



REDES NEURAIS ARTIFICIAIS

Prof. Marcelo Dib

Redes Neurais Artificiais

Motivação

A Inteligência Computacional busca, através de técnicas inspiradas na Natureza, o desenvolvimento de sistemas inteligentes que imitem aspectos do comportamento humano, tais como: aprendizado, percepção, raciocínio, evolução e adaptação.

Redes Neurais Artificiais

Motivação

- Criar máquinas capazes de operar independentemente do homem ;
- aprender sozinhas ;
- interagir com ambientes desconhecidos ;
- possa ser chamada de autônoma, inteligente ou cognitiva;
- capacidade de lidar com eventos inesperados ;

Redes Neurais Artificiais

Motivação

- Neurofisiologistas e psicólogos estão particularmente interessados em compreender o funcionamento do sistema neural humano.
- As características de resposta a estímulos apresentada por neurônios individuais, bem como redes de neurônios, são alvo de estudo dos neurofisiologistas, enquanto os psicólogos estudam funções do cérebro no nível cognitivo e estão interessados na utilização de técnicas baseadas em Redes Neurais Artificiais para criar modelos detalhados do comportamento humano.

Redes Neurais Artificiais

Motivação

- Inspirados na habilidade apresentada pelos seres humanos e outros animais no desempenho de funções como o processamento de informações sensoriais e a capacidade de interação com ambientes pouco definidos, os engenheiros estão preocupados em desenvolver sistemas artificiais capazes de desempenhar tarefas semelhantes.
- Habilidades como capacidade de processamento de informações incompletas ou imprecisas e generalização são propriedades desejadas em tais sistemas.

Redes Neurais Artificiais

Motivação

- Uma **Rede Neural Artificial (RNA)** é um sistema de processamento de informação inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência;
- Utiliza um sistema que possui circuitos que simulam o cérebro humano, inclusive seu comportamento, ou seja, aprendendo, errando e fazendo descobertas ;
- Os modelos neurais artificiais têm como principal fonte de inspiração as Redes Neurais Artificiais biológicas;

Redes Neurais Artificiais

Motivação

- Os modelos neurais artificiais oferecem um paradigma atrativo, pois “aprendem” a resolver problemas através de exemplos.
- Isso significa que se a rede aprende a lidar com um certo problema, e lhe é apresentado um similar, mas não exatamente o mesmo, ela tende a reconhecer esse novo problema, oferecendo a mesma solução.
- Aprendem a reconhecer padrões de informações estabelecidos no meio ambiente ;

Redes Neurais Artificiais

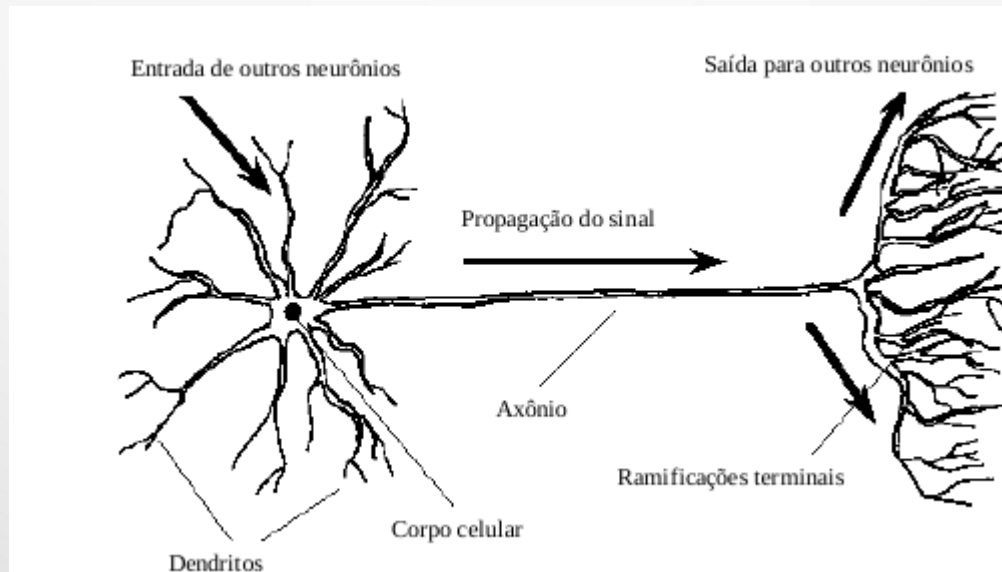
Motivação

→ A característica mais significativa de redes neuronais está em sua habilidade de aproximar qualquer função contínua ou não contínua com um grau de correção desejado. Esta habilidade das redes neuronais as tem tornado útil para modelar sistemas não lineares.

Redes Neurais Artificiais

Motivação

- O neurônio é uma célula no cérebro cuja principal função é coletar, processar e disseminar sinais elétricos;
- A Figura apresenta um modelo de um **neurônio biológico** com a sequência de propagação dos sinais pela célula;



Redes Neurais Artificiais

Motivação

- 10 bilhões de neurônios ;
- 60 trilhões de conexões (sinapses);
- O sistema nervoso é formado por um conjunto extremamente complexo de células, os neurônios ;
- O cérebro humano possui cerca de 10^{11} neurônios e mais de 10^{14} sinapses, possibilitando a formação de redes muito complexas ;
- Neurônios têm papel essencial na determinação do funcionamento e comportamento do corpo humano e do raciocínio;

Redes Neurais Artificiais

Motivação

- Neurônios são formados:
 - pelos dendritos, que são um conjunto de terminais de entrada;
 - recebem estímulos pelo corpo central (soma) coleta, combina e processa informações ;
 - pelos axônios que são longos terminais de saída transmitem os estímulos ;
- Neurônios se comunicam através de sinapses;

Redes Neurais Artificiais

Um Breve Histórico

As Redes Neurais Artificiais passaram por um interessante processo de evolução, marcado por um período de grande atividade seguido por anos de estagnação nas pesquisas e pelo ressurgimento do interesse científico como consequência do desenvolvimento de novas tecnologias e fundamentos teóricos;

