

CAPÍTULO 3

REDES NEURAIS ARTIFICIAIS

3. Redes neurais artificiais.

3.1 Apresentação.

Em nosso segundo capítulo trataremos da composição musical gerada por redes neurais artificiais (RNAs). Teremos uma estrutura semelhante ao capítulo anterior, abordando inicialmente os conceitos e noções filosóficos, revisitando o mecanicismo, o funcionalismo e as representações mentais, agora de uma nova perspectiva. Posteriormente, estabelecemos um panorama histórico e taxonômico das principais arquiteturas conexionistas (ou RNAs) e os processos matemáticos envolvidos.

3.2 O estudo da mente no enfoque das RNAs – o paradigma conexionista.

Podemos seccionar as abordagens computacionais da ciência cognitiva em dois principais grupos: a IA e as RNAs. Da mesma forma que a primeira, as RNAs também têm sua origem localizada durante a cibernética do anos 40, novamente pela influência dos trabalhos de Alan Turing. Devido a sua natureza puramente lógica, a MT possibilitou diferentes caminhos implementacionais, um desenvolvido por von Neumann que levou ao CSD e a IA, como vimos, e outro que teve como fundadores o neuropsiquiatra Warren McCulloch e matemático Walter Pitts, que criaram a primeira modelagem conexionista (GARDNER, 1995, p. 159-160).

3.2.1 Contexto histórico.

Vamos nos remeter novamente aos anos 40, época da cibernética, das conferências Macy e do congresso Hixon. Época onde se funda uma ciência da mente que relativamente opõe-se ao behaviorismo e ao mesmo tempo a psicologia popular, ou seja, buscando considerar estados internos direcionadores do comportamento sem o mentalismo comum na psicologia popular (DUPUY, 1995, p. 51). De forma geral, tais

estados internos foram considerados através da noção de *feedback*, formalizada graças aos desenvolvimentos da lógica dos anos 30. Von Neumann, por um lado, propôs o modelo cognitivista da mente pela criação de uma máquina *artificial* que implementa uma máquina *lógica* (uma MT); e como vimos, o CSD não tinha nenhum tipo de correlato com a neurofisiologia de um organismo. Porém, ao longo do desenvolvimento da IA, e de teorias como a do processamento de informação, são bastante freqüentes analogias entre o CSD e o sistema nervoso, especialmente o cérebro; até hoje se encontram abordagens, por exemplo, da percepção auditiva, que postulam modelos explicitamente computacionais (CSD). McCulloch e Pitts, por outro lado, não iriam propor um modelo fisicamente implementado (uma máquina artificial), mas uma máquina lógica fisicamente encarnada, estabelecendo correlatos neurofisiológicos do sistema nervoso central com a lógica. Dessa forma, Dupuy (1995, p. 56) diz que o “número de atores” envolvidos na modelagem cibernética da mente não eram apenas dois, o organismo e a máquina, mas eram três: “(...) o organismo em sua estrutura (o cérebro), o organismo em sua função (a mente) e a máquina, com esta última desdobrando-se em máquina lógica (máquina de Turing ou máquina de McCulloch e Pitts) e máquina artificial, material (o computador).”

Em 1943 McCulloch e Pitts publicam seu referencial artigo *A Logical calculus of the ideas immanent in nervous activity*. É neste artigo que encontramos a primeira caracterização de uma rede de neurônios como um sistema formal, operando de acordo com procedimentos lógicos. Portanto, a máquina lógica de McCulloch e Pitts é inspirada na estrutura e no funcionamento cerebrais, incluindo em sua modelagem os tão importantes mecanismos de *feedback*, os quais os cibernéticos consideravam como dirigentes do comportamento. Apesar da inspiração neurofisiológica deste modelo, sua plausibilidade biológica não é tão pertinente quanto se supunha na época. Para os padrões da atual neurociência tal modelagem é considerada como sendo não realista, ou abstrata, por não incluir muitos dos aspectos e propriedades cerebrais de baixo-nível (CHURCHLAND e SEJNOWSKY, 1992, p. 136). Os autores, neste aspecto, faziam as seguintes “postulações físicas” (grifo nosso) sobre seu cálculo (McCULLOCH e PITTS, 1943, pp.24-25):

- 1) “The activity of the neurone is an ‘all-or-none’ process.
- 2) A certain fixed number of synapses must be excited within the period of latent addition in order to

excite a neurone at any time, and this number is independent of previous activity and position on the neurone.

3) *The only significant delay within the nervous system is synaptic delay.*

4) *The activity of any inhibitory synapse absolutely prevents excitation of the neurone at that time.*

5) *The structure of the net does not change with time.”*

Este tipo de caracterização do funcionamento e estrutura de uma rede de neurônios *simile* não está nem próxima daquelas realmente condizente com a de uma população neuronal natural; a atividade de um neurônio ser do tipo tudo-ou-nada, qualquer atividade inibitória impedir a excitação de um neurônio, a estrutura da rede ser fixa ao longo do tempo, são todos exemplos incongruentes com a neurociência. De qualquer forma, com estes tipos de suposições foi inicialmente desenvolvida a modelagem connexionista da mente, que é equivalente a uma máquina de Turing (DUPUY, 1995, p.59), e ainda leva em consideração a estrutura cerebral. A classe de funções calculáveis por uma rede de McCulloch e Pitts é incluída dentre as funções recursivas, que como sabemos é equivalente as funções Turing-computáveis (DUPUY, 1995, p.65). Mas, contudo, essa equivalência é de certa forma restrita, pois o cérebro como um sistema finito de unidades básicas (neurônios) não tem o mesmo poder computacional de uma MT, que tem uma memória infinita.

Os aspectos gerais da lógica envolvida com a cibernética já foram discutidos (seção 1.2.5.1) e agora, então, vamos brevemente descrever alguns dos aspectos que na IA não foram considerados. Sendo McCulloch um neuropsiquiatra, suas pesquisas foram decorrências de pesquisas anteriores sobre o cérebro. Pesquisadores como Ramon e Cajal, e depois Corente de Lo’, todos da Espanha, já haviam estabelecido no começo do século o funcionamento do cérebro com uma rede de neurônios; desde esta época também já se pesquisava analogias de circuitos elétricos e a álgebra Booleana; e Kubie em 1930 havia provado a existência de *feedbacks* na estrutura cerebral, inclusive despertando um posterior interesse da abnegada psicanálise (pelos cibernéticos) dos anos 40 (DUPUY, 1995, pp.55-60). Na época ainda havia uma grande oposição entre localizacionistas e holistas, sendo os últimos o grupo dos gestaltistas, como Köhler, e, em menor proporção, Lashley, um neurofisiologista. Köhler defendia que para entender-se o funcionamento cerebral precisa-se trabalhar com a física de campo contínuo e não sobre impulsos de neurônios isolados. Tanto Köhler quanto Lashley defendiam uma

perspectiva *top-down* e McCulloch e Pitts com seu modelo neural uma abordagem tipicamente *bottom-up*. Característica essa que continua dominante no conexionismo, ou seja, partir dos elementos básicos para entender o funcionamento do conjunto.

