'Perceptions'?

ESSAY

by

Michael J. C. Gordon

Gonville & Caius

A human being when looked at from the point of view of a behavioural psychologist is an automaton A which gives 'responses' when stimulation of a certain kind is applied. In particular such an automaton can be taught to recognise patterns in the following sense: Say we wish to teach it to give a response "Yes" when shown a picture of the numeral 2 and "No" when shown some other numeral. Suppose we have arranged that "Yes" and "No" are the only responses A can make. We begin by choosing a sequence of patterns P_1, P_2 .... etc such that each P_r is a picture of some numeral, we then stimulate A with P_1, if the required response is given we stimulate it with P_2 and so on until a wrong response is made (i.e. "Yes" if P_r is not a 2 or "No" if P_r is a 2) whereupon A is 'punished' by an unpleasant stimulus. e.g. an electric shock, this process is continued and it is an observable fact that after a certain time the frequency of wrong responses becomes very small. Various attempts have been made to explain how the human visual system can recognise patterns such as the numeral 2, and also various attempts have been made to design an artificial automaton capable of exhibiting the above learning and recognising behaviour, the perceptron and its descendants is one such attempt.

From the above it is clear that 2's have the property that a human can be taught to recognise all 2's by being shown just a finite number of examples of them. E.M. Braverman, a Russian cybernetician has called such collections of objects "images", he defines an image to be a collection of physical objects such that a human can be taught to recognise whether any given object is a member of the collection by being shown just a finite number of the members of the collection. Other examples of images are: Pictures of John Smith, things which smell of rotten eggs etc. An example of a non-image is: The collection of people whose surname is Smith. It should be noticed that the definition of an image depends on the elements of it being recognisable by a human, such images will be called "human-images" to distinguish them from a more general kind of image which we now describe. Let A be an automaton and Q some collection of physical objects such that A can be stimulated by being shown elements of Q, we say Q is an A-image if there exists some training process such that A can be taught to recognise whether something is an element of Q, by showing it only a finite number of elements of Q. In what has just been said the terms "automaton", "physical objects" and "training process" were used very vaguely in order to give them some intuitive significance. However, the rest of this essay will be mainly concerned with the construction of a very simple mathematical model in which these terms are given rigorous definitions. First, let us concentrate on defining entities which will function as "physical objects" in our model. It is important to realise that there are dangers in trying to be more rigorous whilst still being general, consider for example the 'set' $X = \{x \mid x \text{ is a } 2\}$ At first sight this looks like a well defined entity, capable of being used like any other set. In fact X looks like being a nice extensional definition of the property "being a 2", but X is not a set, firstly, because we have no well defined definition of "being a 2" and secondly, even if we did the class of entities we would call 2's is continually changing(e.g. artists are inventing new shapes of 2's). In order to make the idea of a physical object well defined we must impose some artificial physical structure on the sort of object which will feature in the recognition process. Let us restrict our attention to visual patterns and assume each pattern, like a newspaper picture, is made up of a large finite number N of dots arranged in a rectangular array. Assume further that the degree of blackness of each dot can be described by a number $x^r \in [0, 1] \subseteq \mathbb{R}$ so that $x^r = 0$ means the r^{th} dot is white, $x^r = 1$, means r^{th} dot is black and $x^r = \frac{1}{2}$ means the r^{th} dot is a shade of gray midway between white and black etc. Thus each pattern is now thought of as an N-tuple $(x^1, \ldots, x^N)$ where $x^r \in [0, 1]$ and so the universe of physical objects we shall consider is the set

$$P = \{(x^1, \ldots, x^N) \mid x^i \in [0, 1]\} = [0, 1]^N$$

Restricting patterns to be elements of P eliminates the problem of the nature of 'sets' like $\{x \mid x \text{ is a } 2\}$ for now we can interpret this as the real set.

$$\{x \mid x \in P \text{ and nine out of ten people say } x \text{ is a } 2\}$$

This definition of a pattern is the first step in the formalisation of the concept of an image, recall that intuitively we think of an image as dependent on two things, a collection of physical objects Q and an automaton A capable of being taught to recognise elements of Q. Now the set P can be interpreted (in this context) as the collection of all physical objects and so Q will be some subset of P. So far we have made no attempt to formalise the automaton part of an image, this we now do.

