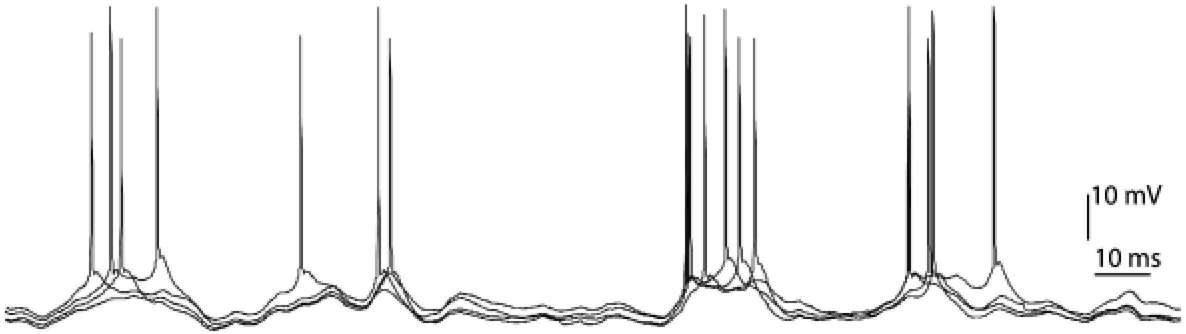
Mathematical Aspects of Neuronal Computation

Ioannis Smyrnakis

Neuron Structure

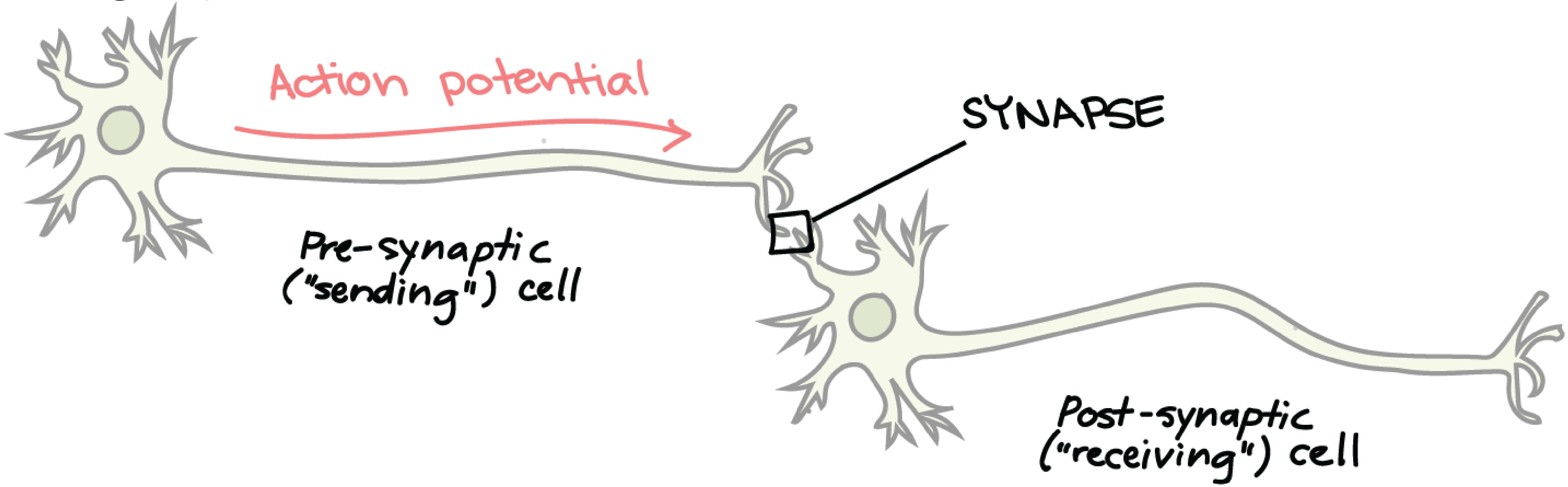
Neuron Activity

- Neuronal activity is through potential surges (spikes) when the underlining membrane potential exceeds a threshold.



Neuronal Communication

- These are transmitted from the axon to the dendrites of the receiving neuron through synapses that have varying conductivities (synaptic strengths).



Neuronal Types

- There are two basic types of neurons, the **<u>pyramidal neurons</u>** which form excitatory connections and the **<u>interneurons</u>** which form inhibitory connections
- There is basically one type of excitatory neurons, the pyramidal neurons. They form long range connections as well as short range ones
- There are many types of interneurons. The precise function of each type of interneuron is not known. Their connections are mostly local. They seem to exert control over the network

Integrate and Fire Model

- The receiving **neuron receives input from a few thousand pre-synaptic neurons.**

- This **input is integrated** by the receiving neuron and when the membrane potential increases a threshold a spike is fired.

- In general, there is also a **membrane potential leakage**, that eventually neutralizes slow input.

Response of Neurons to Stimulus

- **Neurons are not reliable responders**. Neurons respond stochastically to a stimulus. If a particular stimulus is presented to a neuronal receptive field, a neuron may respond or may not respond.
- **However, animals are reliable responders.** When a clear stimulus is presented to an animal, the animal respond consistently.
- **Information Pathways**. This means that there is some kind of integration of many stochastically responding neurons towards a reliable response. **The neurons whose unreliable response is integrated towards a reliable response are said to form an information pathway.**

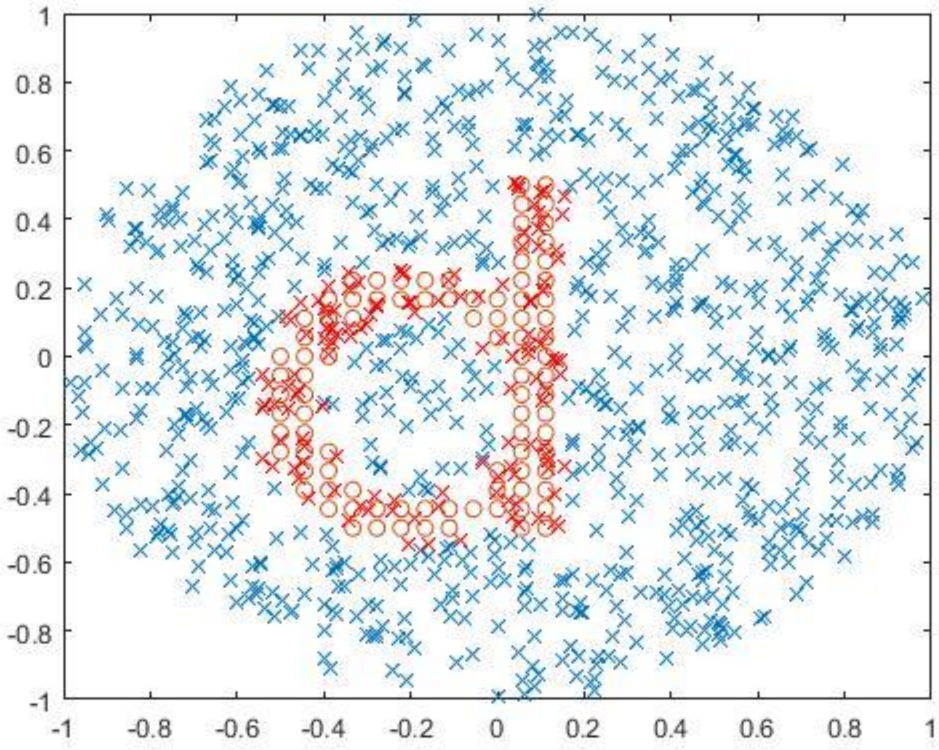
Information Pathways

- Information pathways appear right from the visual input in the retina.

