Redes Neurais Artificiais e Classificação de Padrões

1. O problema do OU-exclusivo

- considere os pontos (0,0),(0,1),(1,0) e (1,1) no plano $\Re^2$, conforme apresentado na figura 1. O objetivo é determinar uma rede com duas entradas $\mathbf{x_i} \in \{0,1\}$ (i=1,2), e uma saída $\mathbf{y} \in \{0,1\}$ de maneira que: $\begin{cases} (\mathbf{x_1}, \mathbf{x_2}) = (0,0) \text{ ou } (1,1) \Rightarrow \mathbf{y} = 0 \\ (\mathbf{x_1}, \mathbf{x_2}) = (1,0) \text{ ou } (0,1) \Rightarrow \mathbf{y} = 1 \end{cases}$

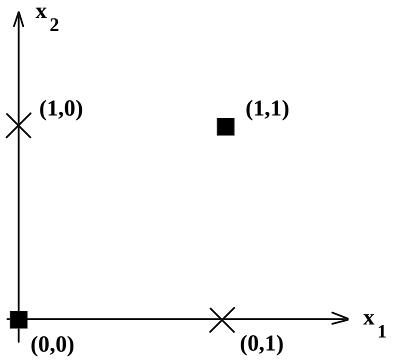


Figura 1 - O problema do OU-exclusivo

- inicialmente será analisado o comportamento de um neurônio tipo perceptron (veja figura 2) no processo de solução do problema exposto acima. A saída **y** pode ser representada na forma:

$$\mathbf{y} = \mathbf{g}(\mathbf{w_1 x_1} + \mathbf{w_2 x_2} + \mathbf{w_0}) \quad \text{onde} \quad \begin{cases} g(\mathbf{u}) = 1 \ \text{ se } \mathbf{u} \geq 0 \\ g(\mathbf{u}) = 0 \ \text{ se } \mathbf{u} < 0 \end{cases}$$

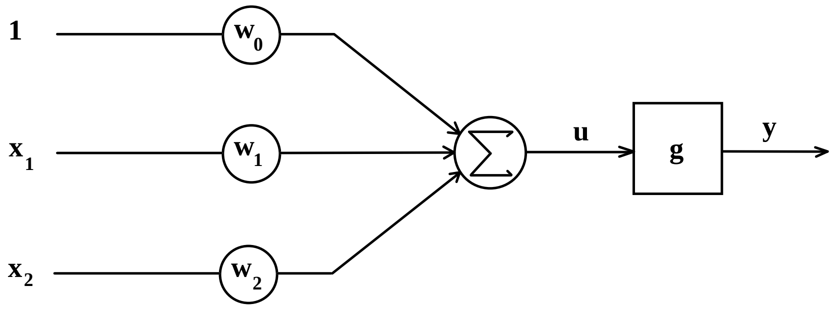


Figura 2 – Neurônio tipo perceptron, com duas entradas (mais a polarização)

- para qualquer valor dos parâmetros $\mathbf{w_0}$, $\mathbf{w_1}$ e $\mathbf{w_2}$, a função $\mathbf{g(u)}$ separa o espaço de entradas em duas regiões, sendo que a curva de separação é uma linha reta.