Redes Neurais Artificiais

Um Breve Histórico

- MCCULLOCH & PITTS (1943) projetaram a estrutura que é conhecida como a primeira rede neural;
- Estes pesquisadores propuseram um modelo de neurônio como uma unidade de processamento binária e provaram que estas unidades são capazes de executar muitas das operações lógicas;
- Este modelo, apesar de muito simples, trouxe uma grande contribuição para as discussões sobre a construção dos primeiros computadores digitais, permitindo a criação dos primeiros modelos matemáticos de dispositivos artificiais que buscavam analogias biológicas;

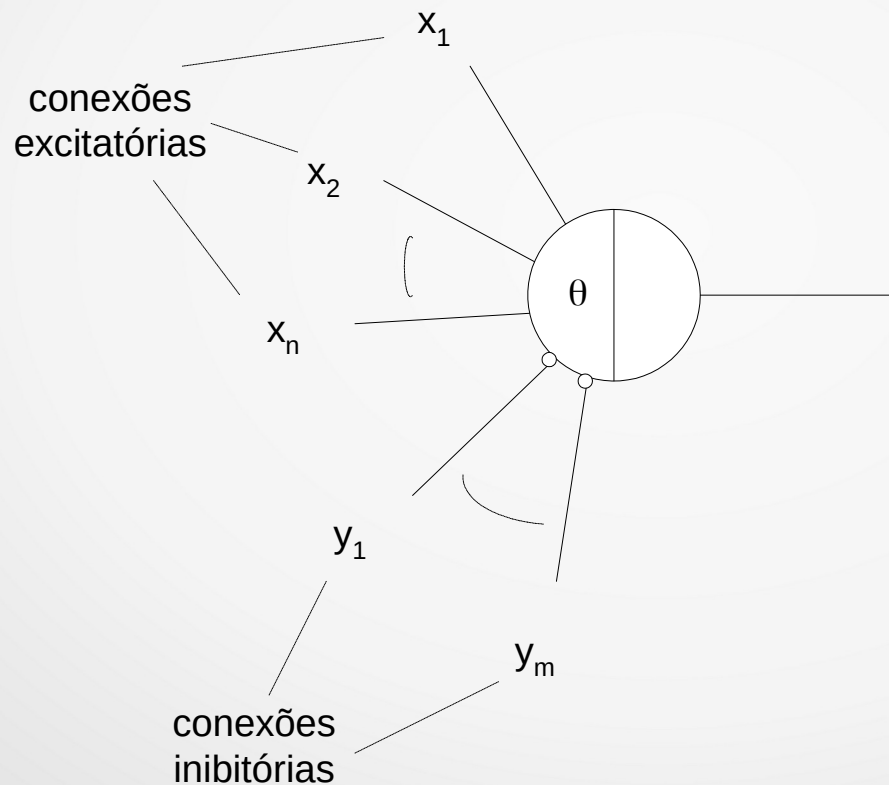
Redes Neurais Artificiais

Um Breve Histórico

- Consiste basicamente de um neurônio que executa uma função lógica.
- Os nós produzem somente resultados binários e as conexões transmitem exclusivamente zeros e uns.
- As redes são compostas de conexões sem peso, de tipos excitatórios e inibitórios.
- Cada unidade é caracterizada por um certo limiar (*threshold*) q .

Redes Neurais Artificiais

Representação funcional de um neurônio



Redes Neurais Artificiais

Um Breve Histórico

- Em 1949, O psicólogo Donald Hebb, demonstrou que a capacidade da aprendizagem em redes neurais biológicas vem da alteração da eficiência sináptica, isto é, a conexão somente é reforçada se tanto as células pré-sinápticas quanto as pós-sinápticas estiverem excitadas;
- Hebb foi o primeiro a propor uma lei de aprendizagem específica para as sinapses dos neurônios.

Redes Neurais Artificiais

Um Breve Histórico

- Em 1957, ROSENBLATT mostrou em seu livro (Principles of Neurodynamics) o modelo dos "Perceptrons".
- Nele, os neurônios (Perceptrons) eram organizados em camada de entrada e saída, onde os pesos das conexões eram adaptados a fim de se atingir a eficiência sináptica usada no reconhecimento de caracteres.
- Rosenblatt também propôs um algoritmo para o ajuste dos pesos do perceptron e provou sua convergência quando os padrões são linearmente separáveis;
- Por volta do mesmo período, B. WIDROW (1962) e seus colaboradores desenvolveram o adaline (Adaptive Linear Element) ;