  - Take for example the letter d presentation on the retina. It excites a number of retinal ganglion cells.

  - The information about letter d is contained in the activity of **all** the excited ganglion cells.

  - **The group of excited ganglion cells forms** the **information pathway** of the letter d presented.



Universality of information pathways

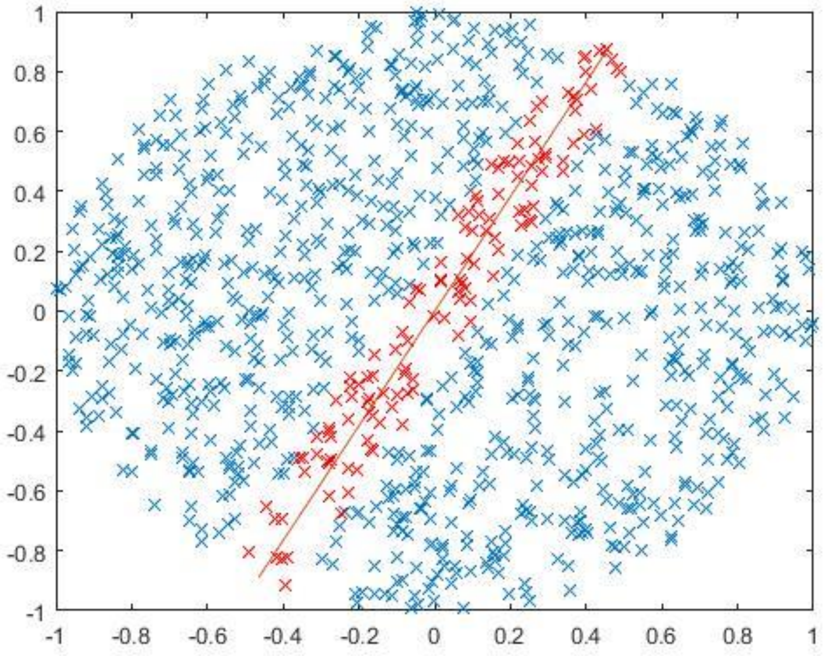
- One of course could imagine that the appearance of information pathways is only at the input layer, however this is unlikely to be true.
- Suppose that one pathway feeds one reliably responding neuron, that responds to a particular letter (letter d).
- First of all, such neurons would have been detected by now, and this has not happened.
- Second, if we have to recognize the word ‘diary’ letter by letter, then we still need all the reliable neurons responding to the letters of the word, and these would form an information pathway.

Question

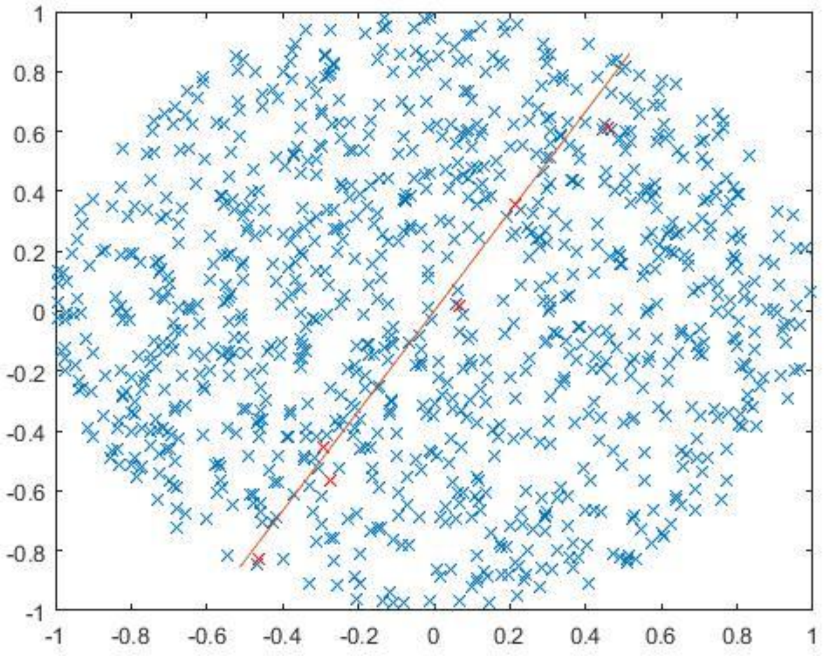
- Can you suggest benefits of stochasticity in neuronal response?

Economy through Stochasticity

- Stochasticity of neuronal response is **<u>necessary for economy</u>** in neuronal activity. To define a line we need much less activity than in the left panel.



Definite Response within distance 0.1



0.1 Prob. of Response within distance 0.1

Stability through Stochasticity

Stochasticity can help **maintain stability** of the neuronal network without loosing information.

- Too much activity →Too much input →Even more activity

- Too little activity →Too little input →Universal silence

Robustness through Stochasticity

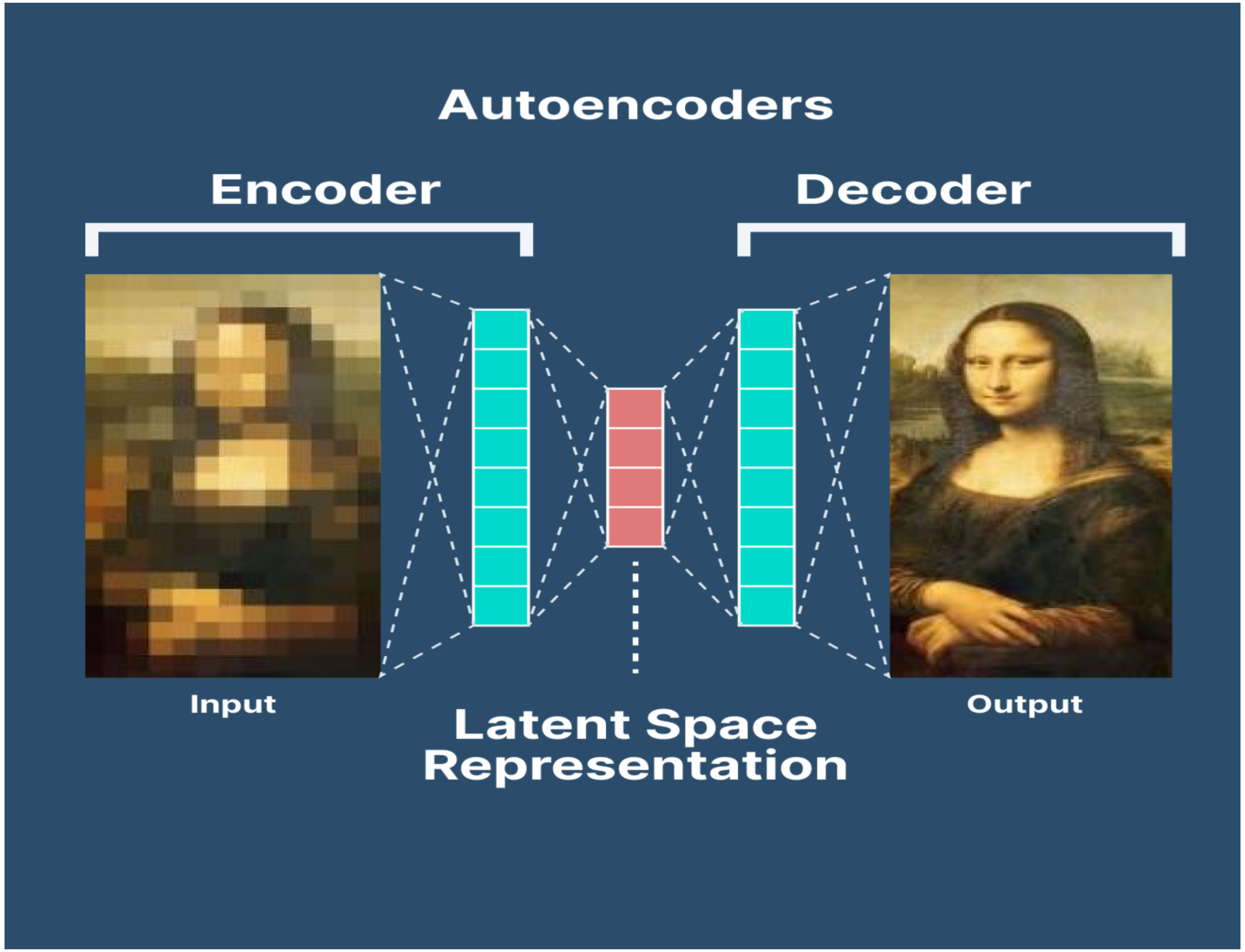
Stochasticity permits **greater robustness of the neuronal network**.

