

# UMA CARICATURA FUNCIONAL DE REDES NEURAIS ARTIFICIAIS

Fernando J. Von Zuben

DCA/FEEC/Unicamp  
Caixa Postal 6101 – CEP: 13083-970 – Campinas – SP – Brasil  
e-mail: vonzuben@dca.fee.unicamp.br

**Resumo** – Redes neurais artificiais são sistemas de processamento de informação formados pela interconexão de unidades simples de processamento, denominadas neurônios artificiais. Os neurônios artificiais recebem essa denominação porque foram originados a partir de um modelo matemático de um neurônio natural. Além de sua natureza multidisciplinar, a computação inspirada em arquiteturas conexionistas não emprega os mesmos paradigmas predominantes na ciência da computação. A motivação que está por trás deste paradigma alternativo de processamento computacional é a possibilidade de elaborar soluções eficazes para problemas de difícil tratamento com base na computação convencional. O avanço verificado nos últimos anos junto à teoria de redes neurais artificiais tem levado invariavelmente ao desenvolvimento de ferramentas de engenharia mais eficazes e à utilização mais eficiente dos recursos computacionais hoje disponíveis, o que implica uma ampliação sem precedentes na capacidade de manipular informação. O grande potencial das redes neurais artificiais só pode ser devidamente explorado com o emprego de procedimentos refinados de análise e síntese, requerendo assim um esforço adicional por parte dos usuários no sentido de aplicar os recursos de processamento disponíveis na medida certa e na situação apropriada. Existem muitas formas de se auxiliar potenciais usuários ou estudiosos da área, dentre as quais se destacam textos com perfil tutorial, os quais buscam posicionar a área de pesquisa frente a todas as demais vertentes de pensamento e procuram também elucidar as principais características dos modelos e das técnicas de aplicação que podem ser adotadas. Este também é o propósito deste texto. No entanto, o estilo de apresentação vai salientar alguns aspectos em detrimento de outros, sempre voltado para a funcionalidade das ferramentas computacionais envolvidas e para a interpretação dos problemas de acordo com perspectivas mais adequadas para tratamento via redes neurais artificiais. A natureza caricaturesca da abordagem está justamente na ênfase desproporcional atribuída a aspectos igualmente relevantes.

**Palavras-chaves:** redes neurais artificiais, análise multivariada de dados, aproximação de funções, auto-organização, dinâmica de relaxação.

## 1 Introdução

Este texto tem a pretensão de apresentar redes neurais artificiais sob um ponto de vista alternativo. Decidiu-se ‘inventar moda’ em lugar de ‘reinventar a roda’. E existe uma explicação para esta iniciativa.

Textos introdutórios e auto-contidos para redes neurais artificiais existem em abundância. Incluem livros, teses, tutoriais, minicursos, artigos de revisão, FAQs (*Frequently Asked Questions*) e muitos outros. Como se está incluindo nesta lista material que não passou por um corpo de revisores, nem todos apresentam atestado de qualidade. De forma genérica, todo texto com atestado de qualidade passará a ser denominado aqui de texto clássico. Quase todos buscam ser auto-contidos e apresentam uma longa lista de referências bibliográficas.

Em contraste, o presente texto segue o caminho oposto: busca a dependência completa e não faz sequer uma menção direta à literatura. Isto não implica que se trata de material totalmente original. Também não é contraditório frente aos textos clássicos, até porque depende deles para se completar. Portanto, não existe nenhum radicalismo. Apenas não se conta a história de forma convencional e se busca enfatizar alguns aspectos funcionais em detrimento de outras análises importantes.

Os riscos de fracasso de tal procedimento poderiam ser elevados caso não houvesse alguma fonte de inspiração consistente e bem fundamentada. E os riscos são baixos, pois existe esta fonte: um curso de pós-graduação. São dezesseis anos, cinco professores e mais de quinhentos alunos de pós-graduação num dos cursos pioneiros de redes neurais artificiais, sob a perspectiva da engenharia de computação. O autor deste texto foi um destes mais de quinhentos alunos e hoje se encontra entre os cinco professores. O curso nasceu, cresceu e amadureceu no mesmo ritmo em que a área de redes neurais artificiais renasceu, recuperou o reconhecimento e enriqueceu-se de fundamentos técnico-científicos, das mais variadas origens.

Quase todas as fórmulas possíveis de serem adotadas para se ministrar a referida disciplina de pós-graduação foram experimentadas. Nenhuma delas resultou em fracasso, mas existem diferenças marcantes em termos de aproveitamento por parte dos alunos. O que vai se procurar retratar neste texto são algumas iniciativas que se destacaram pela eficácia de transmissão de informação relevante, além de terem sido devidamente validadas.

## 2 Motivação

Ao invés de se ocupar em explicar o que são redes neurais artificiais, o texto procura discutir para que servem as redes neurais artificiais. Repare que se recorre à funcionalidade, em detrimento de alguns aspectos conceituais de grande relevância. Ao invés de se ocupar em analisar o potencial das ferramentas computacionais inspiradas em redes neurais artificiais, tomam-se problemas genéricos e de grande interesse prático e procura-se prepará-los para serem tratados via redes neurais artificiais. Em virtude de se empregar uma formalização matemática poderosa, envolvendo o conceito de espaço normado e análise multivariada de dados, a extensão da metodologia a ser descrita é imediata quando se procura resolver problemas similares, até porque o nível de abstração empregado favorece a generalidade da técnica.

Esta política de abordagem deixa a missão de apresentar redes neurais artificiais muito mais leve do que se poderia imaginar. E a leveza está no fato de redes neurais artificiais servirem para muita coisa. Além disso, os problemas que admitem tratamento eficaz são muito populares e despertam o interesse de qualquer leitor.