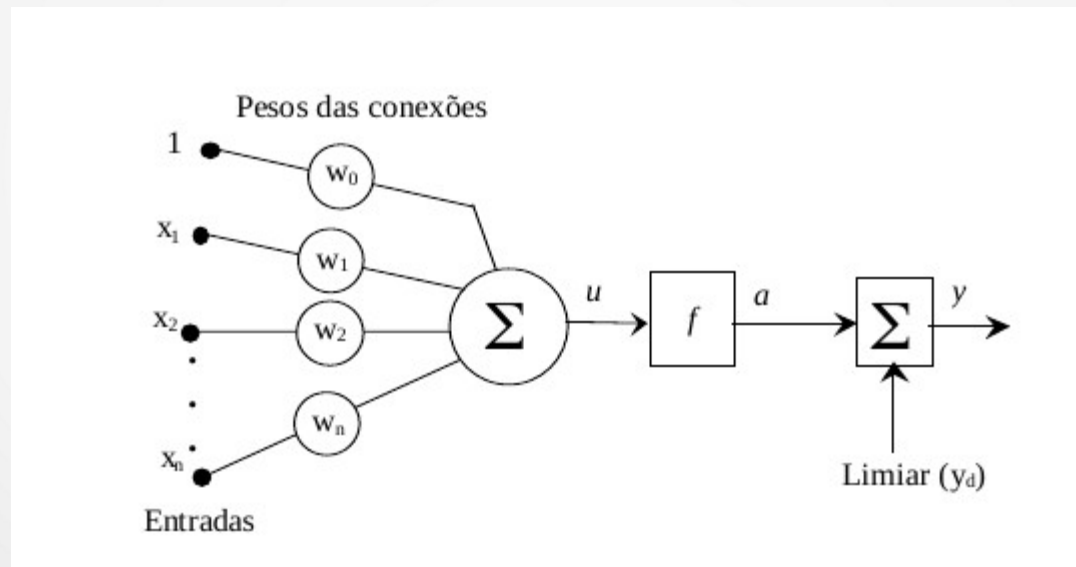
Redes Neurais Artificiais

O perceptron

- Um Perceptron é um modelo matemático que recebe várias entradas, e produz uma única saída binária;
- Ele introduziu pesos, w_1 , w_2 , ..., números reais expressando a importância das respectivas entradas para a saída.
- A saída do neurônio, 0 ou 1, é determinada pela soma ponderada, $\sum w_j x_j$, menor ou maior do que algum valor limiar (threshold) ;
- Assim como os pesos, o threshold é um número real que é um parâmetro do neurônio ;

Redes Neurais Artificiais

O perceptron



Redes Neurais Artificiais

Um Breve Histórico

- Foi constatado por Minsky & Papert que um neurônio do tipo Perceptron só é capaz de resolver problemas com dados de classes linearmente separáveis.
- Após a publicação destes resultados, a maior parte dos pesquisadores da área passou a buscar alternativas dentro do campo da engenharia e, principalmente, da lógica matemática;

Redes Neurais Artificiais

Um Breve Histórico

- Apesar deste êxodo generalizado, um número de pesquisadores continuou a trabalhar com Redes Neurais Artificiais nos anos 70. Os nomes de T. Kohonen (Finlândia), S. Grossberg, B. Widrow e J. Anderson (Estados Unidos), E. Caianiello (Itália), K. Fukushima (Japão) e Ígor Aleksander (Inglaterra) estão associados a este período;
- Nos anos 80, Neurofisiologistas foram Adquirindo um maior conhecimento sobre o processamento de informações nos organismos vivos;
- Avanços tecnológicos tornaram disponível um maior potencial computacional a baixo custo, viabilizando ou facilitando simulações e testes com modelos neurais ;

Redes Neurais Artificiais

Um Breve Histórico

- O ganhador do Prêmio Nobel J. HOPFIELD (1982), do Instituto de Tecnologia da Califórnia, juntamente com D. TANK, um pesquisador da AT&T, desenvolveram grande quantidade de modelos de Redes Neurais Artificiais baseadas em pesos fixos e ativações adaptativas.
- A rede de Hopfield pode ser considerada como um sistema dinâmico com um número finito de estados de equilíbrio, de forma que o sistema invariavelmente irá evoluir para um destes estados ou para uma seqüência periódica de estados a partir de uma condição inicial;

Redes Neurais Artificiais

Motivação

- No entanto, o fato que efetivamente colocou a área de Redes Neurais Artificiais como uma das prioritárias na obtenção de recursos foi o desenvolvimento de um método para ajuste de parâmetros de redes não-recorrentes de múltiplas camadas ;
- Este método, baseado em um algoritmo denominado retro-propagação (backpropagation), tornou-se largamente conhecido após a publicação, em 1986, do livro Parallel Distributed Processing, editado por J. L.M CCLELLAND & D. E. RUMELHART ;

Redes Neurais Artificiais

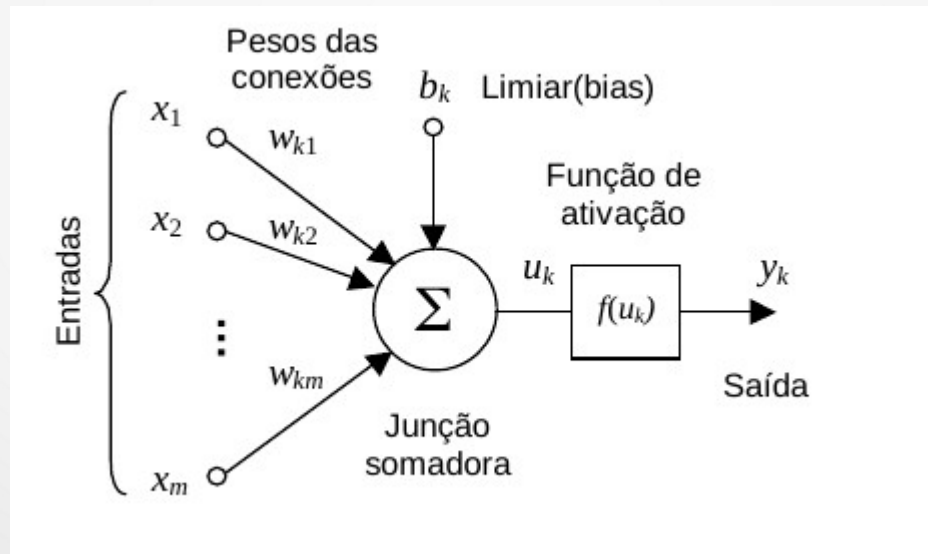
Definição

- Modelo matemático: Simplificações da realidade com o propósito de representar aspectos relevantes de um sistema em estudo, sendo que detalhes de menor significância são descartados para viabilizar a modelagem;
- As Redes Neurais Artificiais artificiais (RNA) têm sido desenvolvidas como generalizações de modelos matemáticos da cognição humana ou biologia neural;

Redes Neurais Artificiais

Definição

- o processamento da informação ocorre em vários elementos chamados neurônios artificiais;



Redes Neurais Artificiais

Definição

- A operação de um neurônio artificial se resume em:
- Sinais são apresentados à entrada (x_1 à x_m);
- Cada sinal é multiplicado por um peso que indica sua influência na saída da unidade (w_k);
- É feita a soma ponderada dos sinais que produz um nível de atividade (u_k);
- A função de ativação $f(u_k)$ tem a função de limitar a saída e introduzir não-linearidade ao modelo.
- O bias b_k tem o papel de aumentar ou diminuir a influência do valor das entradas.
- É possível considerar o bias como uma entrada de valor constante 1, multiplicado por um peso igual a b_k .