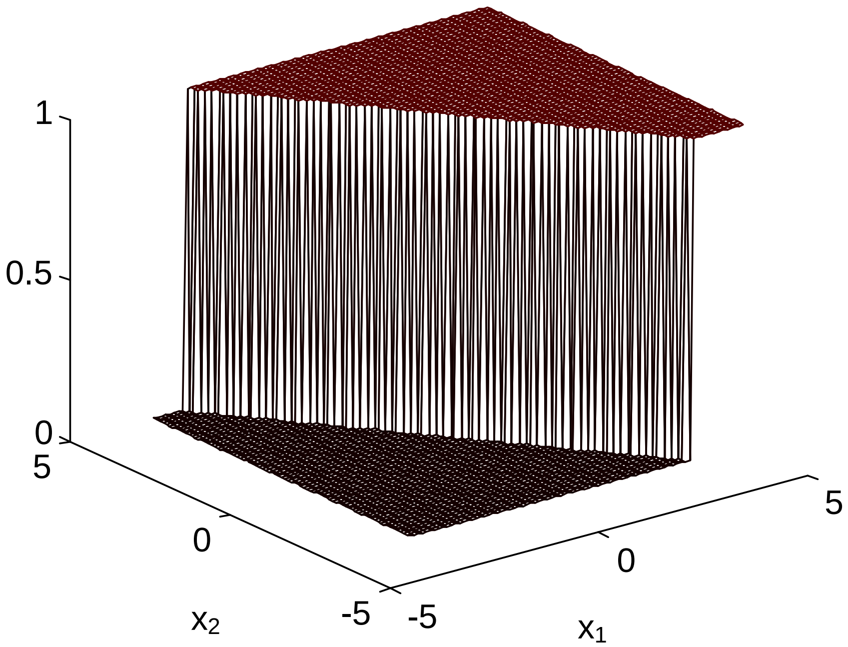


Figura 3 – Mapeamento de entrada-saída para o perceptron da figura 2, com $\mathbf{w_0} = -6$, $\mathbf{w_1} = 4$ e $\mathbf{w_2} = 3$

- no problema do OU-exclusivo (figura 1), pode-se constatar que não existe uma única linha reta divisória de forma que os pontos (0,0) e (1,1) se posicionem de um lado enquanto que (0,1) e (1,0) permaneçam do outro lado da linha.
- logo, pode-se imediatamente concluir que um neurônio tipo perceptron não apresenta grau de liberdade suficiente para resolver o problema proposto, o que foi corretamente constatado por Minsky & Papert, em 1969.
- no entanto, esses autores também acreditavam que não havia razão para supor que redes multicamadas pudessem conduzir a uma solução para o problema proposto. Esta hipótese só foi definitivamente rejeitada com o desenvolvimento do algoritmo de retro-propagação (*back-propagation*), já nos anos 80, o qual permite o ajuste automático de pesos para redes neurais multicamadas, arquitetura necessária para a realização de mapeamentos não-lineares, como será verificado mais adiante.

- considere o problema de mapeamento de uma rede neural tipo perceptron, com uma camada intermediária (veja figura 4), aplicada ao problema do OU-exclusivo.

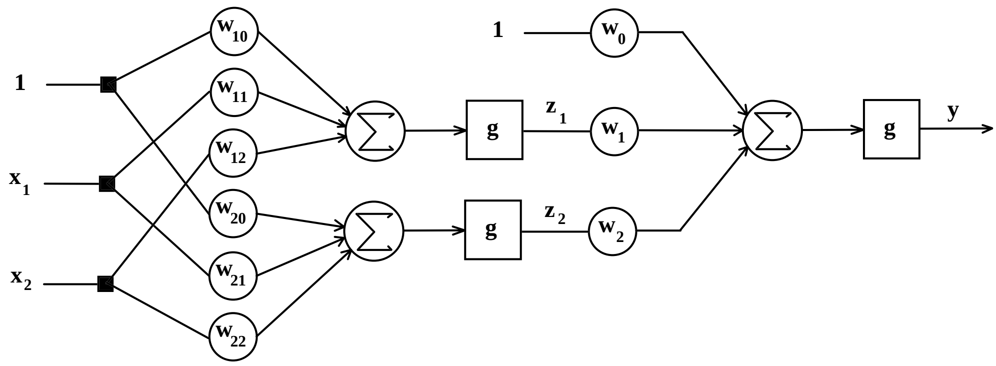


Figura 4 - Perceptron de três camadas (uma camada intermediária)

- a camada de entrada fornece um vetor de entrada ($\mathbf{x_1}$,$\mathbf{x_2}$) para a camada intermediária, enquanto que a camada intermediária produz duas saídas $\mathbf{z_1}$=sgn($\mathbf{w_{10}}$+$\mathbf{w_{11}x_1}$+$\mathbf{w_{12}x_2}$) e $\mathbf{z_2}$=sgn($\mathbf{w_{20}}$+$\mathbf{w_{21}x_1}$+$\mathbf{w_{22}x_2}$). Na camada de saída, o sinal de saída da rede neural é dado por $\mathbf{y}$=sgn($\mathbf{w_0}$+$\mathbf{w_1z_1}$+$\mathbf{w_2z_2}$).