3.2.2 A rede de McCulloch e Pitts (MP)

Os pressupostos neurológicos assumidos por McCulloch e Pitts em sua modelagem foram apresentados logo acima. Vamos, então, estabelecer uma breve descrição do neurônio símile das redes MP. Cada neurônio pode receber várias entradas, inibitórias e excitatórias, que atingindo um certo limiar fazem com que a unidade dispare um impulso (valor binário) adiante.

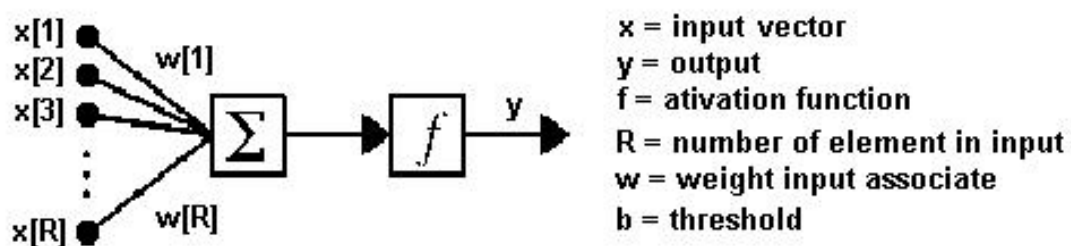


Fig. 3.1. Neurônio típico das redes MP.

Com este tipo de unidade podem ser instanciados vários conectivos lógicos, construindo-se uma rede com um número suficiente de neurônios que tenham limiares adequados. O limiar deste tipo de arquitetura é designado pela quantidade mínima de impulsos excitatórios que uma unidade deve receber. Em seu artigo de 1943 os autores demonstram nove teoremas lógicos possíveis de serem realizados por uma rede MP. Eles ainda afirmam que “*This is of interest as affording a psychological justification of the Turing definition of computability and its equivalents, Church’s λ -definability and Kleene’s primitive recursiveness: if any number can be computed by an organism, it is computed by these definitions, and conversely*” (1943, p.37)¹.

¹ Benante (2001, p.103) ilustra a modelagem lógica de tais redes com o problema do ou-exclusivo.

Ao longo destes quase 60 anos as modelagens conexionistas foram sendo continuadas e inúmeros modelos surgiram, propondo soluções, alternativas e novos paradigmas tanto para a computação quanto para os problemas envolvidos no estudo da cognição. Algumas arquiteturas serão exemplificadas a partir da seção 2.3. Agora, vamos discutir como esta abordagem *bottom-up* do conexionismo foi responsável pela reformulação de conceitos como o mecanicismo, o funcionalismo e as representações mentais, para podermos, depois, prosseguirmos para as propriedades gerais normalmente apresentadas por RNAs.

3.2.3 Mecanicismo e reducionismo.

De fato, a noção de mecanicismo é tão claramente fundamental nas RNAs quanto no modelo cognitivista, mas nas RNAs junto à ela está explicitamente a noção de fisicalismo. Ser um processo mecânico não significa ser necessariamente físico, executado por uma máquina física, mesmo porque a modelagem de McCulloch e Pitts a princípio era apenas lógica. Significa ser um processo regido por leis, que pode ser executado, pelo menos analogamente, num sistema físico. Se entendemos que processos do mundo físico estão sujeitos às leis da mecânica, incluindo suas novas formas, como a mecânica quântica, eles podem ser explicados com termos desta área; da mesma forma que se, talvez, existirem processos não físicos que possuam correlatos físicos ou que são sustentados por objetos físicos, eles podem ser também explicados pelos últimos que, por sua vez, são explicados em termos mecânicos. A isto chama-se *reduccionismo*, reduzir tudo a física. Dupuy (1995) afirma que, devido ao mecanicismo, para a ciência cognitiva, “(...) a física, longe de ser a rival a derrotar, é o modelo a imitar” (p.55).

O modelo científico é uma abstração que procura explicar um processo ou fenômeno, e no caso da IA já se tinha um modelo de antemão proveniente da lógica, que sendo implementado numa máquina física como CSD poderia ter sua pertinência avaliada e, se comprovada, poderia-se postular uma equivalência *funcional* entre os dois sistemas, explicando a mente pelo computador. No caso das RNAs não se partiu de um modelo lógico preestabelecido, mas ele foi extraído do fenômeno físico. Este modelo lógico então podia ser implementado em outro sistema igualmente físico, desde de que

o último simulasse a estrutura e funcionamento do primeiro, e o comportamento dos dois seria equivalente. E ainda, como vimos acima, tanto o modelo lógico da MT quando da RNA se mostraram muito semelhantes, ou mesmo equivalentes. Mas por estabelecer um modelo científico, abstrato, McCulloch não pode ser acusado de dualismo, como muitas vezes o é a IA, pois entendia a mesma coisa por máquina lógica e máquina natural; para ele “a máquina é um ser lógico-matemático encarnado na matéria do organismo” (DUPUY, 1995, p.54). Essa encarnação da máquina na matéria possibilita a visão, além de mecanicista, reducionista do modelo de McCulloch e Pitts:

“Demonstra ele [McCulloch, no artigo de 1943], em princípio, a existência de uma máquina lógica equivalente à de Turing (tendo em vista que tudo que uma pode fazer, a outra pode fazer, e reciprocamente), que pode ser considerada em sua estrutura e em seu comportamento uma idealização da anatomia e da fisiologia do cérebro. Este resultado, na mente de McCulloch pelo menos, constitui um avanço decisivo, já que não é mais o cérebro em sua função (a mente), mas também o cérebro em sua estrutura, o cérebro material, natural, biológico, que é assimilável a um mecanismo e, mais precisamente, a uma máquina de Turing. É assim que o cibernético julga resolver o velho problema da alma e do corpo, ou, em seus próprios termos, do “*embodiment of mind*”. O cérebro e a mente são um e outro uma máquina, e a mesma máquina. O cérebro e a mente, portanto, são um só.” (DUPUY, 1995, p.59)

As modelagens tanto inspiradas na lógica (IA) quanto inspiradas na biologia (RNAs) compartilham a idéia de que a noção de mecanicismo é fundamental para a edificação de suas teorias. Mas as equivalências entre noções filosóficas param por aí.

3.2.4 Funcionalismo.

Como vimos no primeiro capítulo, a noção de funcionalismo, ambos científico e ontológico, sustenta uma equivalência *funcional* entre dois sistemas com distintas realizações físicas. O dualismo de propriedades, ou monismo anômalo, utiliza a noção de funcionalismo, e até depende dela. Como assim o faz a IA, principalmente a forte. Pois, sem o funcionalismo não se pode afirmar que um sistema como um CSD é a realização de uma mente. Mas no caso das RNAs esta noção terá que ser reformulada. Como pudemos ver logo acima, para McCulloch a mente e o cérebro são a mesma

coisa, e, portanto, o modelo lógico abstraído da biologia é dependente da estrutura e funcionamento do cérebro e seus componentes. Sendo dependente, para uma mente ser realizada em outro sistema fisicamente diferente, este deve simular não as propriedades lógicas análogas, mas tem que simular a estrutura e funcionamento da primeira, e as analogias lógicas serão consequência disso. Portanto, aquilo que chamamos de estados mentais não são simplesmente propriedades lógicas, geradas por e que resultam em comportamentos ou outros estados mentais, que independem da natureza física que as sustentam. Ao contrário, tais estados são encarnados na física, dependem da matéria e das leis mecânicas que a governa. Dessa forma, o funcionalismo não pode ser defendido pelo connexionismo em sua forma ordinária.