- Suppose for example that a neuron in the network dies and this neuron is a definite response neuron. Then the information it encodes is lost, unless there are more neurons responding to the same information. However these neurons may not be connected the right way along the line.
- However **if the neuron that died is only part of a pathway with a small probability of response, then the network will continue to operate** without any changes unless the pathway is in a limiting state.

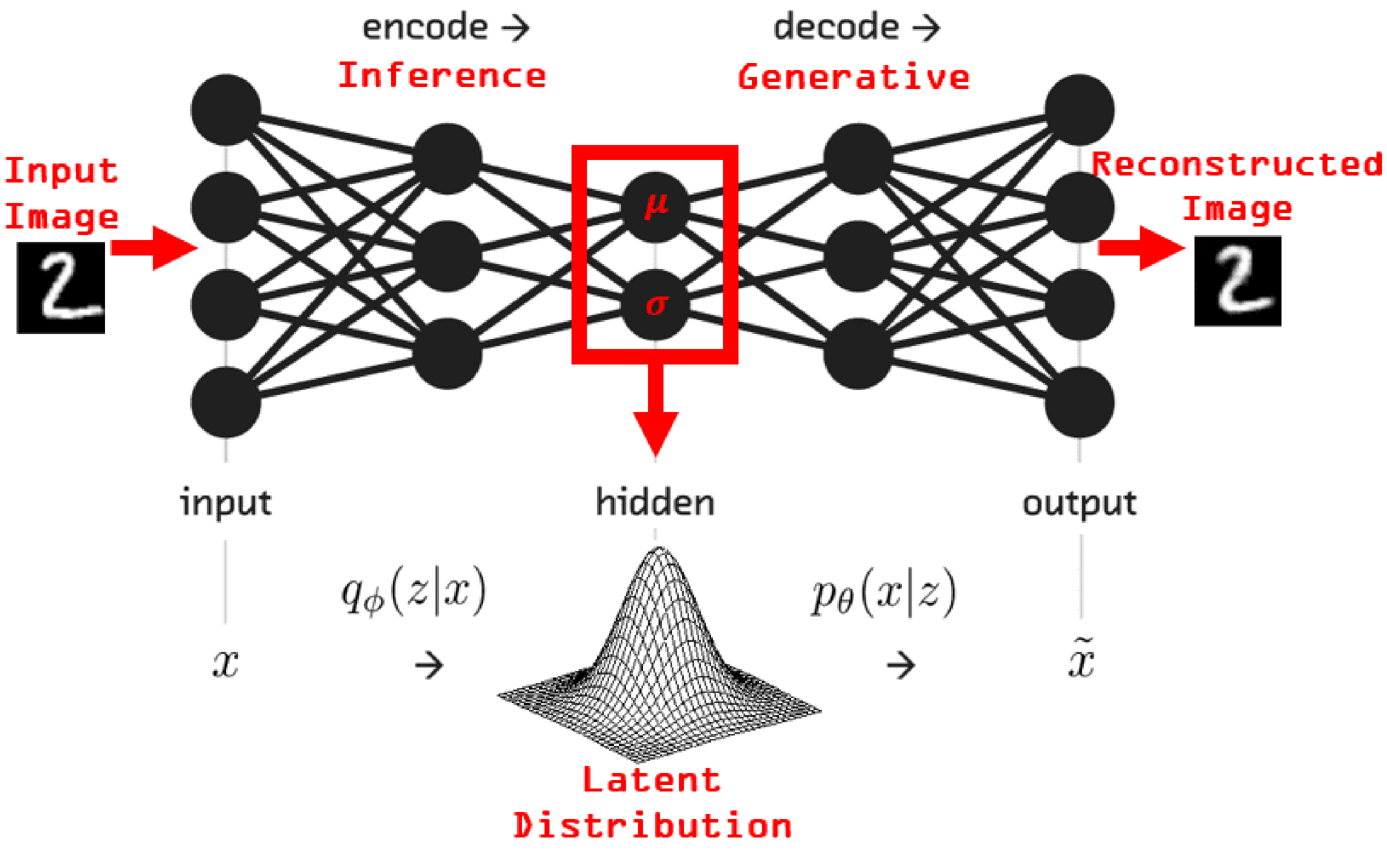
Acyclic Behavior through Stochasticity

- Infinite loops is a problem any programmer faces when programming.
- The brain is a self-programmed module. **<u>How does the brain avoid infinite loops?</u>**
- One way to do so is through stochasticity. **<u>A random perturbation brings the system out of the infinite loop</u>**

Autoencoders



Variational Autoencoders



Imagination through Stochasticity

Information Flow Control through Stochasticity

- Suppose that an animal is presented with an image (possibly a natural image) that has many objects. The information input is enormous, hence we have an **information bottleneck**
- Somehow there should be an information reduction process, while keeping functionality at the same time
- A **controllable network noise level** can keep only the most prominent pathways active above noise level, controlling in this way the amount of information that passes through the information bottleneck

The Information Capacity Question

Suppose that definite response to a signal is encoded in the activity of a pathway of stochastically responding neurons, how many such pathways can an aggregate of N neurons support, so that

a) **<u>The pathways responds to their corresponding signals with probability close to 1</u>**,

b) **<u>The activity of one pathway does not interfere with the activity of another.</u>**

Simplifying Assumptions

- Neurons are **randomly placed** in the volume they occupy.