We assume there exists a set of states S of the automaton A such that each state $s \in S$ determines entirely the input/output behaviour of A. We thus think of A as a mapping

$$A : \{\text{inputs}\} \times S \longrightarrow \{\text{outputs}\}$$

The inputs here will be elements of P (applying an input x will be interpreted as showing x to A) and there will be two possible outputs -1 (corresponding to "No") and 1 (corresponding to "Yes"). The nature of S will be discussed later. Thus:

$$A : P \times S \longrightarrow \{-1, 1\}$$

Since $\{-1, 1\} \subseteq \mathbb{R}$ (the real numbers) we can regard A as mapping P x S into $\mathbb{R}$ rather than onto $\{-1, 1\}$

During the training process as described in the first paragraph we show A elements of P and indicate whether the output given is correct, if it is not then A modifies its internal state $s \in S$ according to some rule. We capture this notion of a rule in a function

$$\sigma_A : P \times \{-1, 1\} \times S \longrightarrow S$$

Where we interpret σ_A as giving the next state $s_{n+1} = \sigma_A(p, a, s_n)$ of A, where A gives output a when it is in state s_n and input p is applied. Given a modification function σ_A, a sequence of patterns $p_n \in P$ and an initial state $s_1 \in S$ we can derive a sequence of states $s_n \in S$ of A determined by

$$s_{n+1} = \sigma_A(p_n, A(p_n, s_n), s_n)$$

this sequence $\{s_n\}$ we call the sequence of states corresponding to $\{p_n\}$ and s_1 via σ_A and it is interpreted as the sequence of states A goes through as it is trained, starting in state s_1, by being shown the p_n in turn and being modified by σ_A.

We can now define rigorously an A-image as a set $Q \subseteq P$ such that there exists a modification function σ_{AQ} such that:

(a) For any sequence $\{p_n\} \subseteq P$ and any initial state $s_1 \in S$ the sequence of states $\{s_n\}$ corresponding to $\{p_n\}$ via σ_{AQ} is eventually constant with value s (i.e. $\exists$ m s.t. $n \geqslant m \Longrightarrow s_n = s$)

(b) $$A(x,s) = \begin{cases} 1 \text{ if } x \in Q \\ -1 \text{ if } x \notin Q \end{cases}$$

Thus an A-image is a set of patterns (Q) such that there exists a training process (σ_{AQ}) such that if A starts in any state and is trained according to σ_{AQ} on any sequence of patterns ($\{p_n\}$) it will, after a finite time, settle down to a state in which it responds "Yes" if and only if an element of Q is applied to its input.

Let us now summarise the model we have set up, we have defined a quadruple (P,S,A,σ_A) where:

$P = [0, 1]^N$

S is a set - called the set of states of A,

$A : P \times S \longrightarrow \{-1, 1\}$

$\sigma_A : P \times \{-1,1\} \times S \longrightarrow S$

Given a sequence $\{p_n\}$ of elements of P, an element $s_1 \in S$ and a σ_A we have defined the corresponding sequence in S to be $\{s_n\}$ where

$$s_{n+1} = \sigma_A(p_n, A(p_n, s_n), s_n)$$

Finally we have defined an A-image to be a subset $Q \subseteq P$ such that there exists a σ_{AQ} such that if $\{p_n\}$ is a sequence in P, s_1 is arbitrary in S and $\{s_n\}$ is the corresponding sequence in S, then

(a) $\exists$ M s.t. $n \geqslant M \Longrightarrow s_n = s$

(b) $A(x,s) = 1 \Longleftrightarrow x \in Q$

Let us now consider the classical problem of pattern recognition in terms of the above model. Assume we are given a finite collection $\{Q_a\}$ of human-images, for example, we might have

$$Q_a = \{x \mid x \in P \text{ \& } x \text{ is a picture of the numeral } a\}$$

We want to construct an automaton A such that Q_a is also a collection of A-images. For if we can find such an A, then we can build a machine to recognise out of which Q_a a given pattern x comes, as follows: We train A in turn on each Q_a and so obtain states $s_a \in S$ such that