A este respeito Gonzalez (1989) estabelece duas formas de funcionalismo: o lógico-computacional e o neuro-computacional. O funcionalismo lógico-computacional é aquele decorrente da IA, segundo Dupuy (1995, p.26), e que tem como principal autor Putnam (1980a e 1980b) (cf. Parte 1.2.3 do presente trabalho). No caso das RNAs temos o funcionalismo neuro-computacional, que diz, em suma, que estados mentais são parcialmente dependentes de um sistema físico específico, o que impossibilita a hipótese das realizações múltiplas de Putnam. É importante frisar o termo parcialmente na afirmação acima, já que se estados mentais forem totalmente dependentes da realização física específica, o cérebro, nenhum outro sistema pode apresentar tais estados a não ser que seja outro cérebro. Putnam defende o funcionalismo ortodoxo (lógico-computacional), mas alguns de seus exemplos são próximos do neuro-computacional, como seu famoso exemplo da terra gêmea, ou o exemplo da nação chinesa dado por Block (1980)². Em tais exemplos se defende uma visão que parece postular a emergência de uma mente devido a complexidade estrutural e funcional da realização física, mas não propriamente da matéria da realização, o que, nos parece, ser compatível com o que Gonzalez chama de funcionalismo neuro-computacional. Portanto, para o connexionismo o funcionalismo é outra noção importante, mas agora alterada, sendo que estados mentais são parcialmente dependentes da realização física, dependentes da complexidade que tal realização apresenta. O que nos permite dizer que, se uma RNA apresentar uma simulação complexa o suficiente, ela pode apresentar

² Nestes exemplos existe a atribuição de estados mentais a sistemas formados por diferentes realizações físicas, como a população da china onde cada indivíduo se comportaria como um neurônio, formando um gigantesco cérebro, ou um planeta alienígena onde minúsculos *marcianos* em diminutas naves simulariam o comportamento de nossas partículas elementares, e se astronautas passassem a viver em tal planeta,

estados mentais. Isso ainda parece pertinente com a distinção de Churchland e Sejnowsky (1992, pp.136-137) de rede neurais realistas e abstratas, para a neurociência. Mas segunda ela, se o que nos interessa for a explicação da mente/cérebro, um modelo muito realista pode não ser interessante: *“models that are excessively rich may mask the very principles the models where built to reveal. In the most extreme case, if a model is exactly as realistic as the humam brain, then the construction and the analysis may be expensive in computational and human time and hence impractical”* (1992, p.137). Acreditamos que eles, os modelos realistas, seriam interessantes na medida em que se quer construir uma mente artificial, mas para explicações científicas sem esta pretensão Churchland et al. dizem que o importante é: *“(...) make the model simply enough to reveal what is important, but rich enough to include whatever is relevant to the measurements needed”* (1992, p.137).

3.2.5 Representações mentais.

Representações mentais podem ser entendidas como objetos internos a nossa mente que representam coisas que estão fora dela. Na perspectiva da IA, vimos que elas são entendidas como tendo uma natureza estritamente simbólica, ou seja, não há correspondência direta entre o símbolo e o que ele representa, a não ser através de convenções arbitrárias. Um dos problemas da representação mental simbólica é que ela não apresenta correlações entre o nível simbólico de processamento e o nível físico de implementação, mesmo por estar ligada ao funcionalismo que defende a idéia de realizações múltiplas. E ainda, representações simbólicas apresentam uma correlação um-para-um entre o que é representado e o que representa (HASELAGER, 1997, p.45). Pois, de maneira semelhante à reformulação conexionista do funcionalismo, que passou a defender que o nível mental é dependente da estrutura do nível material e dos processos que ali ocorrem, as representações mentais serão também reformuladas.

Para as RNAs, as representações usualmente denominadas de representação distribuída. Da década de 50 até a atualidade as neurociências gradativamente deixaram a teoria da codificação local (*local coding theory*) em função representação distribuída

com o passar do tempo seus corpos seriam compostos em vez de átomos terráqueos de seres alienígenas. Devido a isso negaríamos estados mentais a eles, pergunta Putnam com esse exemplo?

(ou *vector coding* ou *multidimensional representation*) (CHURCHLAND e SEJNOWSKY, 1992, pp.157-162), e essa nova postura é, inclusive, condizente com as arquiteturas de RNAs. Em primeiro lugar, a representação distribuída ocorre sobre um conjunto ou população de neurônios que representam um objeto (enquanto que no *local coding*, cada único neurônio seria responsável por representar uma característica específica do objeto). Ou seja, a representação não é um objeto único, mas está espalhada sobre um conjunto deles interagindo, formando um espaço multidimensional. Outra característica importante de ser destacada é que a população de neurônios é sempre a mesma para diferentes representações, fato que van Gelder chama de sobreposição (1991). E nenhum conteúdo de um objeto é representado por um único neurônio, como aponta Haselager (1997, p.45) em acordo com van Gelder (1992). Quanto a isso, Churchland e Sejnowsky (1992, p.165) discordam parcialmente: “(...) *in some instances, the vector components may correspond to features such as the color, shape, motion, etc. of an object, but in other cases there may be no identifiable microfeature correspondig to any component.*” De qualquer forma, outra característica importante das representações distribuídas é que elas apresentam redundância, visto que várias unidades, em geral, representam a mesma coisa. Isso leva a vantagens como redução da taxa de sinal-ruído da membrana celular, amplificação de sinais fracos pela atuação conjunta da várias células, e tolerância à falha, isto é, que a destruição de uma das células não impede o funcionamento da rede como um todo (CHURCHLAND e SEJNOWSKY, 1992, p.165).

Um dos fatos que apontamos acima é o da sobreposição de diferentes representações na mesma população neuronal. Pois bem, falta explicar como é que as representações são mantidas na rede enquanto não estão ativadas. Num dado momento existe uma configuração de ativação dos neurônios envolvidos, e isto é a representação distribuída de algo. Mas esta mesma população neuronal também pode representar *n* outras coisas em outros momentos. Estas outras representações são outros padrões de ativação, que são armazenadas na forma de pesos sinápticos, ou configurações de pesos. Cada neurônio de uma rede tem ligações com outros neurônios, chamadas de conexões sinápticas, e os sinais enviados de um para outro podem ser mais fortes ou mais fracos. Tais pesos possibilitam a recuperação de outros padrões de ativação perante um estímulo específico para determinado padrão. Através destes processos podem ser armazenados inúmeros padrões.

Neste momento, tendo passado pelas noções de mecanicismo, funcionalismo e representação mental, na perspectiva do connexionismo, acreditamos que podemos seguir para uma descrição geral do modelo conexionista da mente, falando de aspectos que ainda não tratamos, tal qual a aprendizagem, poder de generalização, poder de classificação, entre outras.