A ampla utilização de redes neurais artificiais e sua eficácia no tratamento de problemas de grande interesse foi conquistada passo a passo. Parte do mérito se deve ao próprio amadurecimento da área de pesquisa. A outra parte sem dúvida é motivada pela própria evolução da natureza dos problemas que hoje desafiam a computação.

Ganhos de desempenho na presença de incrementos de complexidade dos problemas é uma tendência que não conduz a uma equação de fácil solução. Com isso, os requisitos de desempenho hoje em vigor eliminam a possibilidade de emprego de técnicas convencionais de solução, muito conhecidas e que produziam resultados satisfatórios junto a cenários menos exigentes.

A mudança de cenário criou a deficiência, e a deficiência criou a necessidade de novas ferramentas computacionais, mais flexíveis e menos fundamentadas conceitualmente que as técnicas convencionais. A maior responsabilidade dos usuários de redes neurais artificiais está justamente em detectar a deficiência e explorar apropriadamente a flexibilidade desta ferramenta computacional.

Como a tendência é a acentuação do quadro retratado acima, mais e mais investimentos irão ser devotados à área de redes neurais artificiais, assim como a outras ferramentas computacionais de perfil funcional semelhante. Fica configurado, portanto, um estado de coisas que justifica plenamente o investimento em redes neurais artificiais.

## 3 Para que servem as redes neurais artificiais?

A pergunta presente no título desta seção é tão genérica quanto freqüente. Por ser demasiadamente genérica, não admite uma resposta capaz de satisfazer a todos os seus potenciais formuladores, certamente interessados em respostas que os levem a tomar decisões quanto a investir ou não no estudo e aplicação de redes neurais artificiais. No entanto, por ser muito freqüente, merece ao menos uma tentativa de tratamento, desde que o leitor das linhas que seguem esteja consciente de que a exposição vai se concentrar em parte de toda a história. Talvez seja a parte mais importante e decisiva, mas não vai ser toda a história.

De partida, é conveniente considerar aspectos genéricos dos problemas que normalmente são tratados com sucesso com o emprego de redes neurais artificiais. São normalmente problemas de análise multivariada de dados, em que três situações básicas podem se manifestar, conforme será descrito nas subseções a seguir.

Mas antes, vale mencionar que a linha de argumentação a ser utilizada para responder à pergunta do título desta seção poderia ser outra, ou seja, poderia partir das propriedades de processamento das redes neurais artificiais, alvos de estudo da neurocomputação. Assim, no lugar de usar os problemas como motivação, poderiam ser consideradas as ferramentas de solução e seu potencial de aplicação. Via esta linha mais conceitual, mesmo sem resolver um único problema, seria possível estabelecer comparações convincentes entre redes neurais artificiais e ferramentas alternativas de computação. Como já anunciado, isto não será realizado aqui.

### 3.1 Situação 1: Treinamento Supervisionado

Os dados correspondem a relações de entrada-saída de algum(ns) fenômeno(s) observado(s), do(s) qual(is) se pode extrair amostras de seu comportamento. Se fenômenos dinâmicos estiverem envolvidos, aspectos temporais estarão presentes. Caso contrário, apenas aspectos da distribuição espacial dos dados estarão em jogo.

A informação contida nas amostras pode advir de relações de causa e efeito, atributos e classes, sintomas e diagnósticos, ações de controle e reações da planta sendo controlada. Normalmente, estas relações expressam associações não-lineares entre as variáveis independentes (vetor de entrada) e as variáveis dependentes (vetor de saída), sendo que o propósito geralmente é sintetizar um mapeamento não-linear de entrada-saída a partir das amostras disponíveis.

Repare que pode haver múltiplas entradas e múltiplas saídas. Uma vez dispondo deste mapeamento multidimensional, ele pode então ser utilizado para explicar as relações de entrada-saída. Por exemplo, no caso das amostras serem de atributos e classes, este mapeamento vai desempenhar o papel de um classificador de padrões.

Em síntese, existe alguma lei que rege o comportamento de entrada-saída e as amostras retratam a aplicação desta lei para alguns casos específicos. Assim, com base nos casos específicos, extrai-se um mapeamento que explica os casos específicos e, por interpolação, explica também todos os outros casos, dentro de um domínio de interesse em que as amostras foram obtidas.

A qualidade do processo de amostragem é condição necessária para o sucesso de todo o processo. O mapeamento não-linear pode ser obtido a partir de alguns modelos de redes neurais artificiais, embora outras técnicas de aproximação de funções e regressão não-linear possam ser consideradas aqui.

### **3.2 Situação 2: Treinamento não-supervisionado**

Os dados correspondem a listas de atributos de algum(ns) fenômeno(s) observado(s). Agora, as amostras já não expressam relações entre variáveis independentes e variáveis dependentes. Cada amostra descreve apenas uma lista de atributos que caracterizam um evento ou objeto.

Supondo que todos os atributos podem ser expressos numericamente, então as amostras podem ser representadas por pontos num espaço com tantos eixos de coordenadas quanto atributos, ou seja, a dimensão do espaço é dada pelo número de atributos. A necessidade de expressão numérica dos atributos não representa uma restrição prática importante, pois atributos simbólicos normalmente admitem a conversão para atributos numéricos discretos.

A informação que se quer extrair a partir das amostras pode ser de diversas naturezas. Em linhas gerais, sabe-se que pontos próximos no espaço de atributos correspondem a amostras com forte semelhança nos atributos, assim como se sabe que pontos uniformemente distribuídos no espaço de atributos correspondem a amostras que apresentam o mesmo nível de correlação de atributos entre si.

Sendo assim, pode ser interessante responder às seguintes questões: as amostras expressam algum agrupamento de pontos? Se sim, quantos? Após discriminá-los, em que tipo de atributos as amostras de cada agrupamento se assemelham mais, ou seja, o que caracteriza cada agrupamento? Finalmente, também após discriminar os agrupamentos, para cada nova amostra a ser classificada de acordo com seus atributos, em qual agrupamento ela deve ser classificada?