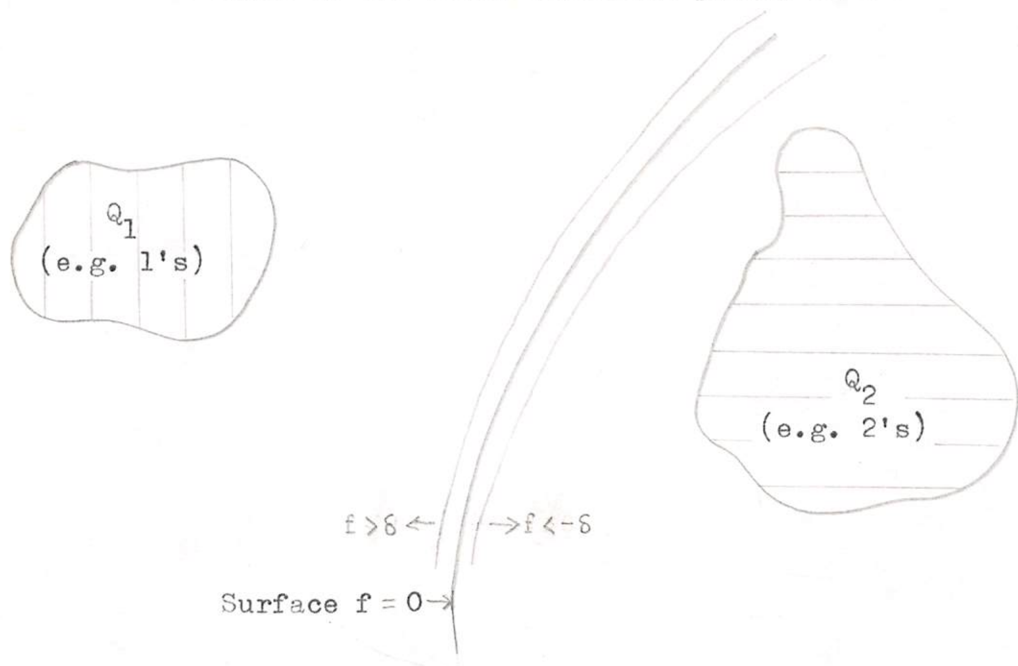
$$A(x,s_a) = 1 \Longleftrightarrow x \in Q_a$$

Then given $x \in \bigcup_a Q_a$, to decide out of which Q_a it came we just evaluate $A(x,s_a)$ for each a and find the one(s) for which this is 1.

It has been widely held that it is possible to define an inner-product on $\mathbb{R}^N$ (e.g. the natural inner-product) such that the induced metric on $P \subseteq \mathbb{R}^N$ has the property that patterns which look the same in P are near together with respect to this induced metric and patterns which are far apart in P look different. Furthermore it is held (especially by the Russians who invented images) that human-images form 'compact clusters' in P ("compact" as used here has no topological significance) so that, for example, if one looks at the points in P corresponding to the human-images: numeral 1, numeral 2, one would see something as in Fig. 1 below:-

Fig. 1

Points of the paper represent points of P



It can be convincingly argued that this assumption concerning the structure of human-images is misleading, however, if we assume it, we can give an interesting heuristic derivation of the structure of a well known type of automaton known as a perceptron. Let us restrict ourselves to designing a device for discriminating between two images Q_1 and Q_2 which form well separated, disjoint clumps in P with respect to some metrical structure. From the diagram above it seems reasonable to assume Q_1 and Q_2 can be separated by a nice well behaved function, $f : P \longrightarrow \mathbb{R}$, by this it is meant that f satisfies

$$\left.\begin{array}{ll} x \in Q_1 \Longrightarrow f(x) > \delta > 0 \\ x \in Q_2 \quad f(x) < -\delta < 0 \end{array}\right\} \quad \text{some } \delta > 0 \qquad (1)$$

We now choose a basis $\{\phi_r(x)\}$ of the vector-space of well behaved functions mapping $\mathbb{R}^N \longrightarrow \mathbb{R}$ such that as n increases $\phi_r(x)$ becomes more 'jagged'. The assumption is then made that there exists an m such that there is a separating function f which can be expanded in a Fourier series with respect to $\{\phi_r(x)\}$ with all Fourier coefficients zero for $n \geqslant m$. Intuitively this is saying we can find a separating function f which has no very 'jagged' harmonics, i.e.