3.2.6 Propriedades dos modelos conexionistas.

Os modelos conexionistas, além da inspiração biológica, apresentam algumas propriedades enquanto sistemas computacionais. Algumas delas são compatíveis com a atividade cerebral, tendo-se em conta que normalmente são modelos simplistas ou abstratos deste. Outras características não o são de maneira tão explícita, mas demonstram grande poder computacional, são ferramentas robustas computacionalmente.

Segundo Churchland e Sejnowsky (1992, p.171), a computação de uma RNA é vista como um processo de convolução de vetores por uma matriz (gerando novos vetores), onde os vetores são o espaço multidimensional de ativação (representação atual) e a matriz o espaço multidimensional dos pesos (representações armazenadas). Isto é uma transformação vetor-vetor, onde a arquitetura da rede com os pesos são o substrato da transformação. Dessa forma a informação é integrada e transformada. Os autores ainda afirmam (1992, p.171), sem deixar de frisar a visão ‘simplista’ destes modelos artificiais, que “*in nervous systems the weight configuration is a kind of living matrix*”. Ainda, esta matriz viva é de uma ordem de entre 10^{11} a 10^{14} neurônios, com entre 10^3 e 10^4 conexões cada um, números incomparáveis com as atuais arquiteturas de RNAs.

3.2.6.1 Aprendizagem.

Uma das mais importantes características de uma RNA é o processo de aprendizagem³, também chamado de fase de treinamento da rede. Ou seja, a rede pode ‘aprender pela experiência’. O processo que possibilita tal capacidade é o fato de que a rede pode, por meio de algoritmos, reconfigurar os pesos de suas conexões. Isso é uma modificação do espaço multidimensional dos pesos, que possibilita uma melhor codificação de todas as representações sobrepostas, efetivando uma mais adequada transformação de vetores de entrada em vetores de saída. O espaço dos pesos é alterado até que se chegue a melhor configuração, mas a determinação deste ponto do espaço dos pesos só pode ser determinada na prática, isto é, não tem como se determinar sua localização previamente.

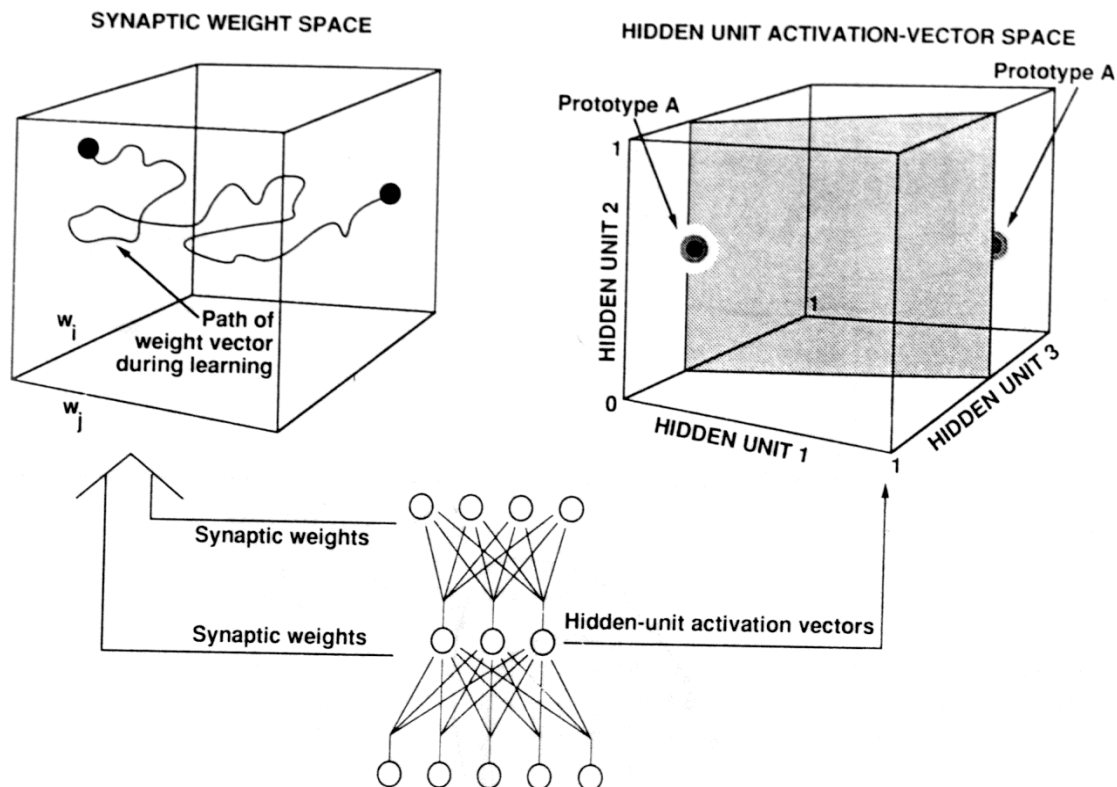


Fig. 3.2. alteração do espaço multidimensional dos pesos pela aprendizagem (In: CHURCHLAND e SEJNOWSKY, 1992, p.169).

É o processo de treinamento ou aprendizagem bem sucedido que garante as outras propriedades que iremos expor em seguida. Tal processo de alteração na configuração da matriz dos pesos pode ser por dois modos: aprendizagem supervisionada e não supervisionada.

- **Aprendizagem supervisionada.**

Nas arquiteturas supervisionadas o desempenho da rede é avaliado por critérios estabelecidos por um supervisor externo. Para cada conjunto de vetores de entrada o supervisor estipula o conjunto de vetores de saída correspondente. Após a fase de treinamento a rede tem que estar apta a correlacionar as entradas com as saídas corretas. Esse processo é realizado por algoritmos especialmente concebidos, e entre eles temos, por exemplo, o *back-propagation* e o *hopfield*. Normalmente, os algoritmos de treinamento operam sobre um valor de *erro* entre a saída obtida e a desejada, calculado por funções matemáticas variadas específicas para cada arquitetura. O processo tende a minimizar o erro pela alteração dos pesos das conexões. Segundo Marar (1997, p.73), “as técnicas de treinamento supervisionado incluem: decidir quando desativar a aprendizagem, decidir como e por quanto tempo apresentar os pares de treinamento [padrões de entrada e saída], além, de suprir informações de erro e performance.”

Na avaliação da rede perante o erro apresentado pode-se determinar locais no espaço dos pesos que geram o menor erro possível, chamado de *mínimo global*. Mas não existe como prever onde estará este ponto no espaço dos pesos. Devido a este fato, muitas vezes na etapa de treinamento a rede pode se encontrar num valor mínimo, mas que posteriormente poderá se perceber que ele não era o mínimo global, mas um local. Para se sair de um mínimo local não existem procedimentos algorítmicos; porém, um bastante comum é a inclusão de ruído nos vetores de entrada para que a matriz dos pesos seja desestabilizada e novamente treinada. Isso não garante que se caia num mínimo global, podendo novamente a rede estabilizar num local. Uma ilustração destes pontos é mostrado abaixo:

³ Estamos nos referindo apenas ao processo de aprendizagem de máquinas conexionistas.