- surge uma questão: existem parâmetros $\mathbf{w_{ij}}$ (i=1,2; j=0,1,2) e $\mathbf{w_k}$ (k = 0,1,2) tais que $\mathbf{y}$ = 0 para as entradas (0,0) e (1,1) e $\mathbf{y}$ = 1 para as entradas (1,0) e (0,1)?
- as saídas da primeira camada ($\mathbf{z_1}$ e $\mathbf{z_2}$) podem ser consideradas como <u>variáveis intermediárias</u> utilizadas na geração da saída $\mathbf{y}$.
- do que já foi visto a respeito de um neurônio tipo perceptron, sabe-se que existem pesos $\mathbf{w_{1j}}$ (j=0,1,2) tais que (veja curva de separação $\mathbf{L_1}$ na figura 5(a)):

(0,1) produza $\mathbf{z_1}$ = 1

(0,0),(1,0),(1,1) produza $\mathbf{z_1}$ = 0.

- de forma similar, existem pesos $\mathbf{w_{2j}}$ (j=0,1,2) tais que (veja curva de separação $\mathbf{L_2}$ na figura 2.5(a)):

(0,1),(0,0),(1,1) produza $\mathbf{z_2}$ = 1

(1,0) produza $\mathbf{z_2}$ = 0

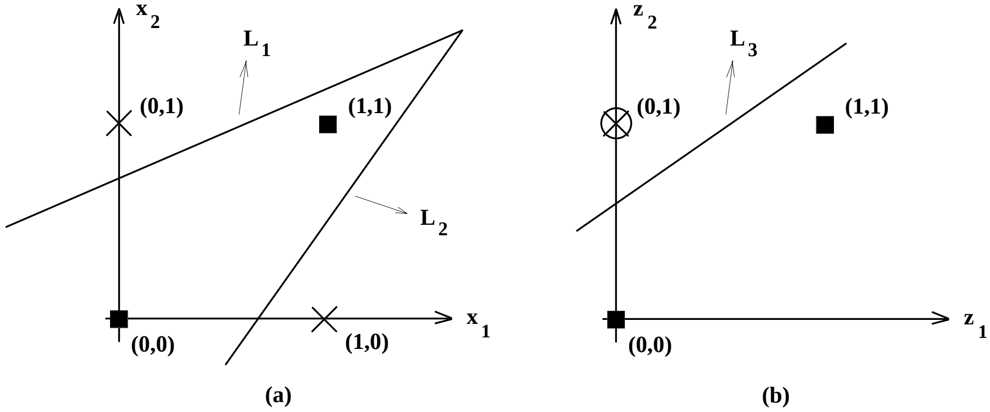


Figura 5 - Realização da função OU-exclusivo

- a discussão acima mostra que existem pesos $\mathbf{w_{ij}}$ (i=1,2; j=0,1,2) de maneira que a entrada (0,1) resulte em $\mathbf{z_1}$ = 1, $\mathbf{z_2}$ = 1, e a entrada (1,0) resulte em $\mathbf{z_1}$ = 0, $\mathbf{z_2}$ = 0, enquanto que (0,0) e (1,1) produzam $\mathbf{z_1}$ = 0, $\mathbf{z_2}$ = 1. Já que (0,0) e (1,1) podem ser separados linearmente de (0,1), como mostrado na figura 5(b) pela curva de separação $\mathbf{L_3}$, pode-se concluir que a função booleana desejada pode ser obtida utilizando-se perceptrons em cascata, ou seja, um perceptron de três camadas.

- isto é possível devido à transformação do espaço de entrada ($\mathbf{x_1}$,$\mathbf{x_2}$), onde os padrões não são linearmente separáveis, no espaço ($\mathbf{z_1}$,$\mathbf{z_2}$), onde os padrões são linearmente separáveis.
- em reconhecimento de padrões é bem conhecido que quando classes de padrões podem ser separadas utilizando-se uma função discriminante não-linear, o problema pode ser transformado em um espaço de dimensão maior onde os padrões são linearmente separáveis.
- por exemplo, se a curva $\mathbf{a_0}+\mathbf{a_1u_1}+\mathbf{a_2u_2}+\mathbf{a_3u_1u_2} = 0$ no espaço ($\mathbf{u_1}$,$\mathbf{u_2}$) separa duas classes de padrões, então um hiperplano (uma superfície linear) no espaço tridimensional ($\mathbf{u_1}$,$\mathbf{u_2}$,$\mathbf{u_1u_2}$) pode também separá-los. Esta foi a propriedade explorada na solução do problema do OU-exclusivo.
- obviamente, existem problemas de classificação muito mais “complexos” que o OU-exclusivo, não apenas por envolverem mais dimensões ou padrões, mas também por considerarem classes disjuntas.