Redes Neurais Artificiais

Definição

→ A saída do neurônio k pode ser descrita por:

$$y_k = f(u_k) = f\left(\sum_{j=1}^m w_{kj} x_j + b_k\right)$$

→ É possível simplificar a notação acima de forma a incluir o bias simplesmente definindo um sinal de entrada de valor $x_0 = 1$ com peso associado $w_{k0} = b_k$;

$$y_k = f(u_k) = f\left(\sum_{j=0}^m w_{kj} x_j\right) = f(\mathbf{w}^T \mathbf{x})$$

Redes Neurais Artificiais

Definição

- os sinais são propagados de um elemento a outro através de conexões;
- cada conexão possui um peso associado, que, em uma rede neural típica, pondera o sinal transmitido;
- cada neurônio (ou unidade) aplica uma função de ativação (geralmente não-linear) à sua entrada de rede (soma ponderada dos sinais de entrada) para determinar sua saída;

Redes Neurais Artificiais

Definição

Uma RNA pode ser caracterizada por três aspectos principais:

- o padrão de conexões entre as unidades (arquitetura);
- o método de determinação dos pesos das conexões (algoritmo de treinamento ou aprendizado) ;
- sua função de ativação;

Redes Neurais Artificiais

Definição

O treinamento de RNA's pode ser dividido em:

- **Supervisionado:** necessita de um “professor” durante a fase de aprendizagem, que antecede a utilização (execução) da rede;
- **Não-supervisionado:** direcionado por correlações existentes nos dados de entrada e, portanto, não necessita de um “professor”;

Redes Neurais Artificiais

Definição

→ **Por Reforço:**

→ No aprendizado supervisionado, a medida de desempenho é baseada no conjunto de respostas desejadas usando um critério de erro conhecido, enquanto que no aprendizado por reforço a única informação fornecida à rede é se uma determinada saída está correta ou não.

→ A idéia básica tem origem em estudos experimentais sobre aprendizado dos animais. Quanto maior a satisfação obtida com uma certa experiência em um animal, maiores as chances dele aprender.

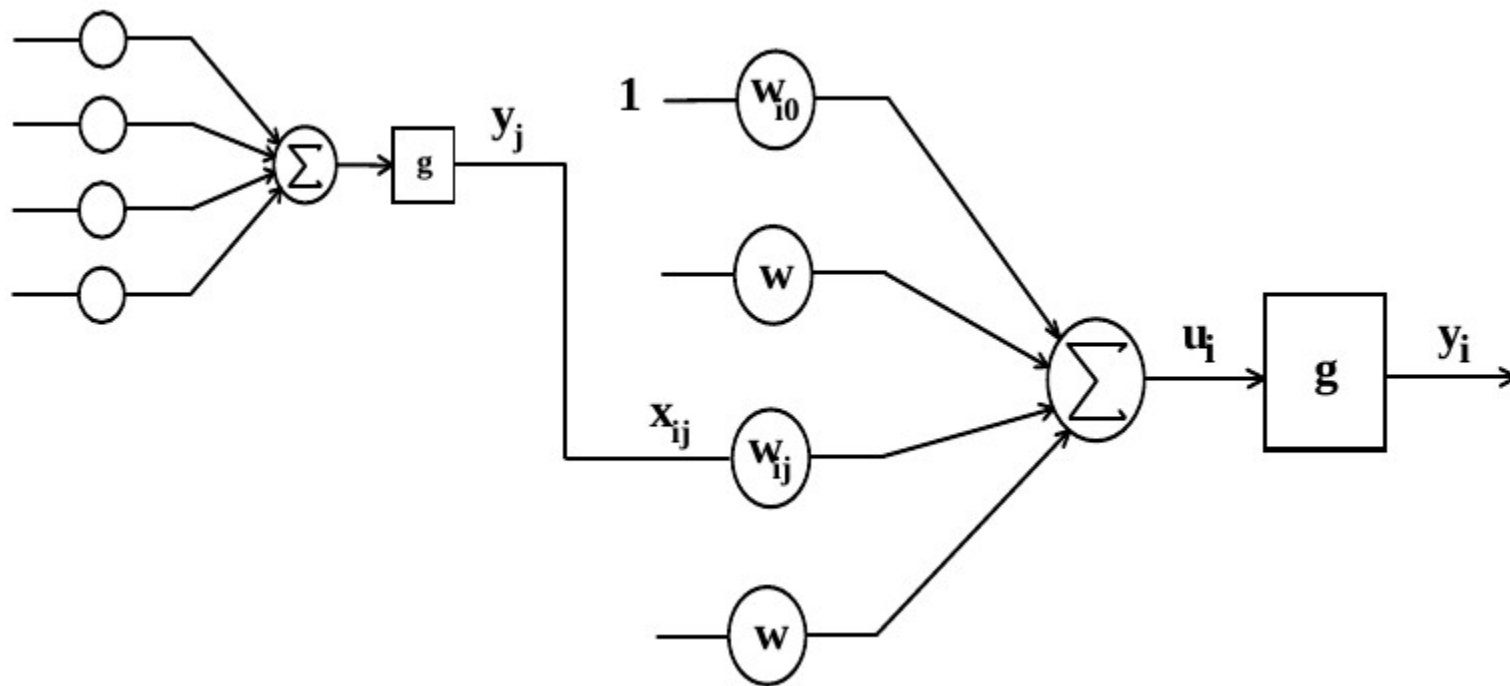
Redes Neurais Artificiais

Redes do Tipo Perceptron de Múltiplas Camadas (MLP)

- Tipicamente, esta arquitetura consiste de um conjunto de unidades sensoriais que formam uma camada de entrada, uma ou mais camadas intermediárias (ou escondidas) de unidades computacionais e uma camada de saída;
- Os sinais de entrada são propagados camada a camada pela rede em uma direção positiva, ou seja, da entrada para a saída;
- Esta arquitetura representa uma generalização do perceptron apresentado anteriormente;

