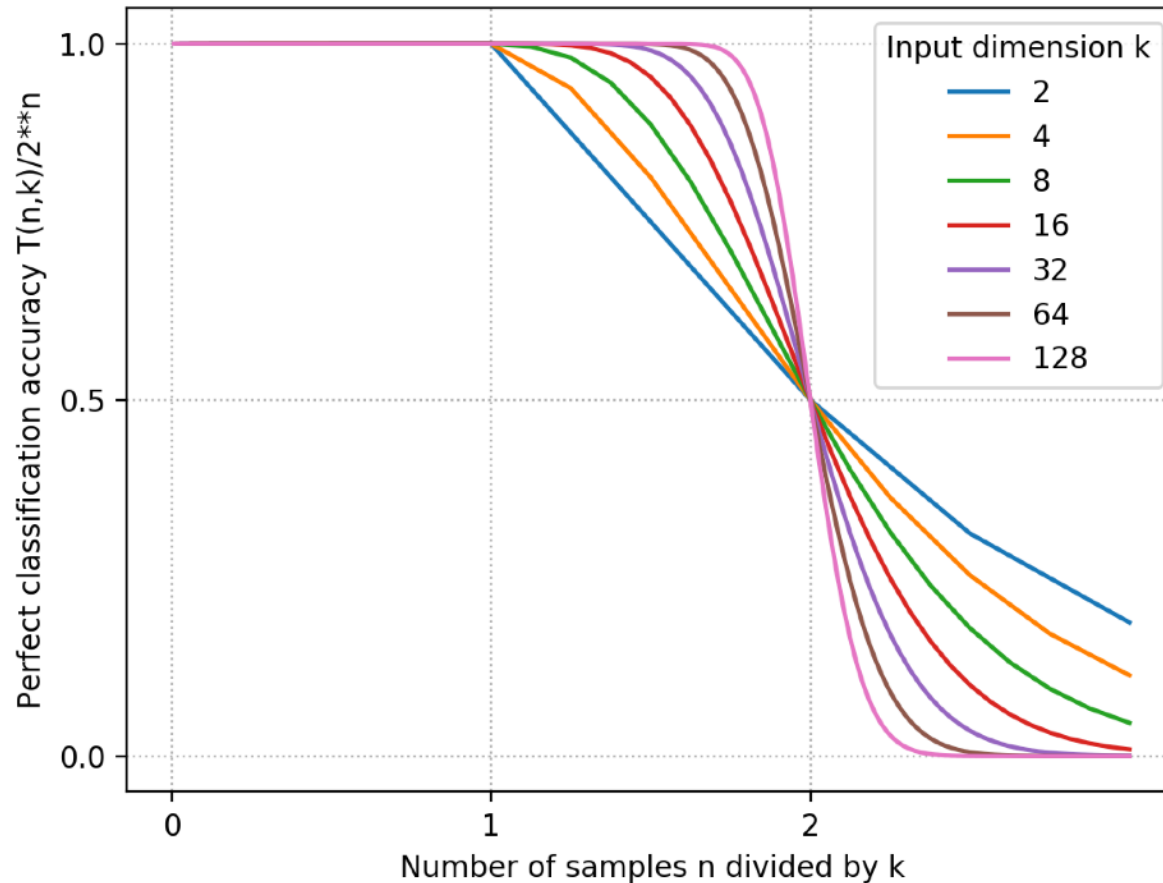
$$2 \cdot 3 \text{ bits} + \min(2, 2 \cdot 3) \text{ bits} + 2 \text{ bits} = 10 \text{ bits}$$

Examples: How many bits of maximal capacity?