2. Um problema mais geral de mapeamento não-linear

- considere, agora, um problema mais geral de classificação de padrões em um espaço de dimensão finita. Com base na figura 6, assume-se que a região triangular (conjunto **A**) corresponde a uma classe **1**, enquanto que o complemento desta região (conjunto **B**) corresponde a uma classe **2**. O objetivo é determinar os pesos de uma rede neural cuja saída é 1 (simbolizando a classe **1**) quando a entrada $(\mathbf{x_1},\mathbf{x_2}) \in$ **A** e 0 (simbolizando a classe **2**) quando $(\mathbf{x_1},\mathbf{x_2}) \in$ **B**.

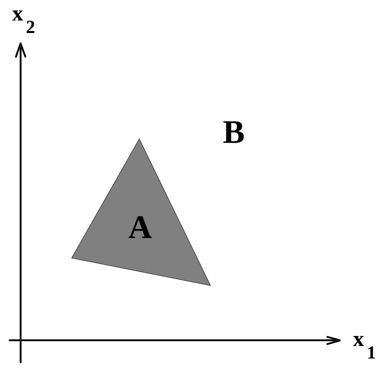


Figura 6 - Um problema de reconhecimento de padrões

- partindo das conclusões extraídas da seção anterior, é possível afirmar que cada um dos três segmentos de reta que delimitam a região **A** pode ser representado por um neurônio tipo perceptron. Tomando-se a função booleana AND das saídas destes três perceptrons, a saída **y** pode ser feita **1** quando $(\mathbf{x_1},\mathbf{x_2}) \in$ **A** e **0** quando $(\mathbf{x_1},\mathbf{x_2}) \in$ **B**. A estrutura desta rede é apresentada na figura 7.

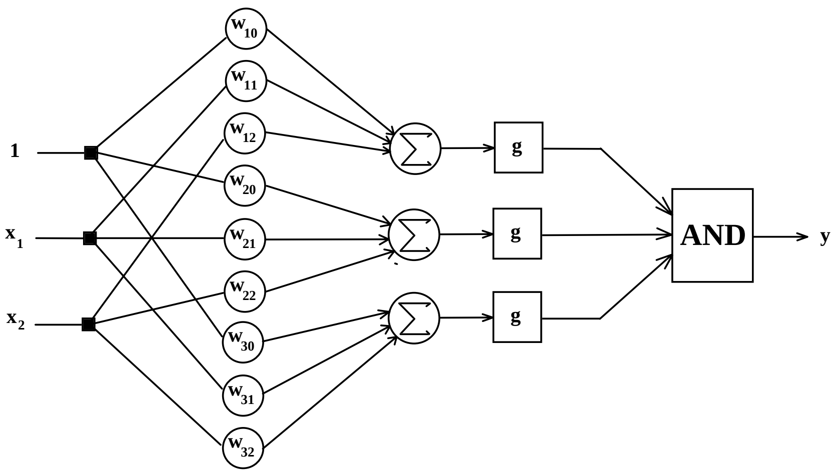


Figura 7 - Estrutura da rede para a solução do problema da figura 6

- a mesma abordagem pode ser estendida para casos em que o conjunto **A** é limitado por um polígono convexo. O número de neurônios necessários na primeira camada da rede é, neste caso, igual ao número de lados do polígono.
- se $\mathbf{A_1}$, $\mathbf{A_2}$ e $\mathbf{A_3}$ são conjuntos disjuntos (desconexos), cada qual limitado por um polígono convexo e o conjunto $\mathbf{A} = \mathbf{A_1} \cup \mathbf{A_2} \cup \mathbf{A_3}$ corresponde à classe **1** e o complemento de **A** (conjunto **B**) corresponde à classe **2**, a mesma abordagem descrita acima pode ser utilizada para reconhecer as duas classes. Neste caso, três redes idênticas àquela apresentada na figura 7 são utilizadas. As saídas $\mathbf{y_1}$, $\mathbf{y_2}$ e $\mathbf{y_3}$ das três redes são tomadas como entrada de uma função booleana OR, cuja saída é 1 (simbolizando classe **1**) quando $(\mathbf{x_1},\mathbf{x_2}) \in \mathbf{A}$ e 0 (simbolizando classe **2**) quando $(\mathbf{x_1},\mathbf{x_2}) \in \mathbf{B}$.
- com base nos resultados obtidos acima, e levando-se em conta que as funções booleanas AND e OR podem ser executadas através de um único neurônio do tipo perceptron, conclui-se que <u>um perceptron de quatro camadas pode executar a</u>