$$f(x) = \sum_{i=1}^{m} c^i \phi_i(x) \qquad (2)$$

The previous paragraph could be made more rigorous but there seems little point as it is only a heuristic justification for assuming (2). Define

$$\Phi : P \longrightarrow \mathbb{R}^m \text{ by } x \longmapsto (\phi_1(x), \ldots, \phi_m(x)) = \Phi(x)$$

Thus with each point $x \in P$ we can associate a point $\Phi(x) \in \mathbb{R}^m$ in such a way that the points in $\mathbb{R}^m$ corresponding to any two disjoint images can be separated by a hyperplane. For this reason $\mathbb{R}^m$ is called the linearisation space. We use these ideas to motivate a concrete realisation of (P,S,A,σ_A). Let $S = \mathbb{R}^m$, $A : P \times S \longrightarrow \{-1,1\}$ be defined by

$$A(x,s) = \begin{cases} +1 & \text{if } (\Phi(x),s) \geqslant 0 \\ -1 & \text{if } (\Phi(x),s) < 0 \end{cases}$$

where for $x,y \in \mathbb{R}^m$ (x,y) denotes the natural inner-product and $\Phi : P \longrightarrow \mathbb{R}^m$ is as above. Thus the automaton A corresponding to a state $s \in S$ just sends points on the -ve side of the hyperplane $\{x \mid (x,s) = 0\}$ determined by s to -1 and points on the positive side to 1. We would like to define σ_A in such a way that if $\{p_r\}$ is a sequence of points such that $p_r \in Q_1 \cup Q_2$ then the corresponding sequence of states eventually becomes a vector s which determines a hyperplane separating Q_1 and Q_2 in the linearisation space, i.e. $\{A(x,s_n)\}$ regarded as a sequence of functions of x 'tends' to a function separating Q_1 and Q_2 as '$n \longrightarrow \infty$'. Intuitatively what we want to do is to increase s_n (i.e. make $s_{n+1} \geqslant s_n$) if $A(p_n,s_n)$ is negative when it should be positive and decrease s_n if $A(p_n,s_n)$ is positive when it should be negative, this motivates the following definition of σ_A

$$\sigma_A(x,a,s) = \begin{cases} s & \text{if } x \notin Q_1 \cup Q_2 \text{ or } (x \in Q_1 \ \& \ a = +1) \text{ or } (x \in Q_2 \ \& \ a = -1) \\ s + \Phi(x) & \text{if } x \in Q_1 \ \& \ a = -1 \\ s - \Phi(x) & \text{if } x \in Q_2 \ \& \ a = +1 \end{cases}$$

We now prove the following theorem which says that under certain conditions A will learn to distinguish members of Q_1 and Q_2 in the sense discussed in the first paragraph of this essay.

<u>Theorem I:</u>

Let Q_1, Q_2 be two subsets of P such that there exists a vector $s \in \mathbb{R}^m$ such that for some $\delta > 0$

$$x \in \Phi(Q_1) = \{\Phi(x) \mid x \in Q_1\} \Longrightarrow (s,x) > +\delta > 0$$
$$x \in \Phi(Q_2) = \{\Phi(x) \mid x \in Q_2\} \Longrightarrow (s,x) < -\delta < 0 \qquad (3)$$

Further assume $\Phi(Q_1)$ and $\Phi(Q_2)$ are bounded in $\mathbb{R}^m$. Then if $\{p_n\}$ is any sequence of points from $Q_1 \cup Q_2$ and s_1 is any point of $\mathbb{R}^m$ there exists a number M <u>independent</u> of $\{p_n\}$ such that $s_{n+1} = \sigma_A(p_n, A(p_n,s_n), s_n)$ for $n = 1,2,3, \ldots$ takes on at most M distinct values (i.e. A makes less than M mistakes).

<u>Proof:</u>

Define $Q = \{x \mid x \in \Phi(Q_1) \text{ or } -x \in \Phi(Q_2)\}$

Let $\{s_{r(n)}\}$ be the subsequence of $\{s_n\}$ obtained by deleting all terms of $\{s_n\}$ such that $s_n = s_{n+1}$ so that $s_{r(n+1)} \neq s_{r(n)}$ for all n. (So, for example if $\{s_n\} = (s_1,a,a,b,a,a,b,c,c,c,c, \ldots)$ then $s_{r(n)} = (s_1,a,b,a,b,)$ where $s_1,a,b,c \in \mathbb{R}^m$ & $r(1) = 1$, $r(2) = 3$, $r(3) = 4$, $r(4) = 6$, $r(5) = 7$, $r(m)$ for $m \geqslant 6$ not defined)