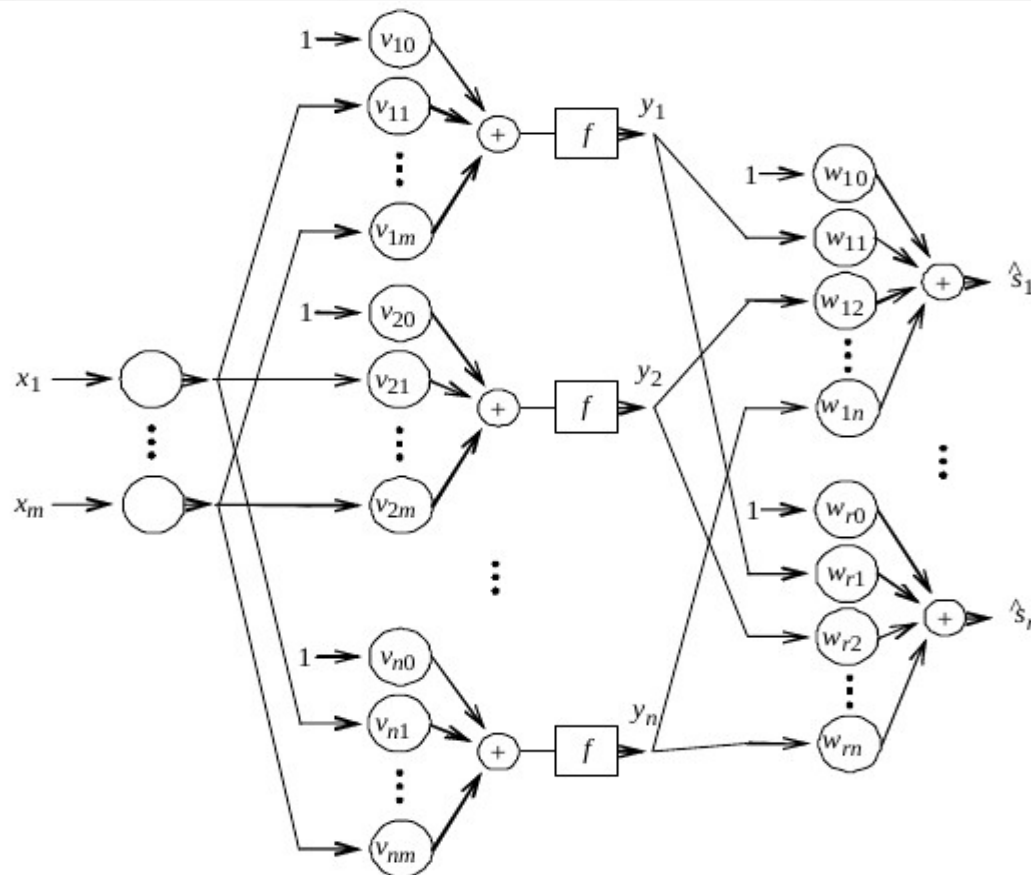
Redes Neurais Artificiais

Redes do Tipo Perceptron de Múltiplas Camadas (MLP) : duas camadas



Redes Neurais Artificiais

Redes MLP : Uma Camada intermediária



Redes Neurais Artificiais

Redes MLP

- As redes do tipo MLP tem sido utilizadas com sucesso para a solução de vários problemas envolvendo altos graus de não-linearidade;
- Seu treinamento é do tipo supervisionado e utiliza um algoritmo muito popular chamado retro-propagação do erro (error backpropagation);
- Este algoritmo é baseado numa regra de aprendizagem que “corrige” o erro durante o treinamento (HAYKIN , 1994);

Redes Neurais Artificiais

Redes MLP

Uma rede do tipo MLP possui três características distintas:

- Função de ativação;
- Número de camadas e unidades intermediárias ;
- Forma das conexões;

Redes Neurais Artificiais

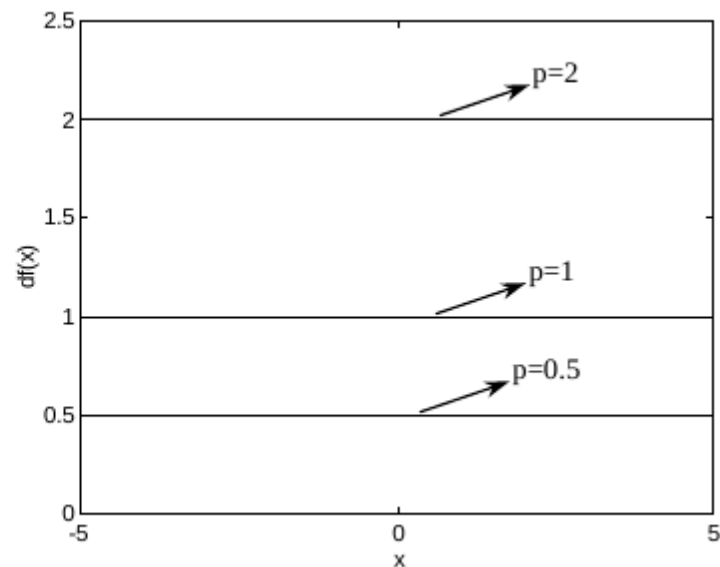
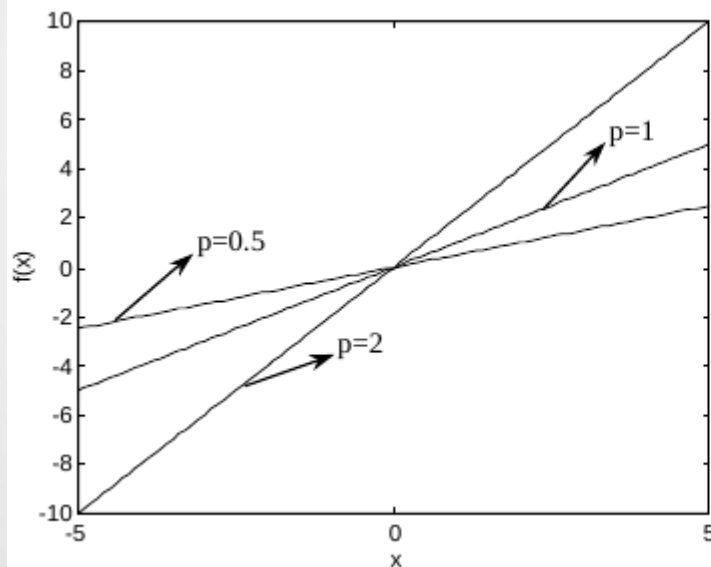
Função de Ativação