A discriminação dos agrupamentos só pode ser conseguida a partir da implementação de processos de auto-organização, o que é conseguido pela aplicação de redes neurais artificiais sob treinamento competitivo ou não-supervisionado, embora outras técnicas de auto-organização possam ser consideradas aqui.

### **3.3 Situação 3: Memória endereçável por conteúdo**

Os dados correspondem a amostras ruidosas ou com conteúdo parcial de memórias previamente armazenadas. O objetivo é recuperar a memória armazenada a partir de sua versão ruidosa ou incompleta presente na amostra, ou seja, realizar o que se conhece por memória endereçável por conteúdo ou memória associativa.

Para tanto, existem algumas redes neurais recorrentes capazes de sintetizar sistemas dinâmicos não-lineares que apresentam como pontos de equilíbrio exatamente as memórias disponíveis. Sendo assim, partindo de uma condição inicial equivalente a uma versão ruidosa ou com conteúdo parcial de uma das memórias, a simples sujeição da rede neural a uma dinâmica de relaxação é suficiente para conduzir à recuperação de uma das memórias armazenadas, pois vai ocorrer seguramente convergência para um ponto de equilíbrio.

A memória recuperada será aquela desejada caso a condição inicial pertença à base de atração do ponto de equilíbrio associado à memória desejada. Se, por exemplo, o nível de ruído for tal que a condição inicial acaba caindo na base de atração de outro ponto de equilíbrio, então a recuperação não será bem-sucedida.

Estudos existem para aumentar as bases de atração de cada ponto de equilíbrio, aumentando assim a eficácia da operação, e também para aumentar a capacidade de memória, ou seja, número de pontos de equilíbrio estáveis que podem ser produzidos com um determinado número de neurônios na rede neural.

## **4 Preparando os problemas para serem resolvidos usando redes neurais artificiais**

Na existência de uma das situações retratadas acima, redes neurais artificiais se transformam em candidatos em potencial para resolver o problema. Mas pode surgir a seguinte pergunta: quando é que eu me defrontei ou como detectar que eu estou me defrontando com alguma das 3 situações acima?

A resposta a esta pergunta é um tanto quanto inusitada: praticamente todos os problemas passíveis de tratamento computacional podem ser vistos como problemas de análise multivariada e, portanto, são candidatos ao tratamento via redes neurais artificiais.

No entanto, que fique bem claro que a possibilidade de emprego de redes neurais artificiais não implica na garantia de bom desempenho, muito menos na garantia de que ferramentas alternativas não produzirão desempenho superior, ou então o mesmo desempenho, mas com menos esforço, ou seja, melhor uso de recursos computacionais. Mas se as redes neurais artificiais não proporcionassem ganhos de desempenho no tratamento de diversos problemas de grande interesse prático, elas não estariam despertando tanto interesse e envolvendo tantos pesquisadores ao redor do mundo. O essencial, portanto, é não superestimar o poder desta ferramenta computacional, nem tão pouco subestimá-lo.

Embora a existência de não-linearidades significativas e a existência de dados de elevada dimensão para serem processados sejam indicativos fortes da viabilidade de se empregar redes neurais artificiais, a melhor sugestão é sempre procurar aplicar também técnicas alternativas de solução, em suas melhores configurações, e comparar os resultados obtidos. Quando as estratégias de solução sob comparação são implementadas de modo a fornecer o melhor desempenho possível, então qualquer que seja o resultado desta comparação, ele não poderá ser rebatido com argumentos elementares.

A maior dificuldade para se implementar esta política de validação do uso de redes neurais artificiais é a necessidade de se dominar outras técnicas de solução que possam ser aplicadas ao mesmo problema, ou então ter condições de demonstrar que não existem técnicas alternativas, dadas as particularidades do problema.

Quando há técnicas alternativas de solução, é possível que elas requeiram que o problema seja modelado de forma diferente, um para cada técnica, para permitir o tratamento.

Como no caso de redes neurais artificiais estaremos modelando todos os problemas como problemas de análise multivariada, algum nível de abstração pode ser necessário para permitir descrever cada problema de interesse em uma forma adequada. É o que passaremos a fazer nas sub-seções que seguem. Iremos nos restringir a alguns casos apenas, que são muito difundidos e que podem ser utilizados como pontos de partida para uma adaptação a outros contextos de aplicação.

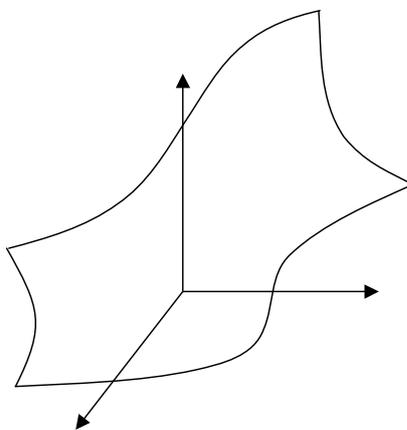


Figura 1– Mapeamento desconhecido a ser aproximado.

#### 4.1 Estudo de Caso 1: Aproximação de Funções

Este caso engloba todos os demais associados à Situação 1 descrita acima, pois toda função é, na verdade, um mapeamento único do domínio para o contra-domínio. Ele será utilizado aqui com o propósito específico de apontar as principais características de um processo de geração de um mapeamento a partir de dados amostrados.

Considere o mapeamento desconhecido apresentado na Figura 1, que é uma função cuja descrição matemática assume a forma:  $f: \mathfrak{R}^2 \rightarrow \mathfrak{R}^1$ . Esta função toma duas variáveis de entrada como argumentos e mapeia em uma variável de saída.