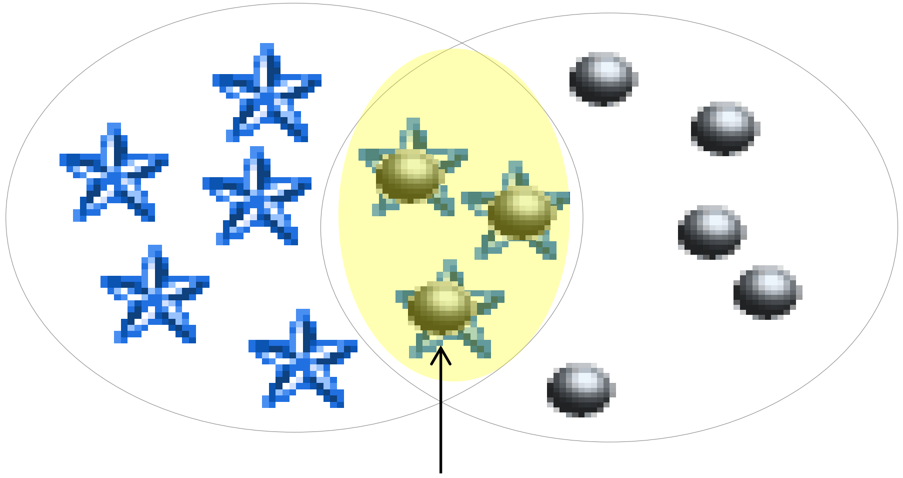
- Neurons respond with probability $\boldsymbol{p}$ to the signal that excites a pathway they belong to, while they respond with probability $\boldsymbol{p_0}$ to either no signal or to a signal that excites a pathway they do not belong to.

- There is a **pathway threshold $\boldsymbol{K}$**. If the pathway has more than K active neurons the pathway is considered to be active, otherwise it is inactive. This pathway threshold is chosen to ensure minimal interference of pathways.

Definite Response and No Interference conditions

- Definite response condition:
$$P(F_i > K_i|S_i) > 1 - \varepsilon$$

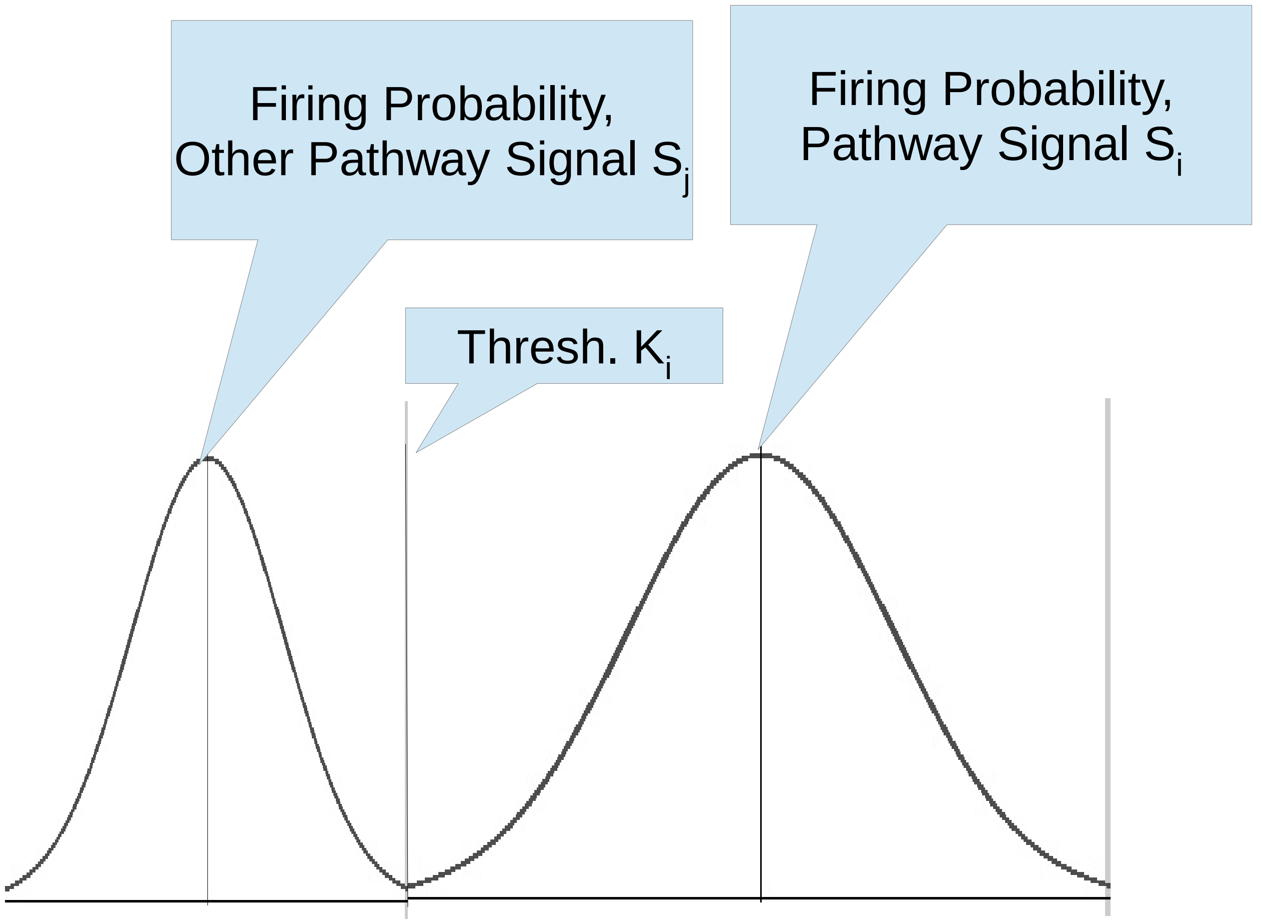
- No interference condition:
$$P\left(F_i > K_i|S_j\right) < \varepsilon$$



Overlap m

S_i: Signal of pathway i
F_i: Firing of pathway i
K_i: Threshold of pathway i
ε: Confidence limit