We show that $|\{s_{r(n)} | r(n) \text{ is defined}\}| \leq M^*$ and this is clearly equivalent to the number of distinct values of s_n being less than $M^*+1=M$. From the definition of $\{s_n\}$ and σ_A it is easy to check that

$$s_{r(n+1)} = \sigma_A(p_{r(n)}, a, s_{r(n)}) = \begin{cases} s_{r(n)} + \Phi(p_{r(n)}) & \text{if } x \in Q_1 \\ s_{r(n)} - \Phi(p_{r(n)}) & \text{if } x \in Q_2 \end{cases}$$

Let $\{x_n\} \subseteq \mathbb{R}^m$ be the sequence defined by

$$x_n = \begin{cases} \Phi(p_{r(n)}) & \text{if } p_{r(n)} \in Q_1 \\ -\Phi(p_{r(n)}) & \text{if } p_{r(n)} \in Q_2 \end{cases}$$

then $x_n \in Q$. Thus $s_{r(n+1)} = s_{r(n)} + x_n$ so that:

$$s_{r(n)} = \sum_{r=1}^{n-1} x_r + s_1 \qquad (4)$$

set $a = \inf_{x \in Q} \dfrac{(s,x)}{\|s\|}$

$b = \sup_{x \in Q} \|x\|$, $b < \infty$ by assumption,

then $x \in Q \Rightarrow (s,x) > \delta$ by (3) so $a \geq \dfrac{\delta}{\|s\|} > 0$ (5)

Also by the definition of "inf":

$$\frac{(s,x)}{\|s\|} \geq a \;\forall x \in Q \text{ so } \frac{(s,x_n)}{\|s\|} \geq a \;\forall n, \qquad (6)$$

Summing (6) over n and using (4) gives: $\dfrac{(s, s_{r(n)} - s_1)}{\|s\|} \geq (n-1)a$ (7)

but by the Cauchy-Schwarz inequality $\|s\| \, \|s_{r(n)}\| \geq (s, s_{r(n)})$ so from (7)

$$\|s_{r(n)}\| \geq \frac{(s,s_1)}{\|s\|} + (n-1)a \qquad (8)$$

Now from (4) $s_{r(n)} = s_{r(n-1)} + x_{n-1}$

$$\Rightarrow \|s_{r(n)}\|^2 = \|s_{r(n-1)}\|^2 + 2(s_{r(n-1)}, x_{n-1}) + \|x_{n-1}\|^2$$

$\Rightarrow \|s_{r(n)}\|^2 \leq \|s_{r(n-1)}\|^2 + b^2$ as $(s_{r(n-1)}, x_{n-1}) \leq 0$ by def^n. of $\{s_{r(n)}\}$ & σ_A

$$\Rightarrow \|s_{r(n)}\|^2 \leq \|s_1\|^2 + (n-1)b^2 \qquad (9)$$

(8) & (9) $\Rightarrow \|s_1\|^2 + (n-1)b^2 \geq \|s_{r(n)}\|^2 \geq \left[\dfrac{(s,s_1)}{\|s\|} + (n-1)a\right]^2$ (10). Since $a > 0$ by (5)

(10) for sufficiently large n cannot hold, let M^* be the first integer such that

$$\|s_1\|^2 + (M^*-1)b^2 < \left[\frac{(s,s_1)}{\|s\|} + (M^*-1)a\right]^2$$

thus $\{s_{r(n)}\}$ must terminate before $r_{(n)} = M^*$. (10) is independent of $\{p_n\}$ and so the maximum number of terms, M^*, in $\{s_{r(n)}\}$ is independent of $\{p_n\}$ Q.E.D.