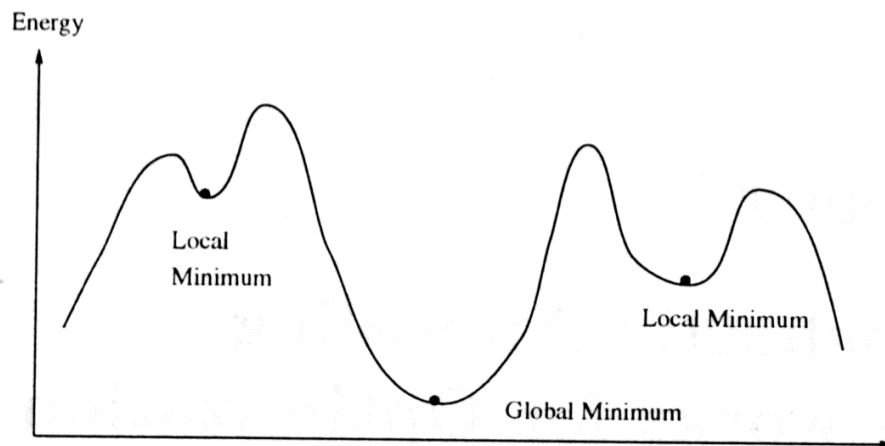


Fig. 3.3. mínimo global e mínimo local (In: KUNG, 1993, p.312).

- **Aprendizagem não supervisionada.**

Nas arquiteturas não supervisionadas não existe um supervisor que determinará a saída correta de uma rede para uma dada entrada. Tal processo é também chamado de auto-organizado, onde as informações úteis estão todas contidas no conjunto de entrada e na matriz de pesos continuamente alterada durante o treinamento. Para cada padrão do conjunto de treinamento a rede auto-organizada gerará uma saída pelas propriedades coletivas dos padrões (MARAR, 1997, p.73). Segundo o mesmo autor, “o processo de treinamento extrai propriedades estatísticas do conjunto de treinamento e agrupa vetores similares em classes”. Por essa característica de agrupamento de dados por meios estatísticos as redes não supervisionadas são muito aplicadas como classificadores, ou extratores de características de um conjunto de amostras. Duas arquiteturas típicas das RNAs não supervisionadas são a ART (*Adaptive Resonant Theory*) de Grossberg e SOM (*Self-organizing Maps*) de Kohonen. (SIMPSON, 1990, pp.13-18)

3.2.6.2 Generalização.

A capacidade de generalização é fundamental para as aplicações que utilizam RNAs. É devido a esta capacidade que a rede pode se comportar adequadamente perante estímulos a que nunca foi apresentada. Pois, a capacidade de generalização está associada a um *processo indutivo*, onde a partir de um conjunto de amostras a rede é treinada e, tendo convergido (atingido uma situação estável e adequada próxima ao

mínimo global), então, pode ser utilizada perante amostras ausentes deste conjunto inicial. Ou seja, ela pode induzir fatos baseada em outros semelhantes e já conhecidos, o que podemos chamar de generalização.

Esta propriedade é uma das mais interessantes das RNAs, e é encontrada em todas as arquiteturas conexionistas. Em relação aos sistemas da IA, como vimos, um dos meios para obter-se generalização em tais sistemas é através de cadeias de Markov, mas estas também são facilmente implementadas em redes neurais, como as Markov model ou Hidden Markov model (KUNG, 1993, pp.239-268). Mas as RNAs vão além de processos de Markov, e podemos afirmar que arquiteturas conexionistas apresentam generalização não apenas para determinação de um espaço probabilístico condicionado por eventos anteriores (como cadeias de Markov) mas para qualquer tipo de procedimento envolvendo categorização e classificação.

3.2.6.3 Classificação e extração de características.

Processos de classificação e categorização são uma das mais freqüentes aplicações de RNAs. Para este fim é determinado um espaço multidimensional de ativação, onde partes deste espaço correspondem às categorias de objetos, similares em suas características. “(...) *similarity between objects represented can be reflected by similarity in their representations, that is, proximity of positions in activation space*” (CHURCHLAND e SEJNOWSKY, 1992, p.169). A eficiência da classificação depende da eficiência do algoritmo de treinamento utilizado, resultando numa classificação mais grosseira ou com mais acuidade. Num nível comportamental, conforme afirmam Churchland e Sejnowsky (1992, p. 169), isso possibilita que humanos possam distinguir entre maçãs e plumas, ou entre diferentes tipos de maçãs vermelhas. Um exemplo do espaço de ativação para distintas categorias é dado a seguir.

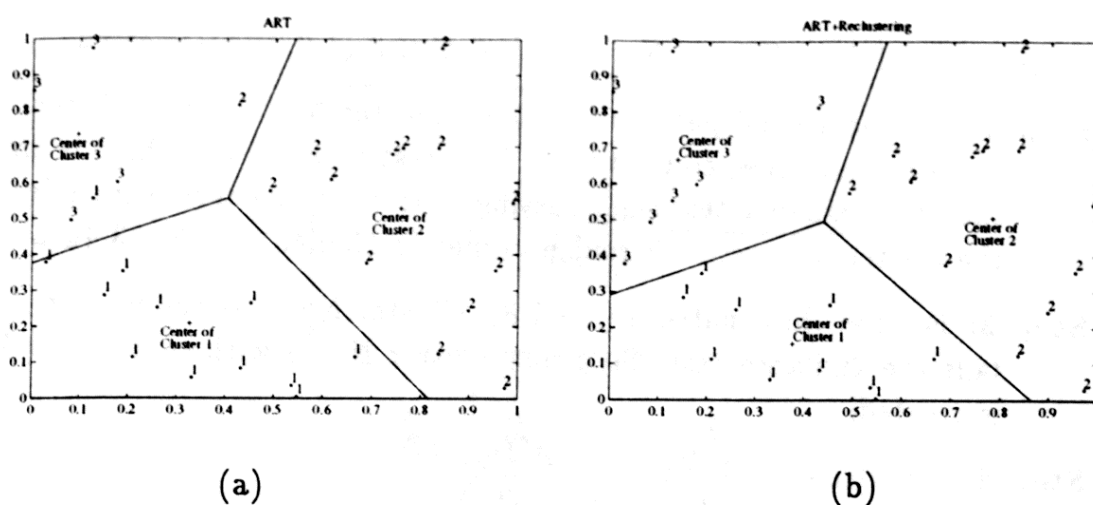


Fig 3.4. espaços de ativação com três categorias (In: KUNG, 1993, p.85).