<u>tarefa de reconhecer elementos de conjuntos desconexos quando os conjuntos são limitados por segmentos lineares</u>.

- situações mais gerais, onde as regiões não são convexas podem também ser adequadamente mapeadas utilizando-se um perceptron de três camadas, já que qualquer região não-convexa pode ser representada como uma união de regiões convexas.
- no entanto, <u>a seleção automática dos pesos para redes com múltiplas camadas não é uma tarefa elementar</u>. A principal dificuldade provém da natureza descontínua da função de ativação **g** utilizada (função sinal).
- uma solução para este problema será proposta quando for apresentado o algoritmo de treinamento para redes neurais multicamadas, denominado algoritmo de retro-propagação, o qual requer que a função de ativação dos neurônios seja diferenciável, ao menos até 1ª ordem.

3. Mapeamentos não-lineares genéricos

- na seção anterior, verificou-se que redes neurais de três camadas, compostas por unidades processadoras com função de ativação do tipo sinal, são capazes de discriminar classes representadas por uma seqüência de segmentos lineares.
- no entanto, muitos são os exemplos de classes que não podem ser adequadamente descritas por seqüências de segmentos lineares e, portanto, não podem ser discriminadas por este tipo de rede neural.
- por exemplo, considere o problema de classificação de padrões pertencentes a duas classes distintas, separáveis por uma circunferência, conforme apresentado na figura 8.
- se for utilizada uma rede neural com função de ativação do tipo sinal, a melhor solução que pode ser obtida é uma aproximação da circunferência por trechos de segmento de reta.

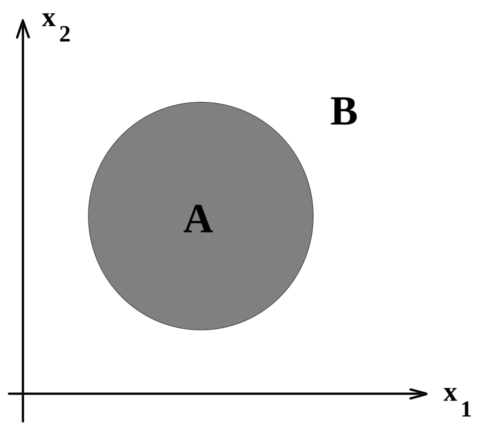


Figura 8 - Padrões separáveis por uma circunferência

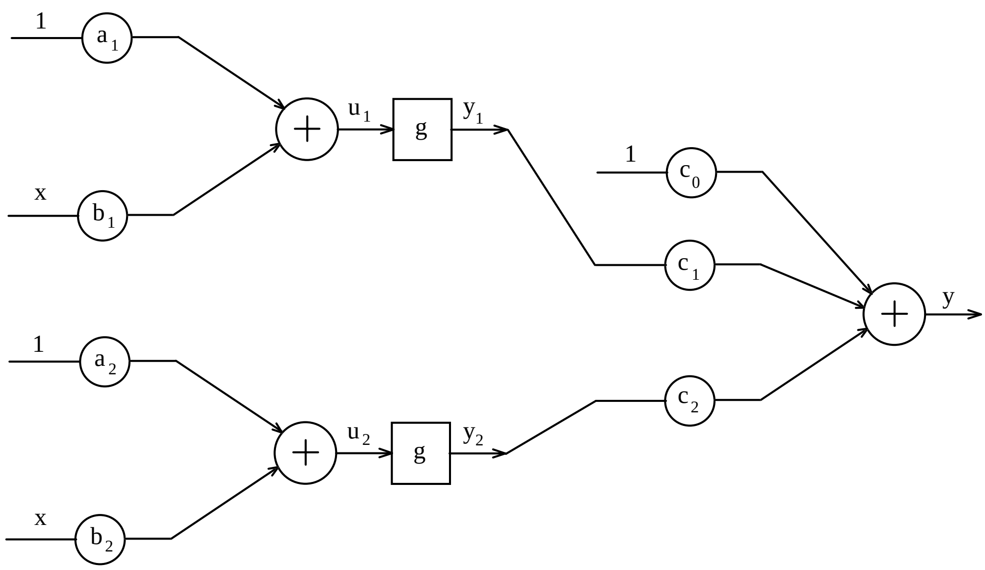
- a estrutura da rede neural é a mesma apresentada na figura 7, sendo que quanto maior o número de neurônios na primeira camada (camada de entrada), melhor será a aproximação. Esta relação é indesejável, pois cria uma dependência assintótica (com taxas de convergência muito baixas) entre a dimensão da rede e a capacidade de classificação.