Optimal Choice of Threshold



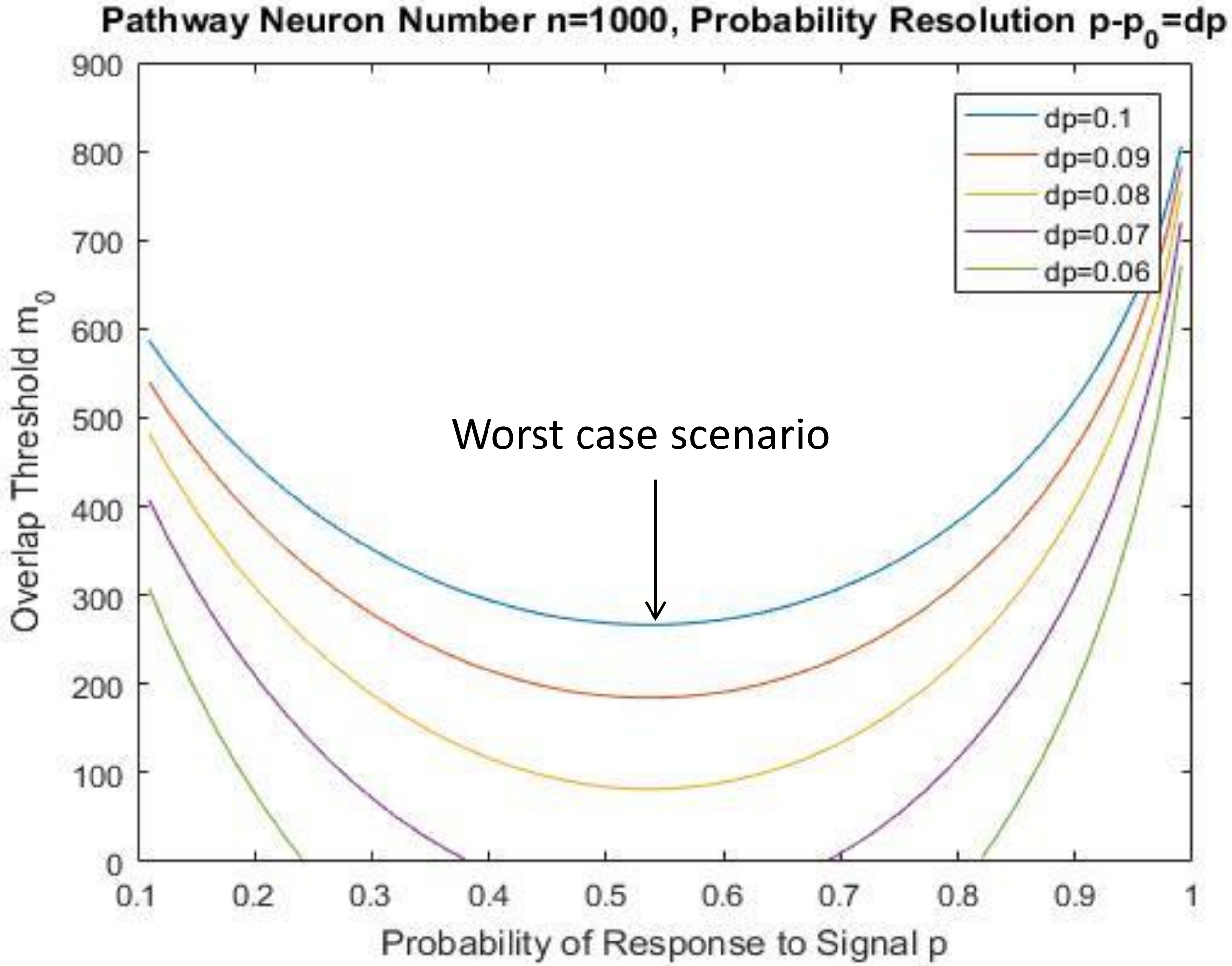
Optimal Threshold Outcome

- Under optimal choice of threshold, and for $\varepsilon = 0.01$, the definite response and the non-interference conditions collapse to the single condition

$$\frac{(n-m)(p-p_0)}{\sqrt{(n-m)p_0(1-p_0)+mp(1-p)}+\sqrt{np(1-p)}} > 2.33$$

- n: Number of neurons in the pathway
- m: Number of neurons in the overlap of two pathways.

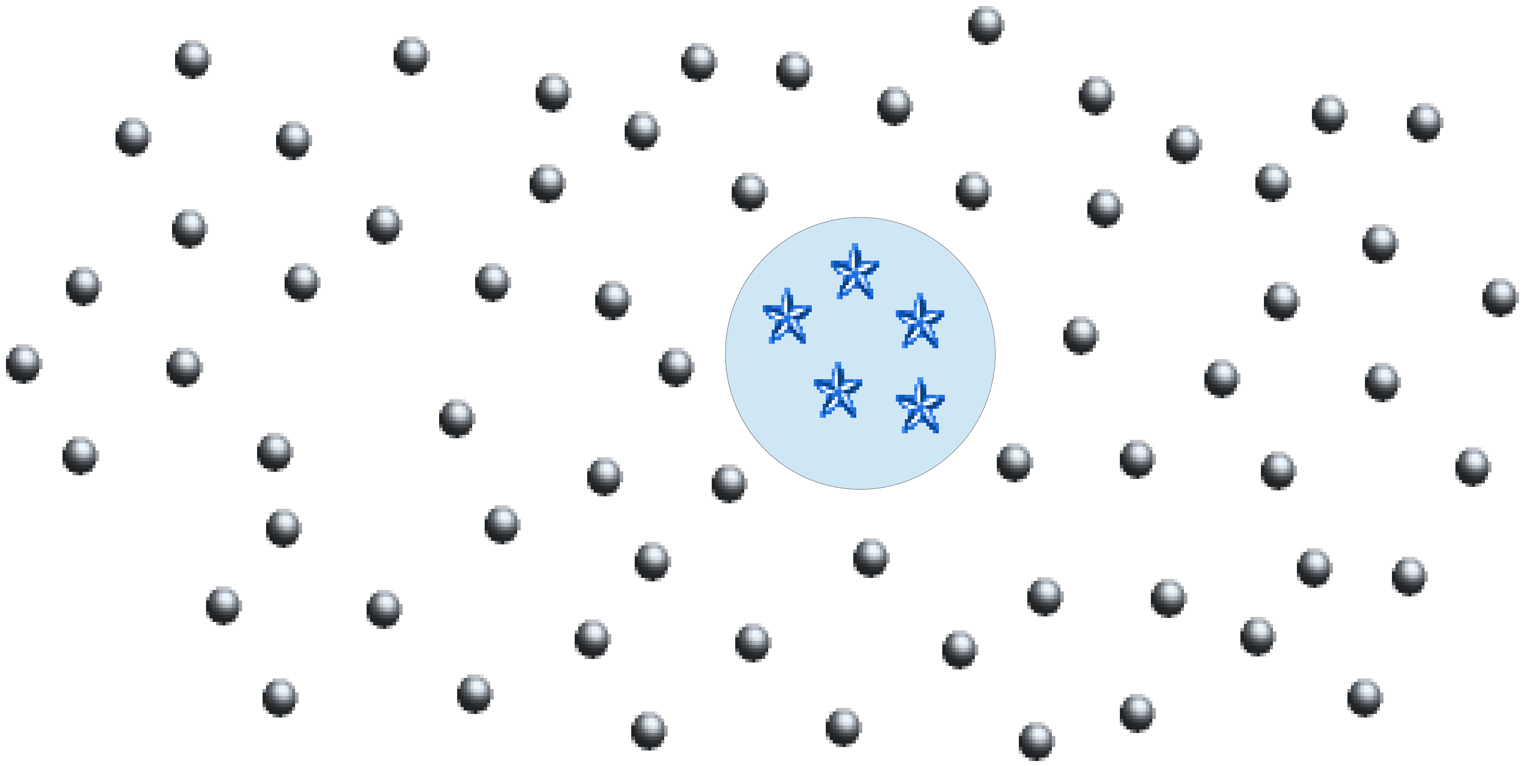
Maximum allowed Pathway Overlap m_0



Pathway Packing Models Examined

- Nearest Neighbor Pathway Model,
- Random Selection Pathway Model,
- Random Selection Pathway Model with Cutoff Radius.