A Figura 2 ilustra o problema de descobrir qual é a relação funcional entre as duas variáveis de entrada e a variável de saída, restrita a uma região fechada e limitada do domínio, denominada aqui de região de operação.

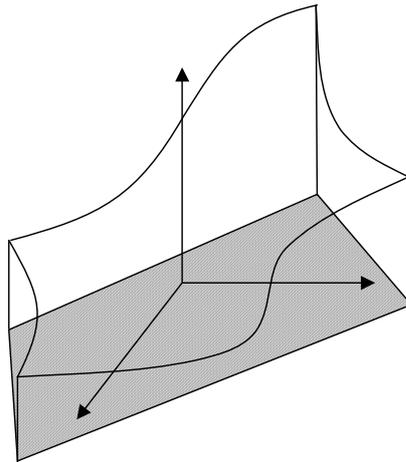


Figura 2 – Exemplo de região de operação.

Considere também que, embora a função seja desconhecida, existem amostras que expressam o comportamento desta função em alguns pontos da região de operação. A densidade de amostragem deve ser suficientemente elevada para permitir expressar as principais características do mapeamento ao longo de toda a região de operação. No entanto, muitas vezes não se pode dispor da quantidade e qualidade de amostras que se deseja, em virtude dos custos associados ao processo de amostragem, dentre outros fatores. Estamos considerando aqui a situação mais adequada, sem perda de generalidade. A Figura 3 apresenta o possível resultado do processo de amostragem, suposto aqui ser imune a ruído, sendo que o mapeamento foi excluído justamente para indicar apenas a informação disponível.

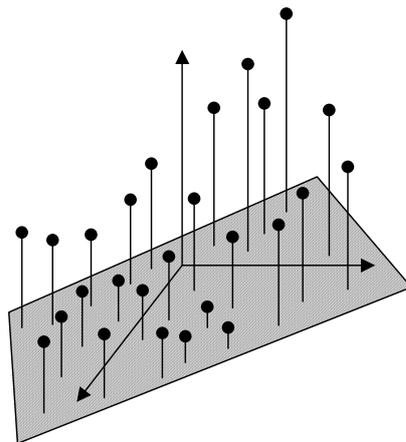


Figura 3 – Amostras expressando o comportamento da função para pontos específicos da região de operação.

É unicamente com base nestas amostras em um espaço multimimensional que vai se buscar sintetizar um mapeamento que representa a melhor aproximação. Repare que não se está querendo aqui realizar a interpolação dos dados, e sim a aproximação dos dados. E há uma grande diferença entre estes dois procedimentos, a qual será ilustrada no caso de uma função  $f: \mathcal{R}^1 \rightarrow \mathcal{R}^1$ , isso após alguns comentários importantes.

Independente das dimensões do espaço onde reside o mapeamento, está demonstrado que as redes neurais artificiais são aproximadores universais, ou seja, são capazes de reproduzir qualquer mapeamento contínuo em uma região de operação fechada e limitada, com o grau de precisão que se queira. Isto implica que a flexibilidade da ferramenta que se quer empregar para realizar o mapeamento é elevadíssima e precisa ser explorada sem exageros, na medida certa.

Consideremos então o caso de uma região de operação no  $\mathcal{R}^1$  e uma função  $f: \mathcal{R}^1 \rightarrow \mathcal{R}^1$  desconhecida, conforme ilustrado na Figura 4(a). O processo de amostragem produziu os pontos apresentados na Figura 4(b). No entanto, para qualquer conjunto finito de pontos amostrados, eles sempre vão representar uma informação local e o comportamento da função em pontos não-amostrados pode assumir infinitos perfis.

Isto implica que resolver o problema de sintetizar o mapeamento a partir das amostras é mal-condicionado, por aceitar infinitas soluções. A requisição de interpolação das amostras pode levar à solução proposta na Figura 4(c), evidentemente inadequada para representar o mapeamento original. No entanto, quando se abre mão de reproduzir exatamente os dados,

forçando a suavidade do mapeamento, então soluções mais adequadas podem ser obtidas via aproximação de funções, como ilustrado na Figura 4(d).

Repare que a solução volta a ser única quando se adiciona um outro objetivo ao problema. Originalmente, o único objetivo era minimizar o erro entre as amostras e a predição fornecida pelo mapeamento sintetizado. Agora, deve-se também maximizar a suavidade do mapeamento. Mas vale salientar que a solução é única para cada critério de suavidade, podendo variar para critérios diferentes.

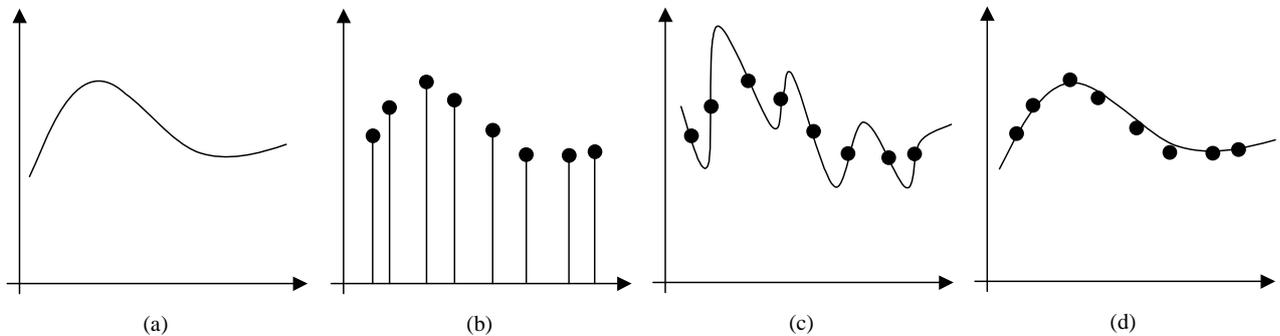


Figura 4 – (a) Função a ser aproximada; (b) Amostras disponíveis; (c) Resultado de um processo de interpolação; (d) Resultado de um processo de aproximação.