- O modelo de cada unidade da rede pode incluir uma não-linearidade na sua saída.
- É importante enfatizar que a não-linearidade deve ser suave, ou seja, diferenciável, diferentemente da função sinal utilizada pelo perceptron original.
- A função de ativação representa o efeito que a entrada interna e o estado atual de ativação exercem na definição do próximo estado de ativação da unidade. Geralmente, esta função é monotonicamente não-decrescente;

Redes Neurais Artificiais

Função de Ativação

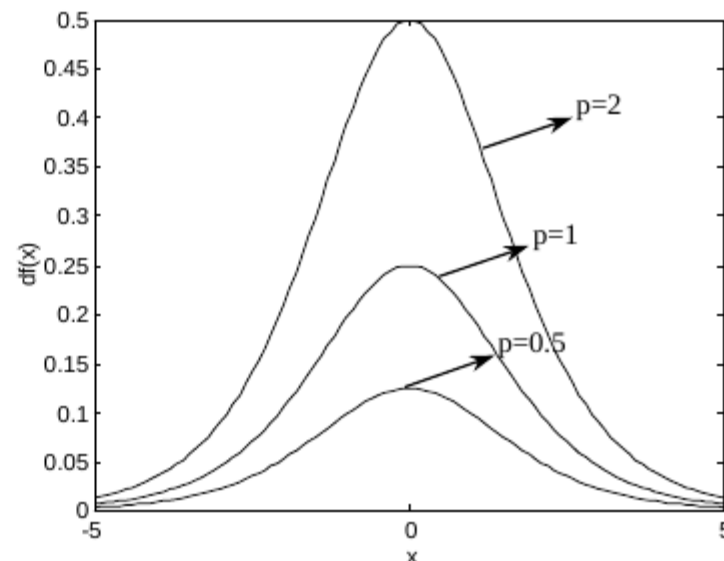
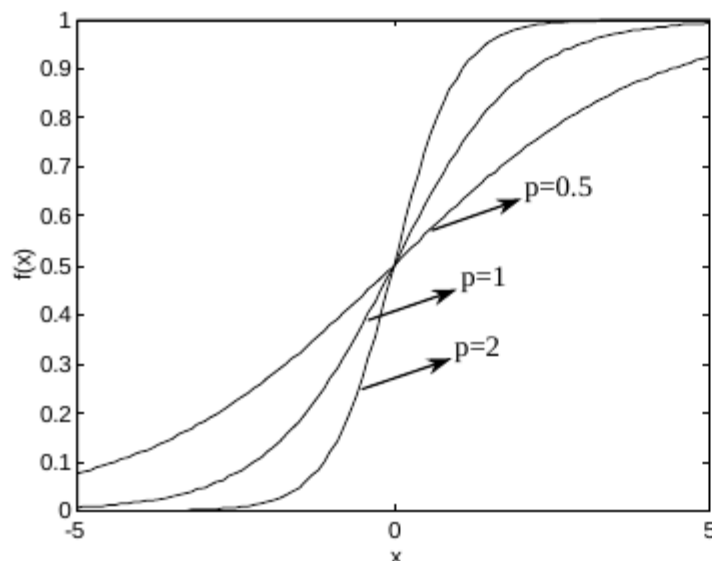
$$f(x) = p \cdot x \quad f'(x) = p$$



Redes Neurais Artificiais

Função de Ativação : Função Logística

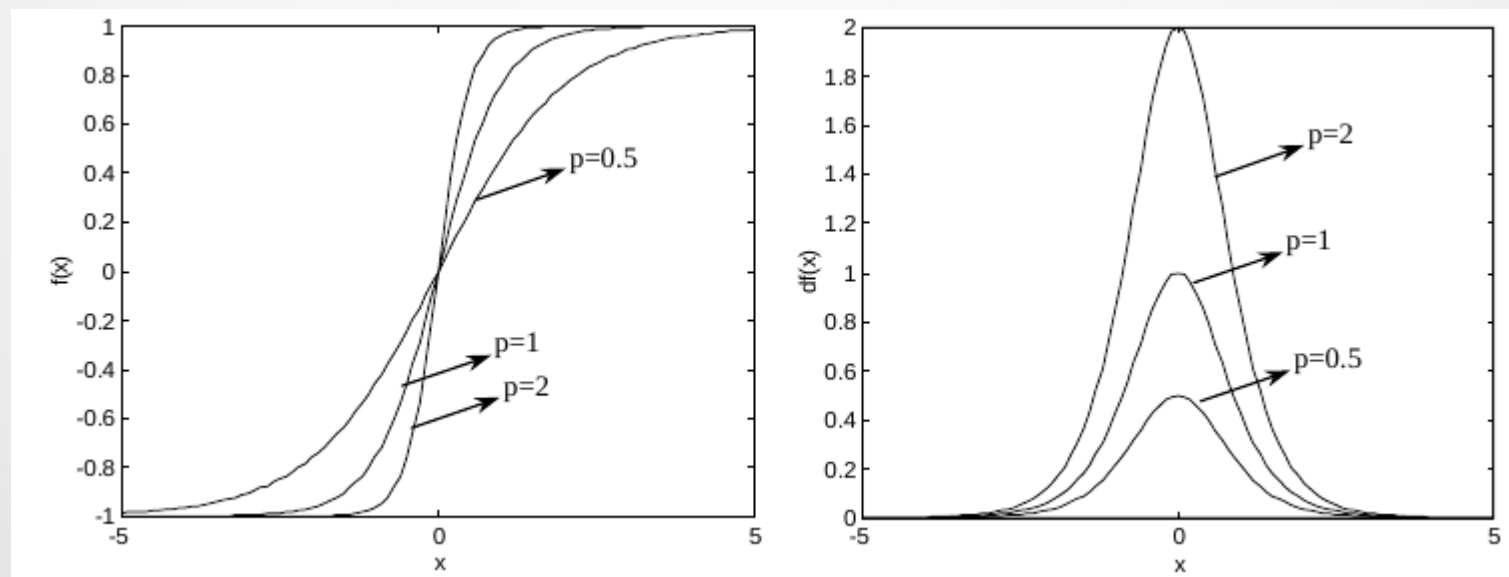
$$f(x) = \frac{e^{px}}{1 + e^{px}} = \frac{1}{1 + e^{-px}}, \quad f'(x) = p f(x) \cdot (1 - f(x))$$