Nearest Neighbor Pathway Model



Information Capacity

- Pathways cannot overlap completely-There is a minimum distance of the pathway centers, D
- This minimum distance is given implicitly in terms of the maximum overlap m_0 through the formula

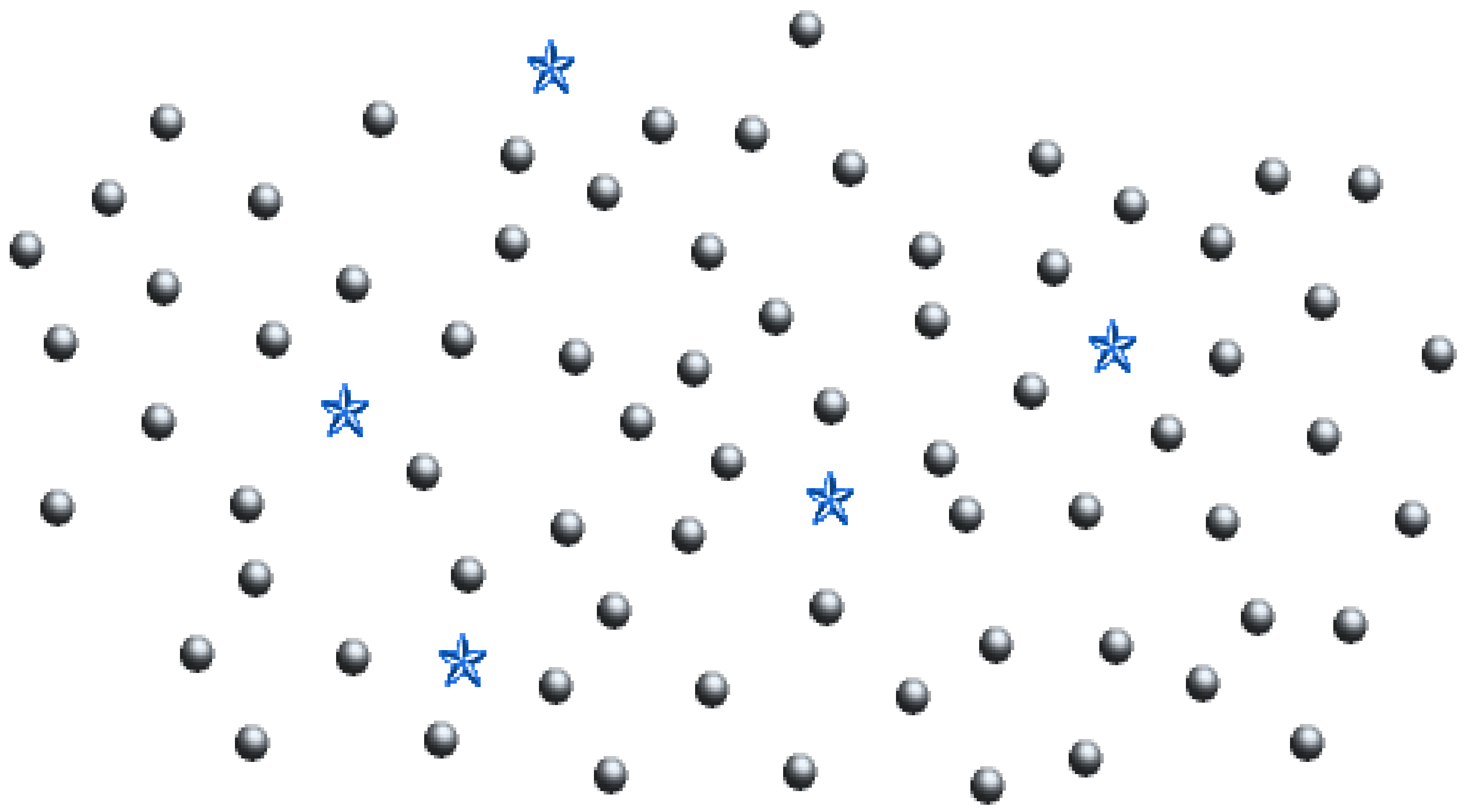
$$m_0 = \frac{1}{12}\pi(4\sqrt[3]{\frac{3n}{4\pi}} + D)(2\sqrt[3]{\frac{3n}{4\pi}} - D)^2$$

- The maximum number of non-interfering pathways is given by

$$N_P = \frac{\mathrm{N}\sqrt{2}}{D^3} = O(N),$$

where N is the overall number of neurons. This is rather small.

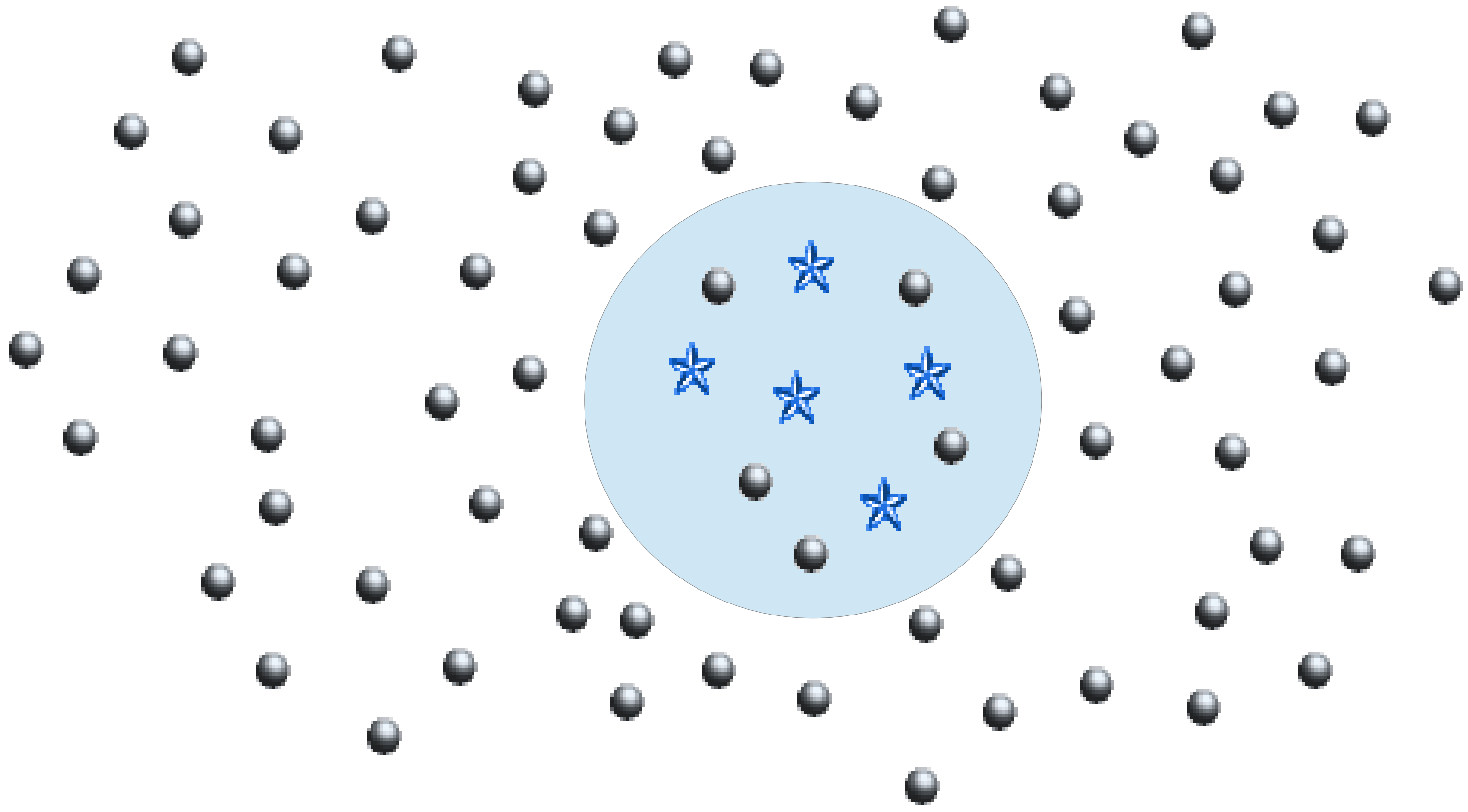
Random Selection Pathway Model



Information Capacity

- Here the pathway neurons are selected at random from the whole aggregate.
- In this model the number of non-interfering pathways is
$$N_P \sim O(N^{m_0})$$
- The number of non-interfering pathways is much larger than in the nearest neighbor pathway model.
- However **<u>locality of the pathways is lost</u>**. This pathway model is inappropriate for early visual areas.