Existem muitas técnicas, geralmente derivadas da estatística, para se definir um nível adequado de suavidade para o mapeamento. Uma abordagem muito comum é utilizar amostras de teste para avaliar o mapeamento obtido. Se tomarmos a capacidade de estimar as amostras de teste como critério de qualidade do mapeamento, então a capacidade de generalização, ou seja, a possibilidade de sucesso do mapeamento em descrever o comportamento da função desconhecida em pontos não-amostrados, será elevada.

A Figura 5, a seguir, ilustra este fato para as duas soluções apresentadas na Figura 4. No caso da Figura 5(a), o mapeamento é inadequado para estimar as amostras de teste, embora apresente erro zero junto às amostras de treinamento. Já no caso da Figura 5(b), apesar do erro de treinamento não ser zero, a capacidade de estimar as amostras de teste é elevada, indicando uma maior capacidade de generalização.

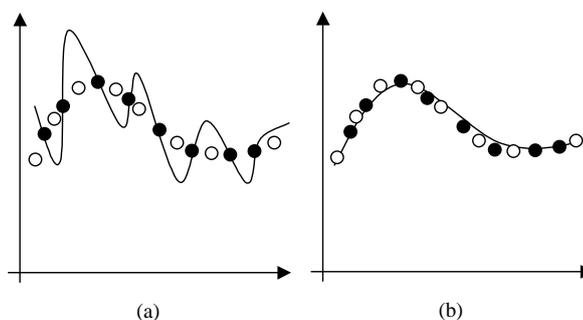


Figura 5 – Comparação de desempenho para dados de treinamento e teste, de modo a medir a capacidade de generalização dos mapeamentos produzidos.

Todos os estudos de caso a seguir, quando se referirem à Situação 1, recairão na mesma problemática abordada nesta subseção.

#### 4.2 Estudo de Caso 2: Classificação de Padrões

**Problema:** A partir de amostras de entrada-saída contendo no vetor de entrada uma lista de atributos numéricos e na saída a classe a que a amostra pertence (também expressa numericamente), obtenha um classificador de padrões que, toda vez que novos vetores de entrada (novas listas de atributos numéricos ou novos padrões) forem apresentados, eles sejam automaticamente classificados em uma das classes disponíveis.

**Solução:** Usando as amostras disponíveis, obtenha uma rede neural artificial que mapeie cada lista de atributos em sua correspondente classe e, em seguida, utilize este mapeador para classificar novas listas de atributos, simplesmente

apresentando estes atributos ao mapeador e observando a saída fornecida. Formulação totalmente equivalente ao Estudo de Caso 1.

A formulação acima é suficientemente genérica para permitir aplicar o procedimento na classificação de textos, imagens, sons e todo tipo de objetos ou pessoas que possam ser descritas por uma lista finita de atributos numéricos. Além disso, quando for possível tomar atributos simbólicos e convertê-los em atributos numéricos, amplia-se ainda mais a aplicabilidade do procedimento.

#### **4.3 Estudo de Caso 3: Identificação de Sistemas Dinâmicos**

**Problema:** A partir de amostras de entrada-saída contendo no vetor de entrada as entradas e o estado atual de um sistema dinâmico e no vetor de saída o novo estado do sistema dinâmico, obtenha um modelo para o sistema dinâmico, capaz de reproduzir a evolução no tempo de seu estado.

**Solução:** Usando as amostras disponíveis, obtenha uma rede neural artificial que mapeie cada vetor de entrada em seu respectivo vetor de saída. Logo, sabendo que a saída da rede neural vai reproduzir o estado do sistema dinâmico no tempo, então ela representa um modelo do sistema dinâmico, ou seja, o resultado de um processo de identificação de sistemas dinâmicos. Formulação totalmente equivalente ao Estudo de Caso 1.

#### **4.4 Estudo de Caso 4: Emulação de Atividades Humanas**

**Problema:** A partir de amostras de entrada-saída contendo no vetor de entrada um conjunto de informações recebidas por um ser humano e no vetor de saída as respectivas ações tomadas pelo ser humano com base nas informações recebidas, obtenha um modelo para a atuação do ser humano, capaz de reproduzir suas ações frente a variados conjuntos de informações recebidas como entrada.

**Solução:** Usando as amostras disponíveis, obtenha uma rede neural artificial que mapeie os vetores de entrada em seus respectivos vetores de saída. Logo, sabendo que a saída da rede neural vai reproduzir as ações de um ser humano, então ela representa um modelo do comportamento do ser humano, ou seja, um emulador. Formulação totalmente equivalente ao Estudo de Caso 1.

A formulação acima é suficientemente genérica para permitir aplicar o procedimento na emulação de controladores de processos industriais, sistemas de diagnóstico médico e muitas outras especialidades, incluindo composição musical e produção de obras artísticas de diversas naturezas. Não é uma tarefa elementar, no entanto, definir o conjunto de informações que está sendo considerado na tomada de decisão em cada caso e a cada instante. Há situações em que obter este conjunto de informações é mais viável que em outras.

#### **4.5 Estudo de Caso 5: Predição de Séries Temporais**

**Problema:** A partir de amostras da evolução de uma série temporal, obtenha um preditor capaz de estimar valores futuros da própria série a partir do comportamento da série temporal até o presente momento.

**Solução:** Usando as amostras disponíveis, realize um estudo estatístico que indique quais são os valores passados da série que mais estão correlacionados com o instante futuro que se quer prever e obtenha uma rede neural artificial que mapeie cada vetor de entrada com estes valores passados em uma saída que deve corresponder ao valor da série temporal naquele instante futuro. De acordo com o tamanho da série temporal, muitos exemplos de entrada-saída podem ser extraídos da própria série. Logo, sabendo que a saída da rede neural vai reproduzir o valor da série temporal em um determinado instante futuro, então ela representa um preditor para a série temporal. Formulação totalmente equivalente ao Estudo de Caso 1.