The above theorem shows that during a training process A will eventually stop making errors, however, the state s in which A eventually comes to rest in, is not necessarily the normal vector of a hyperplane separating $\Phi(Q_1)$ and $\Phi(Q_2)$ for example, if the training sequence $\{p_n\}$ was defined by $p_n = p \in Q_1 \ \forall n$ then A would come to rest in the first state s_n for which $A(p, s_n) = 1$ and it would not follow that for this s_n, and for arbitrary $x \in Q_1$ that $A(x, s_n) = 1$. It can be shown that if p_n is chosen 'sufficiently random' then s will, with probability 1, be the normal vector of a hyperplane separating $\Phi(Q_1)$ and $\Phi(Q_2)$. In proving this the phase "sufficiently random" is replaced by well defined statistical assumptions. The number M which gives an upper bound, over <u>all</u> training sequences, of the number of errors A can make, could be calculated if we knew something about the sizes of $\Phi(Q_1)$ and $\Phi(Q_2)$ since M is only a function of a,b and s_1, however, in practice we do not know anything about a and b. The importance of Theorem I is mainly that it seems intuitively to indicate that we are on the right track in defining (P, S, A, σ_A) as above, however, it is possible to construct a slightly different realisation of (P, S, A, σ_A) such that Theorem I is still true but now it does not intuitively indicate that anything we should call learning is going on, to see this, define P,S,A as before and redefine σ_A. First let

$$S' = \{x \mid x \in S \ \& \ x \text{ has rational coordinates w.r.t. the natural basis of } \mathbb{R}^n\}$$

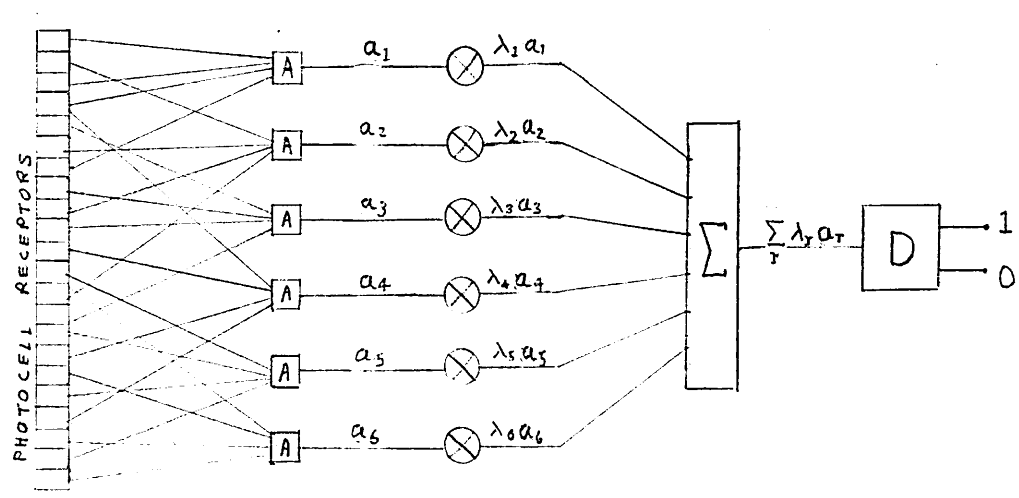
S' is countable so let $S' = \{y_1, y_2, \ldots\}$ then define

$$\sigma_A(p,a,s) = \begin{cases} s \text{ if } (p \in Q_1 \ \& \ a = +1) \text{ or } (p \in Q_2 \ \& \ a = -1) \\ \text{otherwise } y_1 \text{ if } s \notin S' \\ \qquad\qquad\quad y_{n+1} \text{ if } s = y_n \end{cases}$$

Thus training A using this new σ_A merely amounts to trying out each element of S' in turn. Since S' is dense in S it is easy to see that if there exists an $s \in S$ satisfying condition ③ of Theorem I then there exists an $s' \in S'$ which will satisfy ③. Let $y_{M'}$ be the first such element in S' then clearly Theorem I is still true if M is replaced by M' (strictly speaking for Theorem I to be true, in this context, the initial state s_1 must not be y_n for $n > M'$, however if s_1 is chosen at random this occurrence has probability zero).

Around 1960 F. Rosenblatt, an American Scientest proposed a design for a learning machine, known now as a perceptron, shown in Fig. 2

Fig. 2.