Random Selection Pathway Model with Cutoff Radius



Information Capacity

- In this case there are two phases of the model. The **ordered and** the **disordered phase**
- If the **pathways are dense** enough, then the pathway centers can not overlap and the number of pathways is
$$N_P = O(N)$$
- If the **pathways are not dense** enough, then the pathway centers can overlap and the number of pathway is
$$N_P = O(Ne^{am_0}), a > 0$$
- Hence **for localized but not dense pathways, the number of pathways is huge and locality is retained**.

Result concerning Information Capacity

- If pathways are dense, local structures then the number of non-interfering pathways is $O(N)$.
- If the number of pathways are dilute, non-local structures then the number of non-interfering pathways is $N_P = O(N^{m_0})$.
- If pathways are dilute enough, retaining locality, then the number of pathways is $N_P = O(Ne^{am_0})$, and this number increases exponentially in the overlap m_0.
- Note that <u>**if the pathways are dilute enough, then the system has an enormous information capacity**</u>.

Question

- Where does a multilayer perceptron differ from animal neural nets;

Stochasticity

- An artificial trained network on the same input produces the same output

- A natural trained network on the same input can produce different output

Supervised vs. Unsupervised Learning

- Suppose that we have the activity of the input layer A to a number of stimuli presented, multiple times each, and we want to deduce the stimulus presented S_i from the neuronal activity A_i of the input layer.

- If we are given a training set, for which we **<u>know both $\boldsymbol{A_i}$ and $\boldsymbol{S_i}$</u>**, then we have **<u>supervised learning</u>**

- If we are given **<u>only $\boldsymbol{A_i}$</u>**, and we try to cluster these $\underline{A_i}$'s into groups that probably correspond to the same signal, then we have **<u>unsupervised learning</u>**.

Bottom-up Learning is Unsupervised Learning

- In multilayer perceptron **there is a cost function on the output, measuring the classification error** of the network on the training set. The weights of the trained network are determined through the minimization of this cost function. This can only be done if we have a labelled training set hence it is supervised learning.
- In animal neural networks, a set of pictures is presented through the eyes of the animal, and **the picture world is split into objects that are grouped together.** This is unsupervised learning. It is possible to have supervised learning in animals as well, however this requires a "teacher" that characterizes objects to the animal, and in any case top down information from higher brain areas.

Spiking Structure

- In a multilayer perceptron, the input to a neuron B_j of a layer B which comes after layer A, is given by a linear combination of the outputs of layer A neurons, $I_{B_j} = \sum_i w_{ji} O_{A_i}$, where the weights w_{ji} are the weights of the trained network
- In an animal neural net the input to a neuron comes in a time series of spikes from the connected and active neurons, weighted by the synaptic strengths. However, it is known that a lone spike has low probability to pass a synapse, while a spike that comes after another spike has high probability to pass. Hence **there are time filters**
- Probably the animal neural net is set so that statistically surprising weighted cofirings (not likely in spontaneous firing) pass through. Hence **information flow is probability determined**.

Pathway Structure

- In a multilayer perceptron there is pathway to neuron transmission. The input to a layer B neuron is determined by a group of layer A neurons. Hence, we can have only as many group information transmitted as layer B neurons are.
- In an animal neural net, there is probably pathway to pathway transmission, hence information from many more layer A groups can be transmitted. This is probably achieved through the horizontal connections of the various layers.

Graph Structured Data

- Multilayer perceptron can recognize objects but not their structure. If the object is a chair, **the net can recognize the chair or the legs** of the chair, but **not link the objects recognized**

- Animal neural nets deal with structured data. For this net **a chair has four legs attached to it**

Invariances

- Animal neural nets **recognize objects equally well irrespective of position, size or rotational state**

- For a multilayer perceptron **an object in different position or of different size is potentially a different object**. The two may be linked together into one object by the net, however one may be more difficult to recognize than the other, and anyway it is not certain that the two will be linked together

Conclusion

- Stochasticity in neuronal response makes animal neural nets robust, able to control information flow, able to avoid infinite loops. Furthermore stochasticity is necessary for imagination
- Information capacity in animal neural nets is enormous, hence such networks can perform information demanding processes. However there are **limitation to memory coming from parallel processing** and from the **problem of addressing** this memory
- Animal neural nets differ from multilayer perceptron in many ways

Input Structure for Early Visual Areas 1

- It seems that frequently **visual input meets a conglomerate of classifiers** that respond when a signal (possibly an object) is recognized by the classifier.
- This **recognition response is encoded in pathways** that gets activated when the particular signal is present.
- These **classifiers operate in parallel**, and it is possible that there is a huge number of them. In this way a picture is split into objects.
- A famous unsolved problem is the precise way the brain splits a picture into objects.

Input Structure for Early Visual Areas 2

- The above **<u>classifiers form the keyboard of the brain</u>**. When we press a key in the keyboard, a letter is encoded in 8 bits.
- Similarly when an object is present in a picture, the object activates its pathway and the object is encoded in the pathway.

Formation of Classifiers in the Brain

- **Classifiers are formed by learning rules**. These are iterative processes that adjust synaptic strengths so that a group of neurons responds to a particular signal, forming a pathway.
- It is important to note that the **pathways are the outcome of iterative processes, hence they can be complicated**. Recall that iterative processes often lead to fractal structures, like the Mandelbrot set.
- Hence the key to understanding early visual area classifiers is not the search for the structure of particular classifiers, but rather the **search for the right learning rules**.

The Most Famous Learning Rule: Hebbian Learning

- **When two joining cells fire simultaneously, the connection between them (synapse) strengthens**.

- Verified experimentally by Lomo (1966) in the rabbit hippocampus, where he showed long term potentiation of chemical synapses initiated by a high frequency stimulus.

- The **activity** of the network **is balanced by long term depression** of synapses that receive low frequency input.

Winner Takes All technique in Learning

- Suppose that we have two layers of neurons, layer A and layer B. Furthermore suppose that somehow neurons in layer B are connected to a number of neurons in layer A.
- **Winner takes all learning dictates that the most active B neuron (or maybe neurons) increases its synaptic strength with active A neurons.** Less active B neurons may or may not decrease their synaptic strengths with active A neurons.
- Winner takes all technique is **appropriate for unsupervised learning**.

Supervised vs. Unsupervised Learning

- Suppose that we have the activity of neuronal layer A to a number of stimuli presented, multiple times each, and we want to deduce the stimulus presented S_i from the neuronal activity A_i.

- If we are given a training set, for which we **know both $\boldsymbol{A_i}$ and $\boldsymbol{S_i}$**, then we have **supervised learning**

- If we are given **only $\boldsymbol{A_i}$**, and we try to cluster these A_i's into groups that probably correspond to the same signal, then we have **unsupervised learning**.