A formulação acima é suficientemente genérica para permitir aplicar o procedimento na obtenção de preditores para séries temporais estacionárias, ou seja, em que a influência do passado no futuro da série temporal não varia com o tempo. Como exemplo, este procedimento pode produzir preditores para:

- ocorrência de abalos sísmicos;
- cotações em bolsas de valores;
- vazão de rios e represas;
- nível de produtividade em culturas;
- certos comportamentos de massa.

Como nos casos anteriores, a presença de não-linearidades significativas é um indicativo da obtenção de resultados convincentes com o uso de redes neurais artificiais na tarefa de predição. Extensões para permitir o tratamento de séries temporais não-estacionárias também já foram propostas com sucesso.

#### 4.6 Estudo de Caso 6: Agrupamento de Dados

Embora a Situação 1 represente a mais freqüente fonte de problemas junto aos quais redes neurais artificiais têm se mostrado ferramentas de solução promissoras, a Situação 2 é aquela que apresenta o maior desenvolvimento nos últimos anos, em termos de contribuições para a área de redes neurais artificiais.

Deve ficar claro que não são as mesmas redes neurais artificiais que resolvem problemas vinculados à Situação 1 que vão resolver problemas vinculados às Situações 2 e 3. Não iremos entrar nos detalhes das arquiteturas de redes neurais aqui, visto que estamos desenvolvendo uma linha de argumentação e motivação centrada em problemas, e não na caracterização e análise da ferramenta de solução.

Esta seção aborda um problema diretamente associado à Situação 2. A formulação a ser utilizada também será suficientemente genérica para permitir sua aplicação direta a problemas semelhantes.

**Problema:** A partir da disponibilidade de um conjunto de amostras em um espaço multidimensional, em que as coordenadas representam atributos numéricos e cada amostra vai corresponder a um ponto em um espaço multidimensional, possivelmente de elevada dimensão, responda às seguintes perguntas:

- existem agrupamentos nesta distribuição de pontos?
- caso afirmativo, quantos são eles?
- para cada agrupamento, existem sub-agrupamentos?
- o que caracteriza cada agrupamento?
- é possível associar novas amostras aos agrupamentos existentes?

A resposta a estas perguntas está vinculada à classificação dessas amostras de acordo com o valor relativo de seus atributos. Repare que esta tarefa de classificação estará sendo realizada junto a dados possivelmente de elevada dimensão, dificultando sobremaneira a sua execução. Problemas deste tipo se manifestam nos mais diferentes contextos:

- séries temporais de igual dimensão podem ser comparadas entre si quando são interpretadas como pontos em espaço de elevada dimensão, em que cada eixo de coordenada corresponde a um instante discreto da série. Como um exemplo, se as séries correspondem à evolução da reação de pacientes a uma dada medicação em fase de teste, então a existência de diferentes agrupamentos de pontos vai indicar a presença de diferentes perfis de reação predominantes entre os pacientes.
- toda empresa prestadora de serviço pode realizar pesquisas de opinião, ou de definição de perfil de seus clientes, a partir de um conjunto de questões com múltiplas opções de resposta. É evidente que as questões precisam ser devidamente elaboradas para conduzir à extração das informações mais relevantes. Tomemos um exemplo numérico: 10 questões com 4 alternativas de resposta e 1000 clientes respondendo ao questionário. Mapeando as 4 alternativas de cada questão em valores numéricos, como  $\{-1, -0,5, +0,5, +1\}$ , então resultam 1000 pontos no  $\mathfrak{R}^{10}$ . Aqui, foram tomadas coordenadas reais, mesmo sabendo que apenas 4 valores podem ser associados a cada coordenada. Mais uma vez, a existência de diferentes agrupamentos de pontos vai indicar a presença de diferentes perfis de clientela, permitindo que a empresa prestadora de serviços defina políticas de atendimento dedicadas a cada perfil.

**Solução:** Usando as amostras disponíveis, defina múltiplos protótipos, que mais adiante serão associados às classes, e procure posicionar estes protótipos de modo a maximizar a representatividade dos dados a partir dos protótipos. A representatividade está diretamente vinculada à distância entre o protótipo e a amostra. Como normalmente existem mais amostras que protótipos, cada protótipo tende a representar mais de uma amostra. A priori, não é possível decidir que amostra será representada por que protótipo. Sendo assim, recorre-se a um processo de treinamento competitivo para definir automaticamente a posição espacial de cada protótipo de modo a atender aos requisitos do problema.

É importante observar que, diferente de muitas alternativas disponíveis na literatura, o número de agrupamentos não precisa ser definido a priori e o número de protótipos pode ser variável, sendo também definido automaticamente durante o processo competitivo. O treinamento competitivo recorre a estratégias de auto-organização e representa apenas a primeira de duas etapas essenciais para permitir obter respostas às questões elaboradas acima.

A segunda etapa é caracterizada pela análise discriminante, que parte da distribuição obtida para os protótipos e infere o número de agrupamentos, realizando a separação efetiva entre eles.

A Figura 6 ilustra a primeira etapa, ou seja, um possível resultado final do processo de auto-organização de quatro protótipos para uma determinada distribuição dos dados, onde os segmentos de reta, quando somados, produzem um comprimento mínimo.