Muitas vezes são necessários processos que extraem as características desejadas das amostras, estes processos podem ser realizados externamente a rede, como um pré-processamento, ou pode se projetar uma rede especialmente para este fim. Existem arquiteturas de redes compostas de sub-redes, onde uma delas pode realizar a extração de características e outra faz a classificação das amostras. Procedimentos de extração de característica são uma parte importante do conhecimento estatístico e matemático, existindo várias técnicas para este fim. Uma delas, e comumente aplicadas à música, é a transformada de Fourier (ou um algoritmo específico chamado de FFT – *fast Fourier transform*). Existem outras inúmeras técnicas para este fim, inclusive mais robustas computacionalmente, a transformada de Karhunen-Loève (MARAR, 1993), transformadas *wavelet* etc. (MARAR, 1997). O principal objetivo destas funções é a redução do espaço dimensional utilizado para a representação do objeto, extraíndo ‘formas bem comportadas’ de baixa dimensionalidade que mantenham as características importantes do objeto. Isso é interessante pois ao mesmo tempo requer menos memória para armazenar as matrizes de valores e também torna a análise delas menos dispendiosa computacionalmente. Existem RNAs específicas para todos estes métodos de análise, como *wavenets* (para *wavelets*) (MARAR, 1997) ou *principal component neural network* (KUNG, 1993), ou RNAs de Fourier (MARAR, 1997, p. 80).

3.2.6.4 Reconhecimento e reconstituição de padrões estáticos e temporais.

Se RNAs podem classificar objetos certamente podemos dizer que elas reconhecem e distinguem padrões, sejam eles estáticos ou temporais. Se uma rede reconhece adequadamente um padrão é porque ela consegue distinguir características comuns ou similares entre padrões diferentes. Pois, uma rede bem treinada deve ter uma representação adequada e de mais baixa dimensionalidade possível de cada tipo de padrão, de forma que, inclusive, além de classificar corretamente ela pode reconstituir vetores de entrada incompletos ou ausentes. Churchland e Sejnowsky (1992, p.173) nos dão um divertido exemplo para ilustrar esta capacidade de reconstituição:

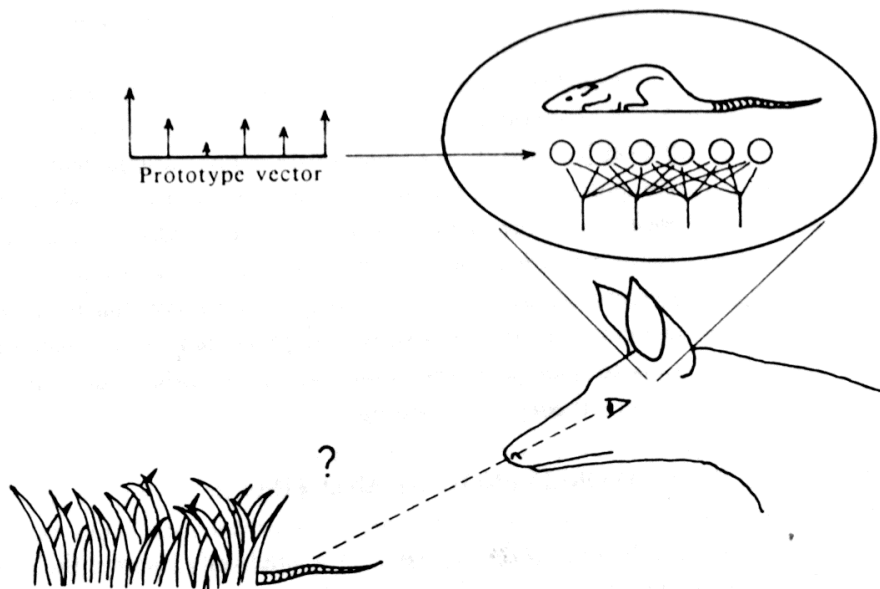


Fig 3.5. Ilustração de idéia de reconstituição de vetores. O *input* para a retina relativo ao rato está limitado à luz refletida apenas do rabo do rato. Porém, o sistema visual do coite completa o vetor, permitindo ao coite suspeitar que existe um rato na vegetação.

Existem dois tipos de reconhecimento e reconstituição de *inputs*: estáticos e temporais. Padrões estáticos estão fora do domínio do tempo, de forma que sua caracterização é completa num dado momento. Nesse sentido uma única configuração de ativação representa completamente o padrão. Mas, no caso de necessidade de reconstituição de inputs incompletos ou distorcidos por alta taxa de sinal-ruído, uma rede bem treinada pode em poucos passos (ou ciclos) recupera-lo. Para tal atividade a maioria das redes *multilayer* (multicamadas) convencionais, assim como redes de

Hopfield e Boltzmann Machines são arquiteturas adequadas (KUNG, 1993, p.9). Um exemplo dessa habilidade é dado a seguir, com uma rede Hopfield (LIPPMANN, 1987, p.9):

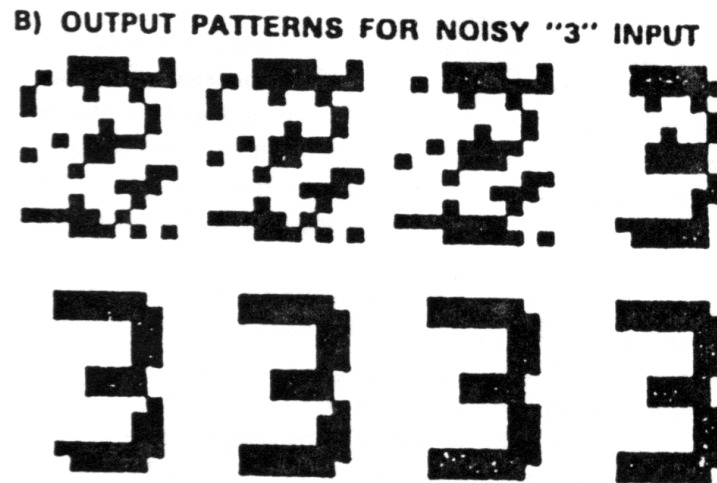


Fig. 3.6. Comportamento de uma rede Hopfield para a recuperação de um input com ruído, em sete iterações.

O mesmo processo é também possível sobre padrões dependentes do domínio do tempo. Estes padrões, para serem apresentados completamente, necessitam de n passos, pois podemos vê-los como uma sucessão de várias matrizes com os vetores de entrada. Ou seja, são padrões não estáticos temporalmente, incorporando as mudanças dos vetores sobre o fluxo temporal para a sua representação completa. Considerando M_t como um conjunto de vetores de entrada no instante t , o padrão completo P será $P = \{M_t, M_{t+1}, M_{t+2}, M_{t+i}\}$, onde $i = n - 1$, e n é a taxa de amostragem. Podemos entender P como uma seqüência de observações de um objeto. No caso da reconstituição de *inputs* incompletos, dado apenas uma parcela dos *samples* de um padrão do tipo P , a rede pode reconstruir aqueles que faltam recuperando o padrão completo. Este tipo de procedimento é extremamente adequado quando se lidando com processamento, classificação e redução de ruído de sinais, como em formas de ondas provenientes de EEG, ECG (eletrocardiogramas) ou sinais de áudio em geral. Arquiteturas adequadas para este tipo de padrão são, por exemplo, *Markov models* e *time-delay dynamic networks* (KUNG, 1993, p.9).