Top Down and Bottom Up Processing

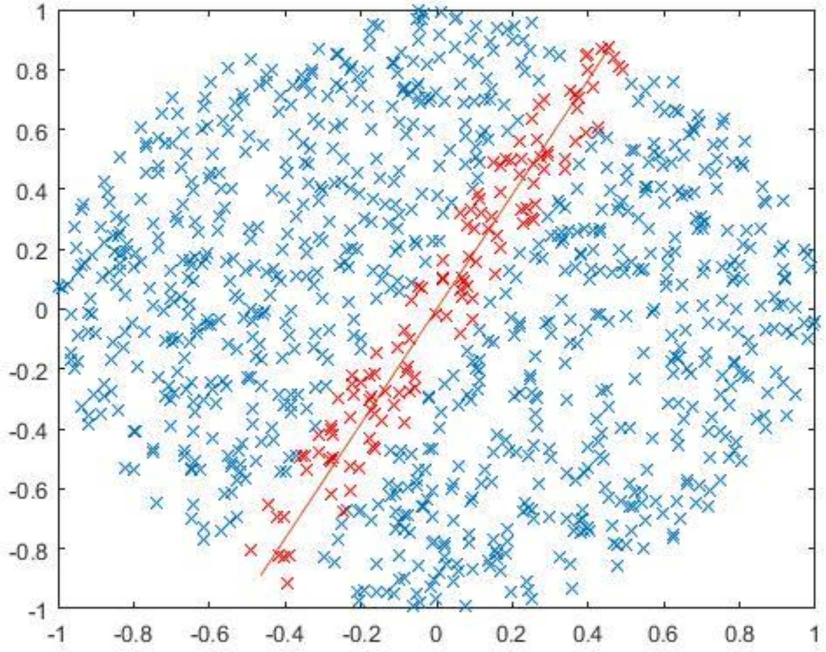
- Bottom up processing in psychology is considered to be processing that **occurs directly on the input without the interference of higher brain areas**. In vision, such processing is the division of an image into objects, but not the identification of these objects. **Bottom up learning is often unsupervised learning.**

- **Top down processing involves feedback from higher brain areas**. Such processing is the identification of an object with the word that corresponds to it. **Top down learning can be supervised**.

A Toy Model for Unsupervised Learning: The Connectivity Matrix Algorithm

- Two layers of neurons, detector layer A and output layer B
- Layer A has 1000 randomly placed neurons within a circle of radius 1
- Layer B has 16 neurons, initially with 150 random connections with layer A
- The signal presented is one of 8 lines that pass through the center of the circle
- A neurons are activated if they are distance 0.1 from the line
- **<u>Input to B neurons is the sum of the synaptic activities of active A neurons connected to the particular B neuron</u>** (initially synaptic strengths are 0.5).

Example Response of Layer A



Learning Algorithm

- Select the B neuron that receives highest input, B_{max} and set the learning rate to ε.
- If an A neuron is active and connected to B_{max}, then increase its synaptic strength by

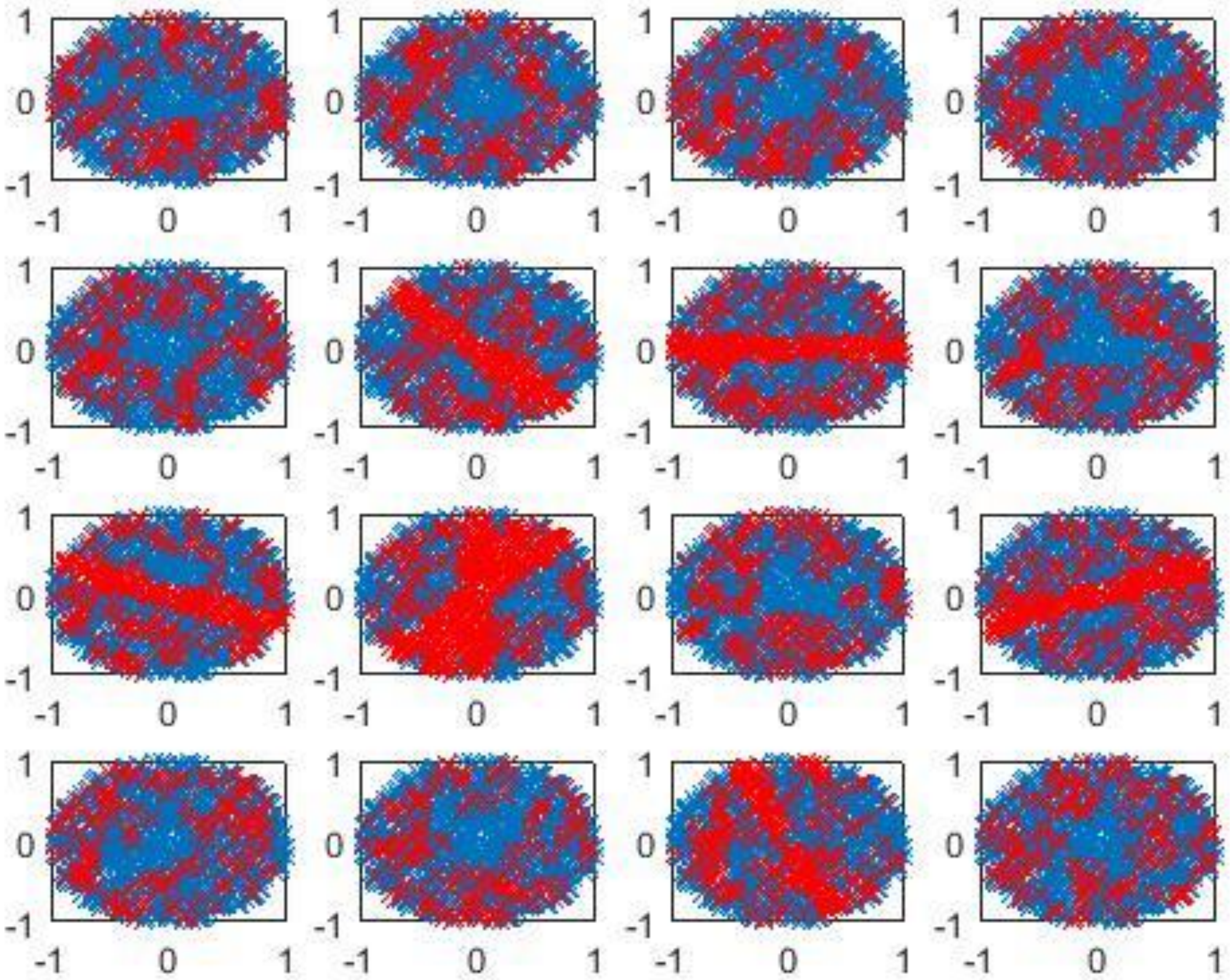
$$s \to s + \varepsilon(1 - s)$$

- If an A neuron is active and connected to another B neuron, then decrease its synaptic strength by

$$s \to s - \varepsilon s$$

- If a synaptic connection is less than 0.1 disconnect the neuron

Outcome of Learning Algorithm



Conclusion

- There is little understanding as yet on the way the early visual areas recognize objects
- Experimentally little more is known beyond the Hebbian rule for learning
- The **<u>information capacity of the brain is huge</u>**, hence it is possible that the brain uses memory greedy algorithms for object recongnition
- A **<u>winner takes all strategy seems to be important in unsupervised learning</u>**
- **<u>A note of optimism:</u>** More precise experimental data are expected in the near future.