Redes Neurais Artificiais

Função de Ativação : Função Tangente Hiperbólica

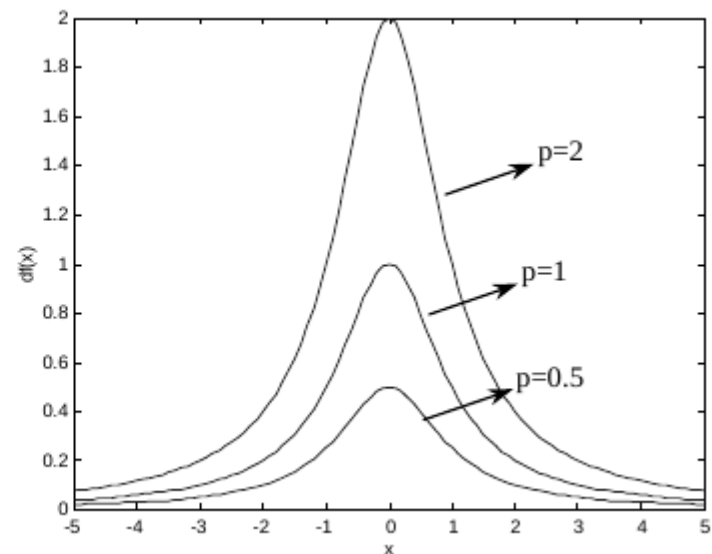
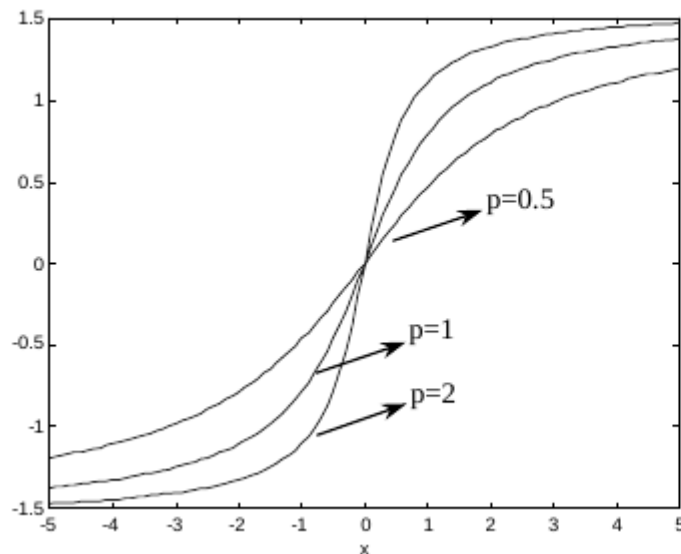
$$f(x) = \frac{e^{px} - e^{-px}}{e^{px} + e^{-px}} = \tanh(px), \quad f'(x) = p(1 - f(x)^2)$$



Redes Neurais Artificiais

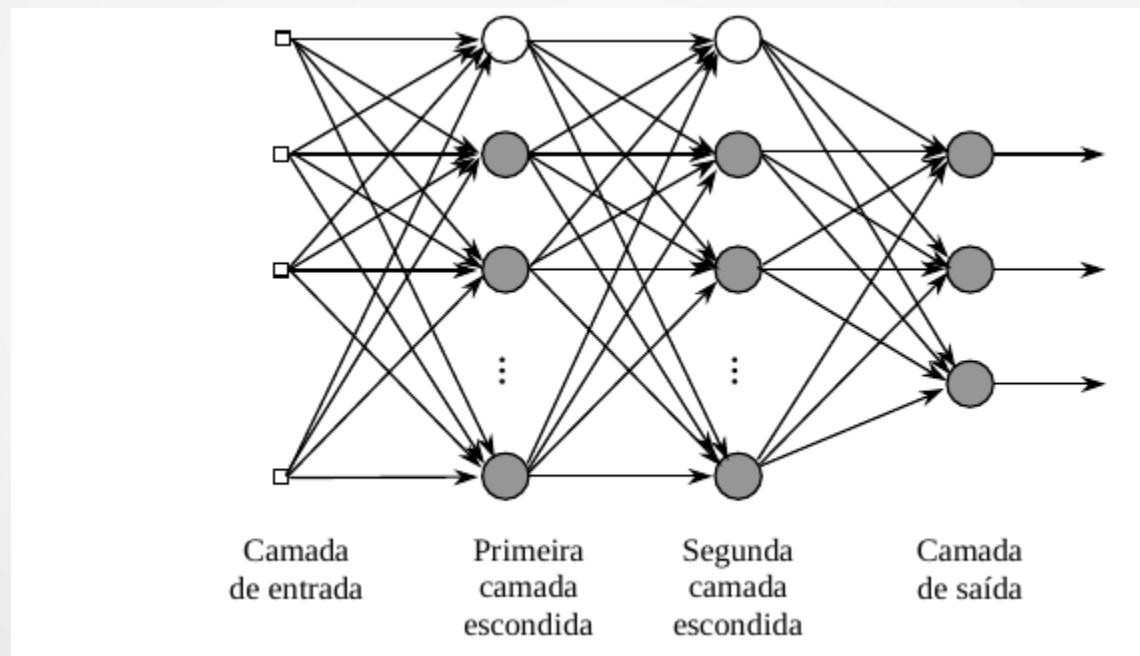
Função de Ativação : Função Arco-Tangente

$$f(x) = \text{atan}(px) \quad f'(x) = p \cdot \frac{1}{1+x^2}$$



Redes Neurais Artificiais

Arquitetura MLP: duas camadas intermediárias



Redes Neurais Artificiais

Nesta rede, dois tipos de sinais podem ser identificados:

- **Sinal funcional:** um sinal funcional é um sinal de entrada (estímulo) que chega na entrada e é propagado positivamente (neurônio a neurônio) através da rede, e aparece na saída como um sinal de saída.
- **Sinal de erro:** os sinais de erro originam-se nas saídas e são retro-propagados (neurônio a neurônio) através da rede.