3.2.6.5 Aproximações de funções – interpolação e extrapolação.

Tendo visto que RNAs podem ser aplicadas à reconstituição e classificação de padrões, logo podemos entendê-las como aproximadores universais de funções. Nesse sentido, as RNAs devidamente treinadas apresentam propriedades tanto de interpolação quanto de extrapolação de funções. Por exemplo, fornecendo apenas alguns pontos discretos uma RNA pode nos oferecer como saída uma função que melhor interpole tais valores, ou, de maneira semelhante, ela pode extrapolar a função gerando valores não determinados previamente, num processo indutivo. Em Marar (1997) temos uma extensiva descrição de arquiteturas supervisionadas aplicadas para aproximações de funções lineares contínuas. Para isto são desenvolvidas famílias (wavelets, radiais) de funções que apresentam o melhor desempenho perante o mínimo erro quadrado versus o número de funções de ativação utilizadas. Marar (1997, p.153) nos dá um exemplo de redes PPS (polinômios potências de sigmóide) utilizadas para a aproximação da forma de onda de sinais de fonemas vocálicos. Este tipo de procedimento pode interessantemente ser aplicado na composição musical computacional, possibilitando a síntese sonora de formas de onda complexas.

Como aproximadoras de funções as RNAs podem estar envolvidas em tarefas de reescalonamento de formas de ondas sobre o domínio do tempo. Abaixo um exemplo de HMM (*Hidden Markov Models*) utilizada para a reescalonamento temporal de formas de onda de ECG amostradas a $60/T$, onde T é o período de um pulso de ECG (KUNG, 1993, p.259):

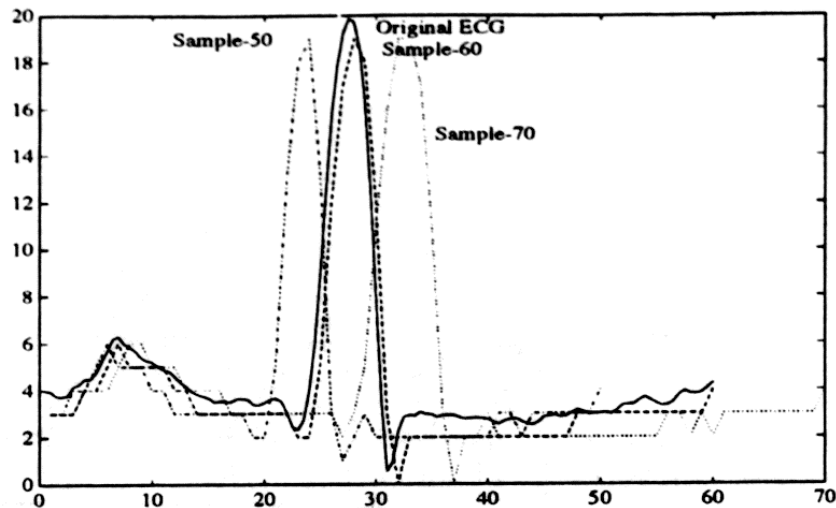


Fig. 3.7. Padrão original (linha sólida) mais três sinais re-escalados temporalmente a $\frac{50}{T}$, $\frac{60}{T}$ e $\frac{70}{T}$.

3.2.6.6 Tolerância à falha.

Por estarem situadas no paradigma PDP (parallel and distributed processing) o poder computacional não está situado sobre uma única unidade de processamento, mas sobre um conjunto delas. Além das questões concernentes às representações distribuídas e sobrepostas em tais sistemas, temos a propriedade de que arquiteturas conexionistas apresentam tolerância a falha, pois se uma das unidades falha ou é de alguma forma destruída a rede ainda pode ter seu poder computacional disponível pelas demais unidades. A eficiência da rede depende da extensão dos danos, como apontam Churchland e Sejnowsky (1992, pp.172-173) com relação ao sistema nervoso:

“Vector-coding and vector-matrix computation are robust; that is, they display functional persistence in the face of damage. Dropping a few neurons from the vector basically just leaves a minor vector-completion task. Dropping many neurons probably means that a recognition impairment can be observed in behavior, such as loss of face recognition (prosopagnosia) or loss of visual recognition of objects (visual agnosia). With relatively minor damage, a network can continue to work, albeit with some loss in accuracy, but it is not stopped dead in its tracks by tiny lesions. Major damage means one route through the matrices is closed, and hence someone might have normal vision acuity, but fail utterly to recognize visually that an object is a dog or a clock. Nonetheless, the sound of barking or

ticking gives instant recognition, because the route for the auditory information is uninterrupted, and auditory vectors are transformed quite normally."

No fim da citação acima ainda percebemos menção a outro ponto relacionado ao processamento paralelo e distribuído: *redundância*. Estando uma representação espalhada sobre um conjunto de unidades processadoras, elas apresentam um certo grau de redundância perante a quantidade total de informação de um padrão.

3.3 Taxonomia das RNAs.

Ao longo da história das RNAs diversas arquiteturas foram criadas com diferentes propósitos e métodos a partir do modelo da rede MP. Com o intuito de clarificar as diferenças e semelhanças entre elas podemos estabelecer categorizações com relação a certas características, como o sentido informacional através da rede, aprendizado supervisionado e não supervisionado, *inputs* binários e de valores contínuos, funções de ativação e de base (*basis function*), número de camadas e estruturas de conexões *intralayer* e *interlayer*, padrões estáticos e temporais, formulações de decisão ou de aproximação/otimização, além de outras possíveis categorias. Lippmann (1987, p.6) propõe uma taxonomia básica das RNAs perante alguns destes pontos, conforme figura abaixo:

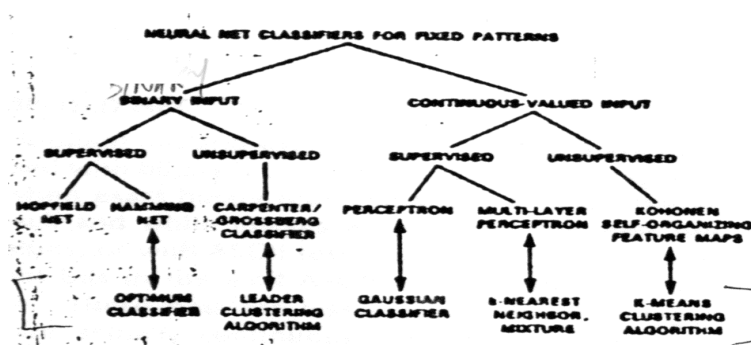


Fig. 3.8. Taxonomia de Lippmann (1987).

Kung (1993, seção 1.3) nos apresenta uma taxonomia mais complexa sobre seis diferentes aspectos; Churchland e Sejnowsky (1992, cap. 3) descrevem de maneira mais

detalhada algumas das possibilidades arquitetônicas de RNAs. Neste intuito, agora, rapidamente passaremos por algumas delas.

- **Aprendizado supervisionado, não supervisionado e pesos fixos.**

As diferenças entre arquiteturas supervisionadas e não supervisionadas já foram apresentadas na seção 2.2.6.1. Apenas a título de recapitulação podemos dizer que o processo de aprendizado, seja ele qual for, ocorre sobre os pesos das conexões, alterando-os em busca de um melhor desempenho perante algum critério. Além destas duas possibilidades existem ainda redes com pesos fixos onde não há aprendizagem (KUNG, 1993, pp.25-26); a rede MP e a de Hopfield são exemplos. Para as RNAs supervisionadas o principal algoritmo de alteração de pesos é conhecido como *back-propagation*. Tal algoritmo foi desenvolvido para as redes MLP (*Multilayer Perceptron*) e consiste num processo de retrogradação do erro pela rede com o intuito de minimizar a diferença entre a saída desejada e a saída obtida (pelo critério do mínimo erro quadrado).

