Remember there were 2 activities that we wanted to do with Bayesian Learning

1. Compute the most probable hypothesis to explain data
2. Predict next data item

$\mathbf{P}(X \mid \mathbf{d}) = \Sigma \mathbf{P}(X \mid h1)\ \mathbf{P}(hi \mid \mathbf{d})$

We are considering all hypothesis when try to guess next candy

If instead choose most likely hypothesis then predict next candy using that alone using the maximum a posteriori hypothesis (MAP)

To choose most likely hypothesis we let hmap = hi s.t. P(hi | **d**) P(hi) is largest
P(hi) is the probability of having this heuristic

Then our prediction of next item is
P(X | hmap)

Maximizing P(D | hi) P(hi)
is equivalent to minizing

1 / (P (d | hi) P(hi))

Which is equivalent to minimizing

Log (1 / (P (d | hi) P(hi))) = -log P(d | hi)P(hi)

= # of bits to describe data **d** with hypothesis hi

+ # of bit describe hi

So MAP is equivalent to choosing shortest string to explain data (formalization of Ockham's Razor).

Maximum likelihood learning is MAP Hypothesis together with assumption that all the hi are equally likely.

Neural Networks

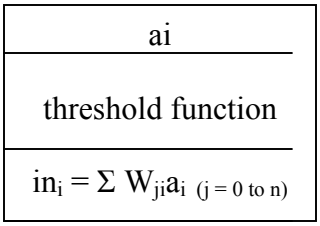
A neuron is the logic computational unit of the brain.

Neuron ->>>>> Brain

One way to try to get intelligent devices is to try to create networks of simulated versions of neurons in a computer.

How to simulate a neuron: edges to other neurons. Each neuron has an activation level ai

ai {-1, 1}
ai [-1, 1]



W_{ij} called bias weights

Threshold function: Usually one of (Step function) book calls it the threshold function.

or a sigmoid function

$f(x) = 1 / (1 + e^{-x})$