- a conclusão que se pode extrair é que, no caso das redes neurais em camadas já apresentadas, funções de ativação lineares por partes, como é o caso da função sinal, vão conduzir a mapeamentos lineares por partes.
- portanto, existe aparentemente uma correspondência muito forte entre o tipo de não-linearidade da função de ativação e a capacidade de mapeamento de redes neurais artificiais para cada tipo de problema.
- se esta correspondência se aplicar a todos os tipos de função de ativação, fica bastante comprometida a utilização de redes neurais em problemas genéricos de mapeamento, pois passa a ser necessário conhecer exatamente a característica do mapeamento para que se possa definir a função de ativação a ser utilizada, ou seja, <u>é preciso aplicar a solução do problema na composição da rede neural que, em princípio, iria solucionar o problema</u>.
- felizmente, requisitos mínimos podem ser impostos às funções de ativação, de tal forma a garantir a existência de solução, independente do problema de aplicação.

- devido à estruturação da rede em camadas, o processamento dos sinais pela rede faz com que os sinais de saída sejam uma composição particular de funções de ativação, as quais devem atender a um conjunto mínimo de propriedades para conferir à rede neural o que se denomina de <u>capacidade de aproximação universal de mapeamentos não-lineares contínuos, definidos em regiões compactas do espaço de aproximação</u>.
- tomando por base métodos de aproximação universal para funções não-lineares, como séries de Taylor (composição de funções polinomiais) e séries de Fourier (composição de funções trigonométricas), não deve surpreender o fato de que redes neurais, cujas funções de ativação são dotadas de algum tipo particular de não-linearidade contínua, sejam capazes de realizar qualquer tipo de mapeamento multidimensional contínuo em regiões compactas do espaço de aproximação.
- embora não estejamos mais restritos ao caso de problemas de classificação de padrões, estes continuam a representar uma das principais aplicações.

4. Redes neurais multicamadas e suas extensões

- nesta seção, vamos apresentar um exemplo ilustrativo de como é o processo de aproximação de mapeamentos não-lineares contínuos utilizando redes neurais multicamadas (apresentaremos apenas o caso de uma camada intermediária), dotadas de função de ativação sigmoidal (portanto, uma função diferenciável).
- já vimos que diferentes expressões para a função de ativação sigmoidal g(.) podem ser escolhidas.
- já vimos também que é recomendável que o(s) neurônio(s) da camada de saída tenham funções de ativação do tipo identidade. Neste caso, considerando *p* neurônios na única camada intermediária considerada, a função de transferência da rede neural pode ser dada na forma: $y = c_0 + \sum_{n=1}^{p} c_n g(b_n x + a_n)$, onde *x* é o vetor de entrada e *y* é a saída da rede (veja exemplo na figura 9, para *p* = 2).