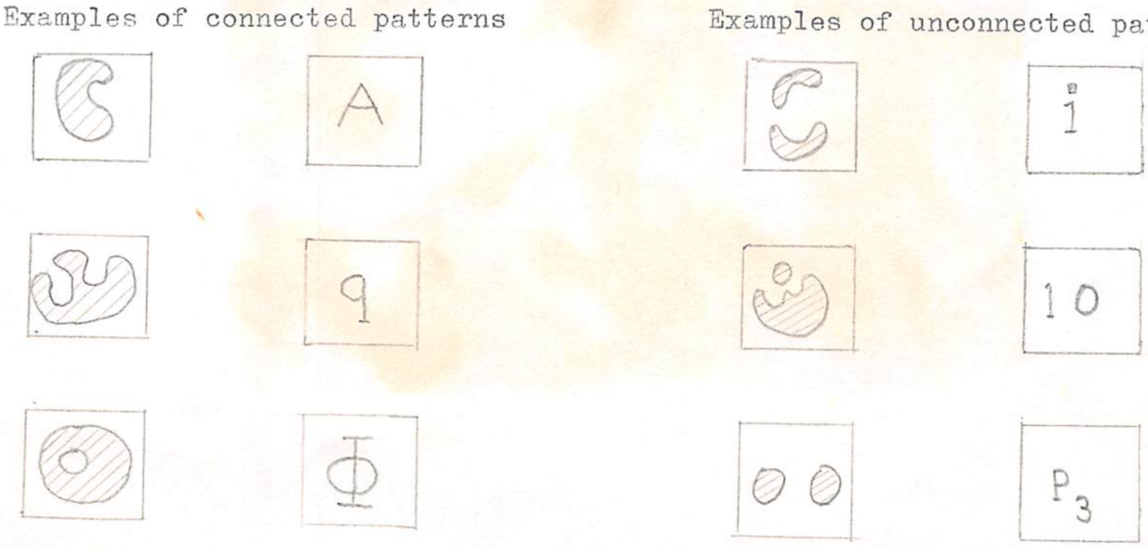


It worked as follows: The photocells were connected, initially in a random manner, to threshold devices known as associative or A-elements. These would compute an output a_r as a function of the inputs from the photocells, these outputs then would be multiplied by constants λ_r (by the devices marked $-\otimes-$ in Fig. 2), and then all the signals $\lambda_r a_r$ would be summed by the box marked Σ, finally, the box marked D would compute the sign of the output $\sum_r \lambda_r a_r$ of Σ and indicate "1" if the sign was positive and "0" otherwise. That is, if a_r was the output from the rth A-element the perceptron would indicate "1" or "0" according as $\sum_r \lambda_r a_r$ was greater or less than 0. The perceptron was 'taught' to descriminate between two classes of patterns by showing it a sequence of patterns chosen at random from the two classes and modifying the variables λ_r in a preassigned way each time it made a mistake. Now if we interpret the A-elements of the perceptron as components of a function Φ mapping the set of outputs from the photocell into $\mathbb{R}^m$ (where there are m A-elements), we see the perceptron is just a physical realisation of the automaton A which features in Theorem I. (With $S = \{(\lambda_1, \ldots, \lambda_m) \mid \lambda_i \in \mathbb{R}\}$) And in fact the modification process σ_A turns out to be very similar to the process Rosenblatt proposed for modifying the variables λ_r in the perceptron. It is interesting to note that Rosenblatt based the design of the perceptron on the then current ideas about the structure of the human visual system, whereas in this essay we have arrived at essentially the same structure by assuming similar patterns form 'compact clusters' in a certain space, (P).

So far we have only considered the structure of human-images and we have shown that <u>if</u> these form 'compact clusters' in P a perceptron-like device can learn to distinguish them, it turns out, however, to be most unlikely that human-images form such nice 'compact clusters'. Marvin Minsky and Seymour Papert have looked at what sort of sets perceptrons can learn to discriminate between, i.e. they have tried to determine what 'perceptron-images' look like. In view of the large number of light receptors humans have in each eye, if the perceptron is to have any chance of being a plausible model of the human visual system each A-element must have inputs coming from only a small fraction of all the receptors. Minsky and Papert have shown that if this is so then no perceptron could be built (let alone taught) to recognise whether a pattern is 'connected' or not in the sense of Fig. 3 below:

Fig. 3

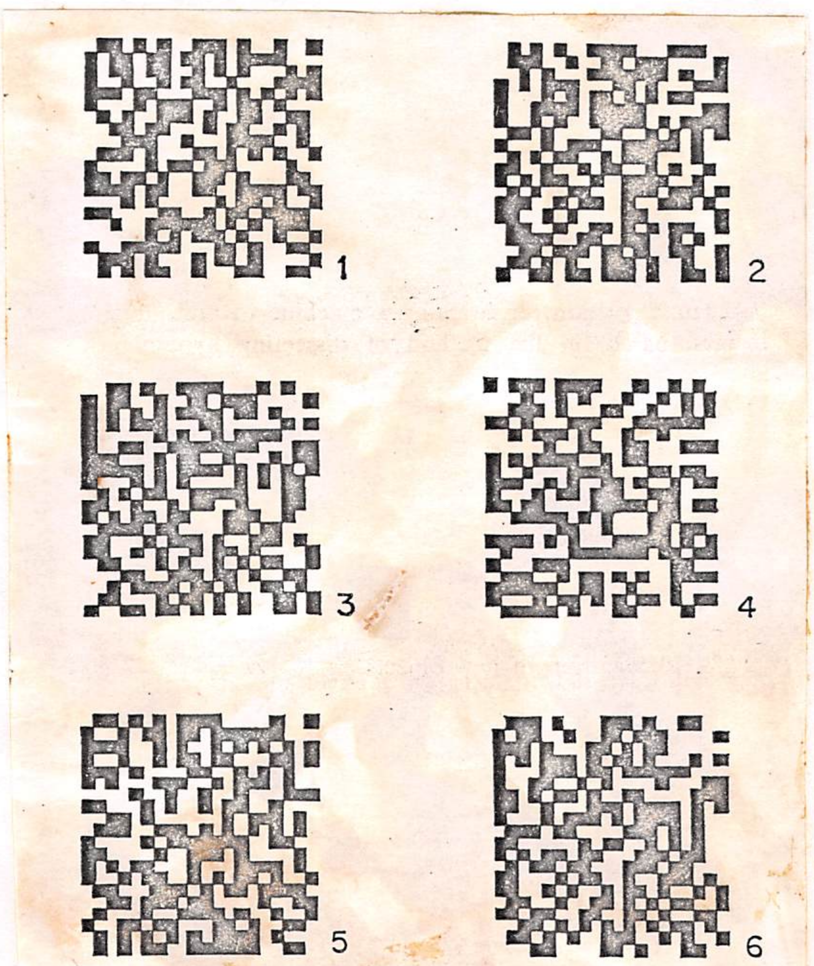
Examples of connected patterns | Examples of unconnected patterns



Since humans can distinguish between connected and unconnected patterns it follows that the eye is not just a perceptron. There are many other properties of human-images which make it seem unlikely that they can be represented as 'compact clusters' in any metric space. For example, the fact that it usually does not change the essential characteristics of the pattern if we translate it around (a 2 at the top of the page is just as much a 2 as a 2 at the bottom of the page, i.e. '2-ness' is position invariant) coupled with the fact that two patterns, one of which is a translate of the other, are not usually near together in P with respect to the sort of metrics we have imposed provides good evidence against any metrical characterisation of pattern similarity. Furthermore, experimental results obtained using perceptrons as pattern recognising devices are not encouraging, for example, it seems that no matter how much 'training' a perceptron is given it can never be taught to get more than about nine out of ten correct recognitions of the numerals. However, although human-images probably do not in general form 'compact clusters' there is experimental evidence (according to A.G. Arkadev and E.M. Braverman) that 'compact clusters' do form human-images, although such images look rather odd in general. Fig. 4 below shows examples from two artificially generated sets of points which form 'compact clusters' in $P \subseteq \mathbb{R}^N$ (where $N = 25 \times 25$) with respect to the Euclidean metric (the degree of blackness of each point has been 'quantized' to either 0 or 1 i.e. white or black).

Fig. 4.

(From "Teaching Computers to Recognize Patterns" by A.G.Arkadev & E.M. Braverman)



2, 3, & 6 come from one compact cluster; 1, 4, 5 from the other.

BIBLIOGRAPHY

Arkadev A.G. and Braverman E.M.: Teaching Computers to Recognize Patterns. Academic Press 1967

Minsky M. and Papert S.: Perceptrons. M.I.T. Press 1969.

Aizerman M.A., Braverman E.M. and Rozonoer L.I.: Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning. Automation and Remote Control (Avtomatika i Telemekhanika) Vol. 25, No. 6. pp. 821-837. June 1964.