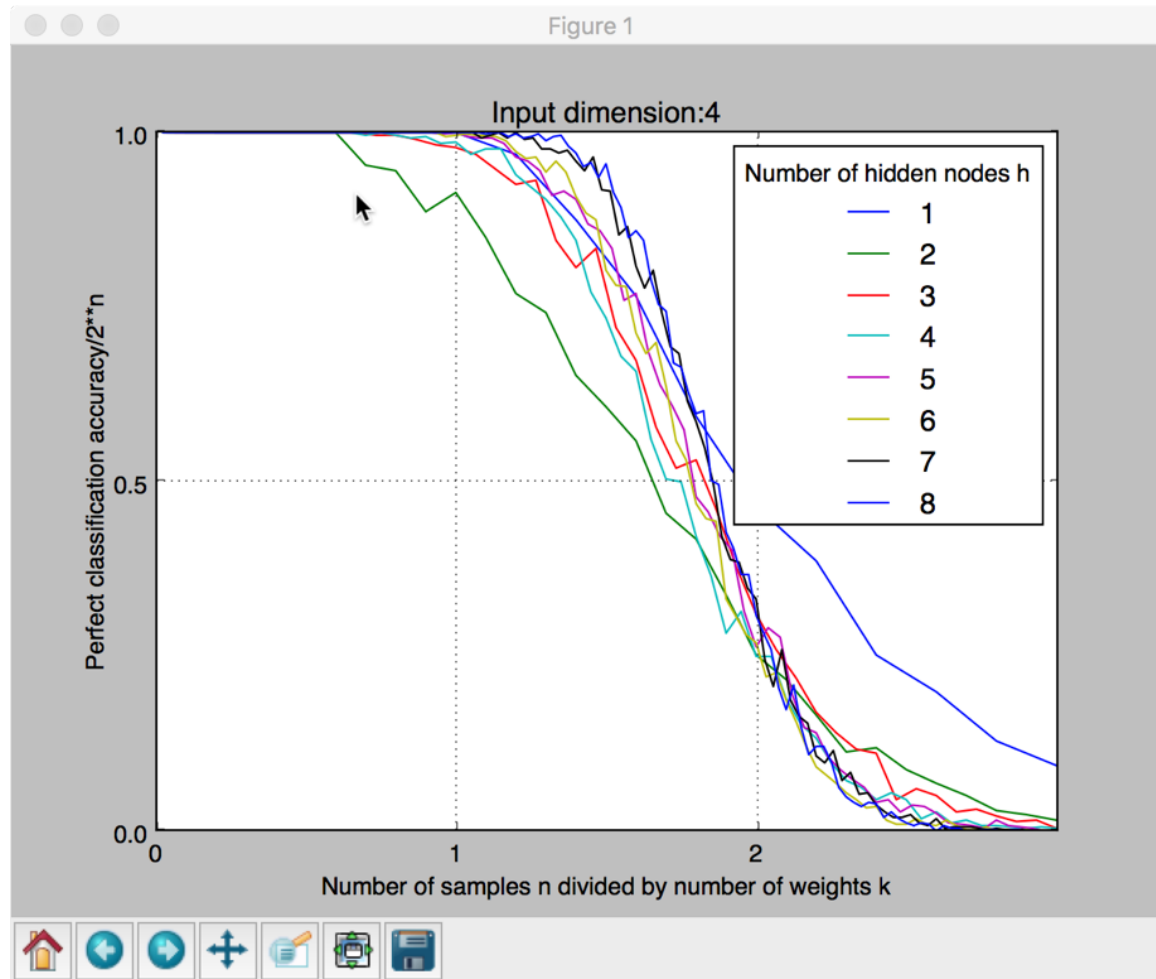


3 bits+4 bits=7 bits

# Characteristic Curve of a Theoretical 3-Layer MLP



# Characteristic Curve of an Actual 3-Layer MLP



# Conclusion

- The lower limit of generalization is memorization. This is, the upper limit for the size of a machine learner is it's memory capacity.
- The memory capacity is measurable in bits.
- Using a machine learner that is over capacity is a waste of resources and increases the risk of failure!
- Alchemy converted into chemistry by measuring: It's time to convert guessing and checking in Machine Learning into science! Let's call it data science?