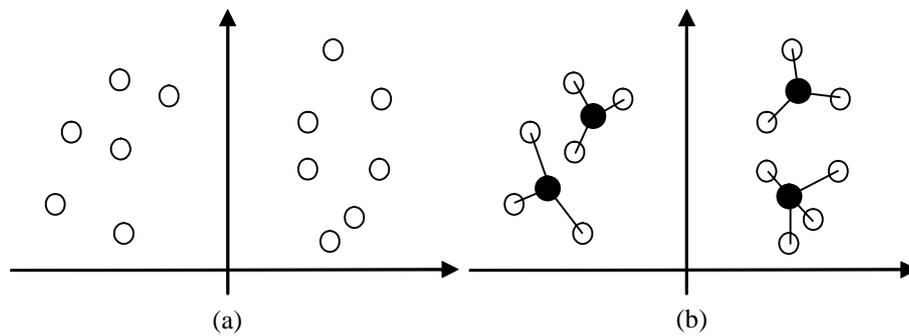


Figura 6 – Resultado de um processo de auto-organização de quatro protótipos, dada uma distribuição de amostras no  $\mathbb{R}^2$ .  
(a) Distribuição das amostras; (b) Posicionamento final dos protótipos e distância das amostras a seus protótipos representantes.

#### 4.7 Estudo de Caso 7: Caixeiro Viajante e Roteamento de Veículos

Problema do caixeiro viajante: Dadas as coordenadas de  $n$  cidades num plano, obtenha o percurso mínimo que passe uma única vez por cada cidade e retorne à cidade de origem.

Problema de roteamento de veículos: Dadas as coordenadas de  $n$  consumidores e de um depósito central e fornecido o número de veículos, minimize o somatório dos percursos dos veículos de modo que todo consumidor seja atendido por um único veículo, sendo que os percursos dos veículos iniciam e terminam no depósito central.

Estabelecendo um critério de vizinhança entre os protótipos e empregando este critério de modo a também minimizar a distância entre vizinhos, o tratamento realizado no Estudo de Caso 6 é imediatamente aplicável aos dois problemas acima. O problema do caixeiro viajante se torna um caso particular do problema de roteamento de veículos, sendo que a Figura 7 apresenta etapas intermediárias do processo de obtenção da solução. Há vizinhança unidimensional entre os neurônios, cada neurônio faz o papel de um protótipo e cada consumidor representa uma amostra a ser representada por um dos protótipos. Um detalhe importante aqui é que o número de protótipos varia automaticamente ao longo do processo de auto-organização, até que haja convergência para o mesmo número de protótipos que consumidores.

Os dois problemas acima são de grande interesse em muitas áreas de atuação, particularmente junto a distribuidores de produtos, desde alimentos a energia elétrica. O projeto de sistemas que requerem o posicionamento espacial de módulos interligados também recai nos problemas acima. A complexidade destes problemas, sob o ponto de vista da teoria de complexidade computacional, os torna intratáveis quando se recorre a algoritmos exatos, ou seja, algoritmos que garantem a obtenção da solução ótima.

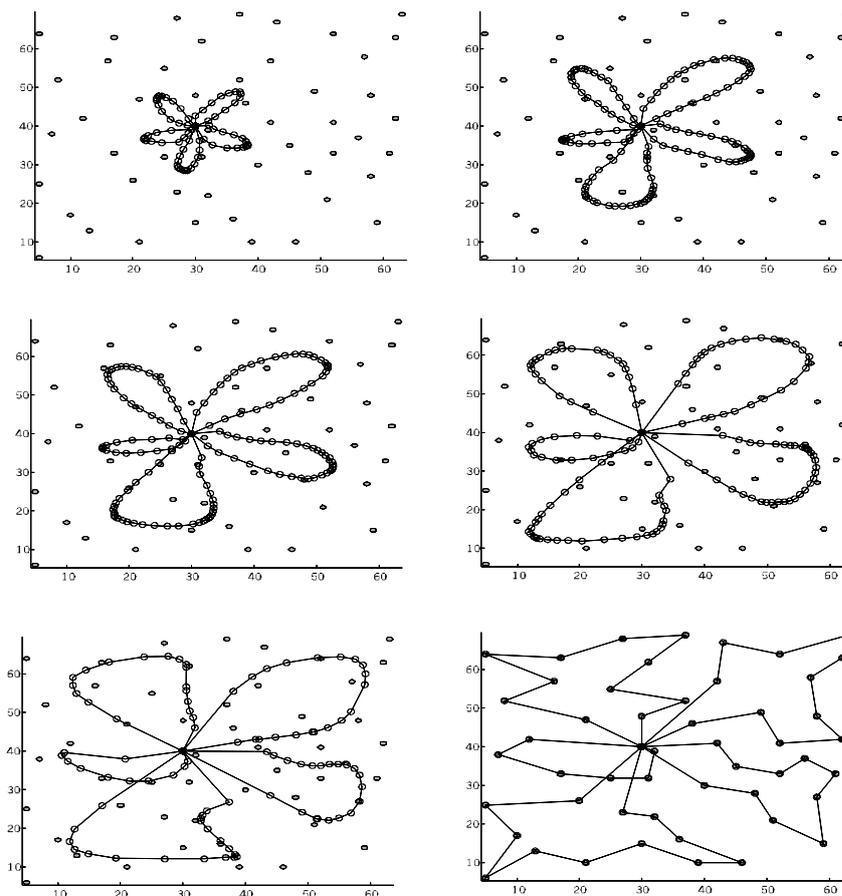


Figura 7 – Vários passos da solução de uma instância do problema de roteamento de veículos, com 5 veículos, usando redes neurais auto-organizáveis. Os pontos interligados representam neurônios ou protótipos (variam em número ao longo da execução) e as ligações entre eles indicam as relações de vizinhança.

O emprego de redes neurais auto-organizáveis não garante a obtenção da solução ótima, mas acaba fornecendo sempre uma boa solução. Para um número elevado de cidades ou consumidores, entre ficar com nenhuma solução ou uma boa solução (mesmo que não seja a ótima), a última alternativa parece ser a mais razoável. Mais uma vez, algoritmos alternativos (também incapazes de garantir a obtenção da solução ótima) existem para propor a solução e comparações de desempenho precisam ser realizadas com critério.