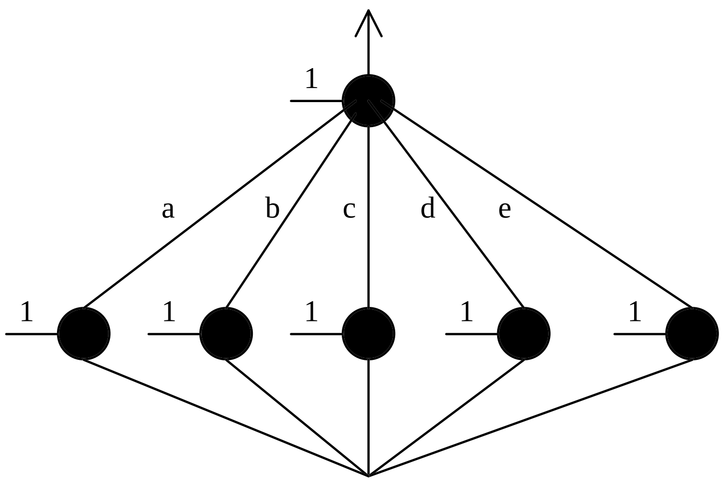


$$y = c_0 + c_1 g(b_1 x + a_1) + c_2 g(b_2 x + a_2) \Rightarrow \begin{cases} \text{a : deslocamento no eixo x} \\ \text{b : inclinação da sigmóide} \\ \text{c : amplitude da sigmóide} \end{cases}$$

Figura 9 - Rede neural multicamada e sua função de transferência

- isto significa que a rede neural com uma camada intermediária realiza um mapeamento que é dado por uma série truncada, tendo a função de ativação g(.) como função básica (função-base).
- é possível então comparar este mapeamento com aquele realizado por uma série de Fourier truncada, dada na forma: $y = c_0 + \sum_{n=1}^{p} c_n \cos(nw_0 x + a_n)$.
- se tomarmos g(.) como sendo cos(.), deduz-se que a rede neural com uma camada intermediária se transforma em uma série de Fourier generalizada, pelo fato de permitir também o ajuste da freqüência da função cos(.).
- as freqüências na série de Fourier são fixas (múltiplos da freqüência fundamental w_0) pois o ajuste deste termo representa uma operação não-linear, sem solução na forma fechada (requer processos iterativos de solução).
- como g(.) pode assumir outras formas, além da função cos(.), muitas outras possíveis extensões podem ser consideradas.

Exemplo: Forma “construtiva” de aproximação de um mapeamento não-linear



$$f(\mathbf{w}) = \underbrace{c_1 g(b_1 x + a_1)}_{a} + \underbrace{c_2 g(b_2 x + a_2)}_{b} + \underbrace{c_3 g(b_3 x + a_3)}_{c} + \underbrace{c_4 g(b_4 x + a_4)}_{d} + \underbrace{c_5 g(b_5 x + a_5)}_{e} + \underbrace{c_0}_{bias}$$

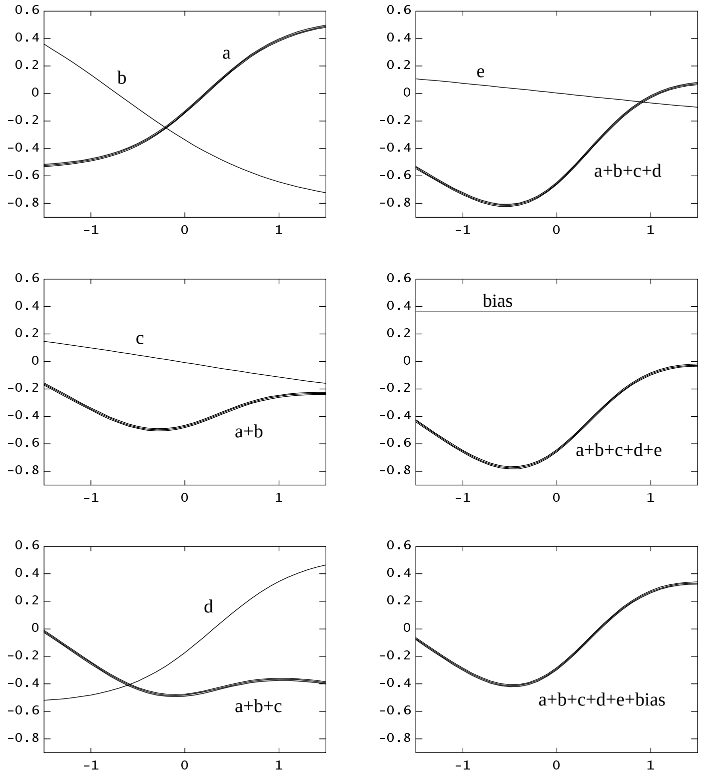


Figura 10 - Composição aditiva de ativações na reprodução de um mapeamento não-linear