- **Funções de base e funções de ativação.**

Funções de base e de ativação são respectivamente as funções que representam a saída da rede como um todo (*net function*) e a saída de um neurônio (*neuron function*). Existem dois tipos de funções de base: lineares e radiais. Uma função linear é uma função do tipo hiperplano, onde a saída da rede é a combinação linear dos *inputs* e dos pesos, separando linearmente o espaço de saída.

$$u(w, x) = \sum_{j=1}^n w_j x_j$$

onde $u(w, x)$ representa a saída da rede, sendo w a matriz dos pesos e x a dos vetores de entrada; os pesos entre as conexões é dado por w_{ij} e x_j representa as entradas de um neurônio. As funções radiais são do tipo hipersfera, e sua saída não linear representa a distância para um padrão referencial.

$$u(w, x) = \sqrt{\sum_{j=1}^n (x_j - w_j)^2}$$

A manipulação da função de base é executada pela função de ativação. As mais comuns delas são a linear, *step* (ou *hard-limiter*), rampa (ou *threshold logic*), sigmóide e gaussiana. Abaixo são mostradas estas funções de ativação.

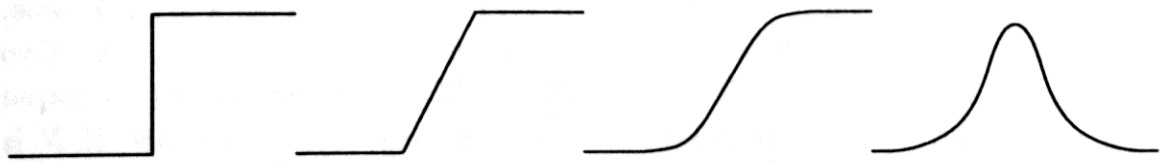


Fig. 3.9. Funções de ativação *step*, rampa, sigmóide e gaussiana.

- **Estruturas de redes neurais**

RNAs podem ter três tipos de camadas (ou *layers*), a camada de entrada (*input*), a camada escondida (*hidden*) e a de saída (*output*). O tamanho de uma rede é determinado pela quantidade de camadas e pelo número de neurônios em cada camada. Estas camadas podem ser conectadas de quatro formas:

feedforward – neste tipo de conexão a informação dos neurônios de uma camada é passada a frente para as camadas seguintes.

feedback – o percurso da informação segue o caminho contrário do tipo anterior, sendo retrograda pelas camadas da rede.

lateral – este tipo de conexão propaga a informação entre os neurônios de uma mesma camada da rede. Um exemplo típico é a rede SOM.

time-delay – informações de momentos anteriores da rede são incorporadas à informações do momento atual da rede.

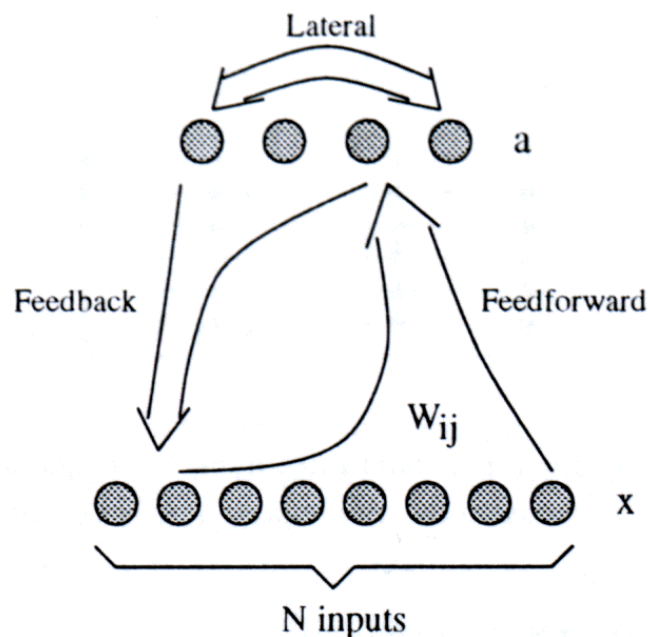


Fig. 3.10. tipos de conexões (In: Kung, 1993, p.31).

Ainda existem dois tipos de estruturas (KUNG, 1993, p.32): ACON (*all-class-in-one-network*) e OCON (*one-class-in-one-network*). As redes do tipo OCON são uma grande rede dividida em sub-redes menores, onde cada uma destas responde para uma ou poucas classes de padrões. Um caso extremo é quando temos apenas uma classe por subnet. Nas redes ACON não existem subdivisões na arquitetura da rede. Dependendo do tipo de problema e da quantidade de padrões que a rede deve reconhecer um ou outro tipo é mais adequado. Quando esta quantidade é alta redes OCON são mais interessante pois requerem um menor número de neurônios na camada escondida para cada classe (normalmente entre 3 e 5), além de apresentarem menor tempo de treinamento e processamento. Mesmo a quantidade total de neurônios de uma rede ACON ou OCON sendo igual, na última é relativamente menor o número de conexões. Porém, o processo de treinamento de uma ACON é mais simples, normalmente pelo algoritmo *back-propagation*.

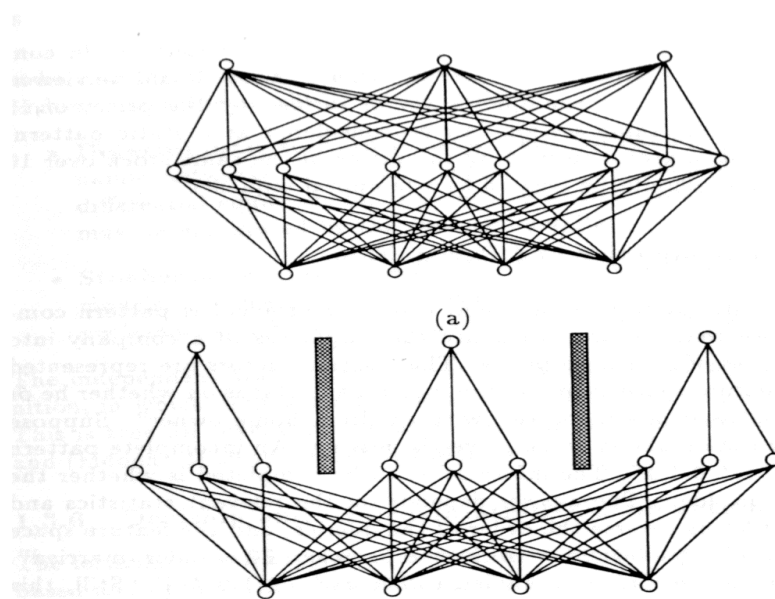


Fig. 3.11. Estruturas ACON e OCON (In: KUNG, 1993, p.34).