#### 4.8 Estudo de Caso 8: Memória Endereçável por Conteúdo

Neste último caso, estaremos tratando um problema vinculado à Situação 3, em que se utiliza dinâmica não-linear de relaxação para se implementar memórias endereçáveis por conteúdo.

**Problema:** Dados  $n$  padrões de referência, que normalmente correspondem a vetores de atributos, frente à apresentação de qualquer novo vetor de atributos na entrada, aponte qual é o padrão de referência a ele associado. Em outras palavras, utilize o conteúdo da entrada para endereçar o padrão de referência correspondente.

**Solução:** Sintetize um sistema dinâmico não-linear que apresente  $n$  pontos de equilíbrio (repare que somente na presença de não-linearidades é que se pode obter múltiplos pontos de equilíbrio), associe cada ponto de equilíbrio a um padrão de referência e, para cada padrão de entrada, utilize os seus atributos para definir a condição inicial do sistema dinâmico. Em seguida, deixe o sistema dinâmico atingir o seu estado estacionário (dinâmica de relaxação), ou seja, convergir para um dos pontos de equilíbrio existentes, e tome o padrão de referência correspondente ao ponto de equilíbrio para o qual houve convergência como solução do problema.

Sintetizar dinâmicas não-lineares com propriedades que se deseja não é uma tarefa elementar, mas foi realizada com sucesso no contexto de redes neurais artificiais. Embora haja alguns detalhes de projeto que levam à uma degradação de desempenho, como a presença de memórias espúrias, ou seja, pontos de equilíbrio adicionais, que não correspondem a padrões de referência, muitas aplicações importantes destas memórias endereçáveis por conteúdo tem sido realizadas. Implementações em hardware são comuns e normalmente envolvem um número elevado de neurônios.

A Figura 8 ilustra uma aplicação bem sucedida dos conceitos descritos acima. Uma imagem binária representando o algarismo 3 foi associada a um ponto de equilíbrio de uma rede neural recorrente, utilizada para sintetizar o sistema dinâmico não-linear. Em seguida, o estado inicial da rede neural foi inicializado com uma versão ruidosa do algarismo 3. Após a fase de relaxação, houve a convergência para o ponto de equilíbrio associado ao padrão de referência. Diz-se então que uma memória foi endereçada pelo conteúdo da entrada.

Vale salientar aqui que, embora a imagem seja bidimensional, é sempre possível realizar um empilhamento das colunas de bits (ou linhas) de modo a obter um vetor. Logo, mais uma vez temos pontos em espaços multidimensionais.

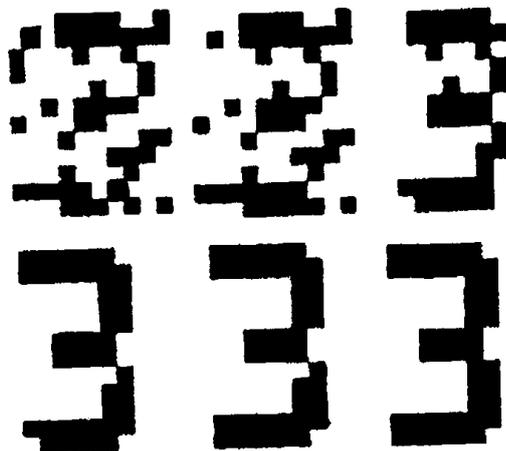


Figura 8 – Recuperação de um padrão de referência a partir de uma entrada ruidosa, usando dinâmica de relaxação.

## 5 Conclusões

Em várias oportunidades ao longo do texto foi mencionada a possível existência de ferramentas alternativas às redes neurais artificiais. A intenção não foi confundir o leitor. Foi simplesmente indicar que redes neurais artificiais representam sempre opções de aplicação, dentre outras. É responsabilidade do usuário saber detectar quando vale a pena investir em uma solução via redes neurais artificiais. Este texto não fornece subsídios adequados para executar tal tarefa.

No entanto, este mesmo texto procura delimitar adequadamente o campo de atuação das redes neurais artificiais. Mais uma vez alerta-se para o fato de que somente parte da história foi contada. Existem muitas outras aplicações e existem muitos outros conceitos diretamente vinculados às aplicações consideradas. Logo, sem uma leitura complementar de textos clássicos da área de redes neurais, não será nada fácil progredir a contento no uso e na análise destas ferramentas para a solução de problemas.

Por último, é mandatório anunciar que a pesquisa em redes neurais artificiais vem apresentando contribuições de extrema importância junto a uma outra área de pesquisa que seguramente vai ocupar a atenção de grande parte dos pesquisadores em matemática aplicada, engenharia e ciência da computação. Tratam-se dos sistemas híbridos (por exemplo, *ensembles* e misturas de especialistas), que buscam resolver problemas recorrendo-se a ferramentas de solução alternativas aplicadas em conjunto, de tal forma a apresentar sempre um desempenho tão bom ou superior ao desempenho da melhor ferramenta quando aplicada isoladamente.

A grande disponibilidade de recursos computacionais e a variedade de propostas de solução para boa parte dos problemas desafiadores dos dias de hoje fornecem o suporte necessário para alavancar esta nova linha de pesquisa. Especialistas em determinar as melhores propostas para se tratar determinados problemas, dentre um elenco de propostas candidatas, devem estar atentos para esta mudança de paradigma, protagonizada pelas redes neurais artificiais.

## 6 Agradecimento e dedicatória

O autor deste trabalho conta com o suporte financeiro do CNPq (processo no. 300910/96-7). Dedico este texto ao meu guru, Prof. Márcio Luiz de Andrade Netto, a quem a área de redes neurais artificiais deve tanto.