Redes Neurais Artificiais

Características Principais

- A camada de entrada geralmente é composta por neurônios sensoriais, ou seja, unidades que não modificam os sinais externos, apenas os distribuem para a primeira camada intermediária.
- As unidades de saída constituem a camada de saída da rede, e as demais unidades constituem as camadas intermediárias.
- As camadas intermediárias são todas aquelas que não fazem parte da entrada e nem da saída.

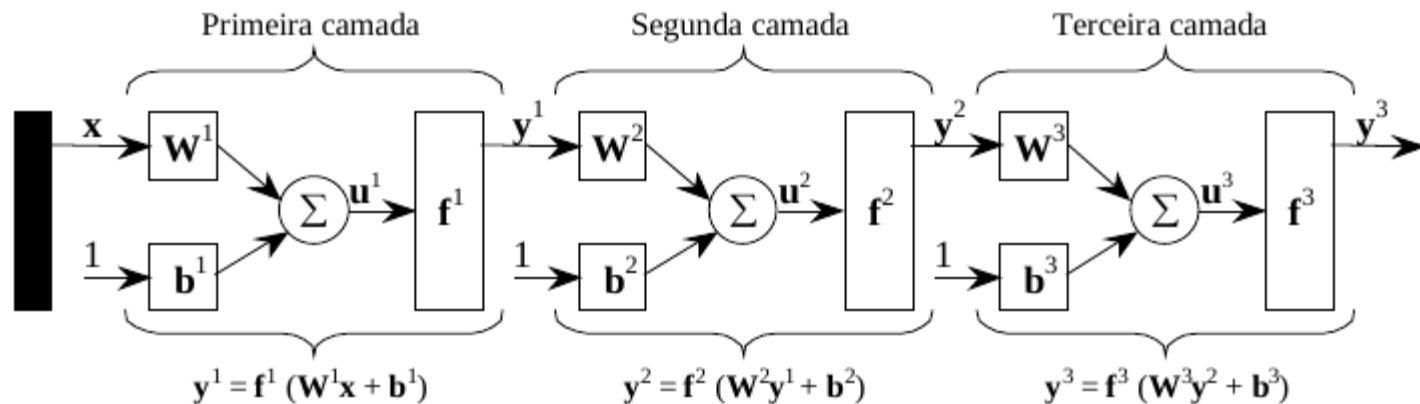
Redes Neurais Artificiais

Cada unidade intermediária ou de saída é responsável por duas tarefas:

- Calcular o sinal na saída da unidade, que geralmente é expresso como uma função não-linear do sinal de entrada e pesos sinápticos associados e
- Calcular uma estimativa instantânea do vetor gradiente, que é necessário para a retro-propagação do erro através da rede.

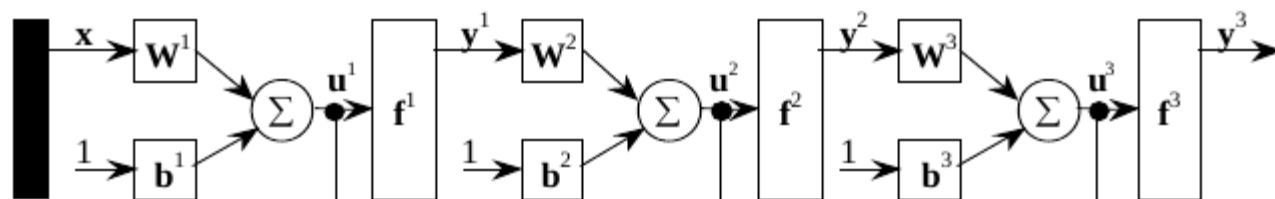
Redes Neurais Artificiais

Características Principais Rede com três camadas

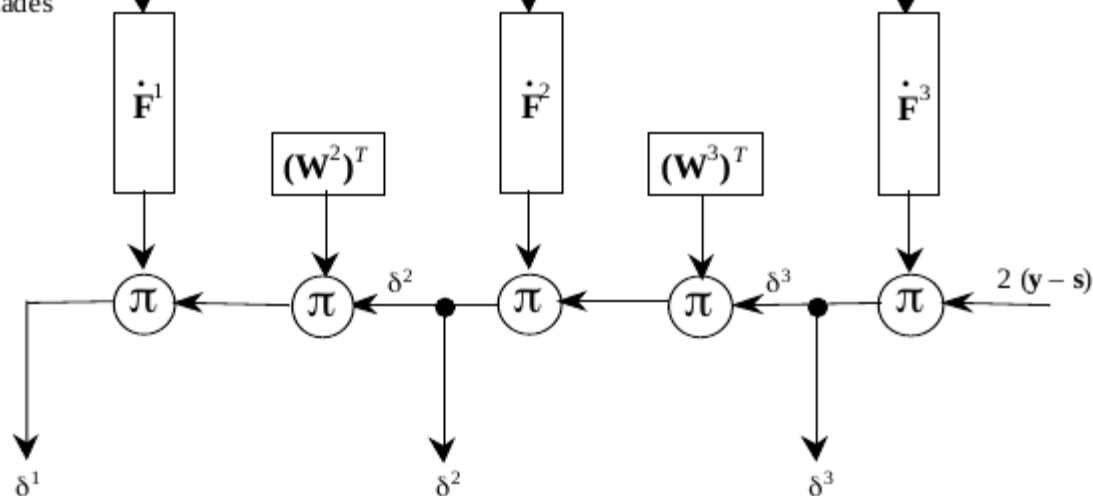


Redes Neurais Artificiais

Propagação dos sinais



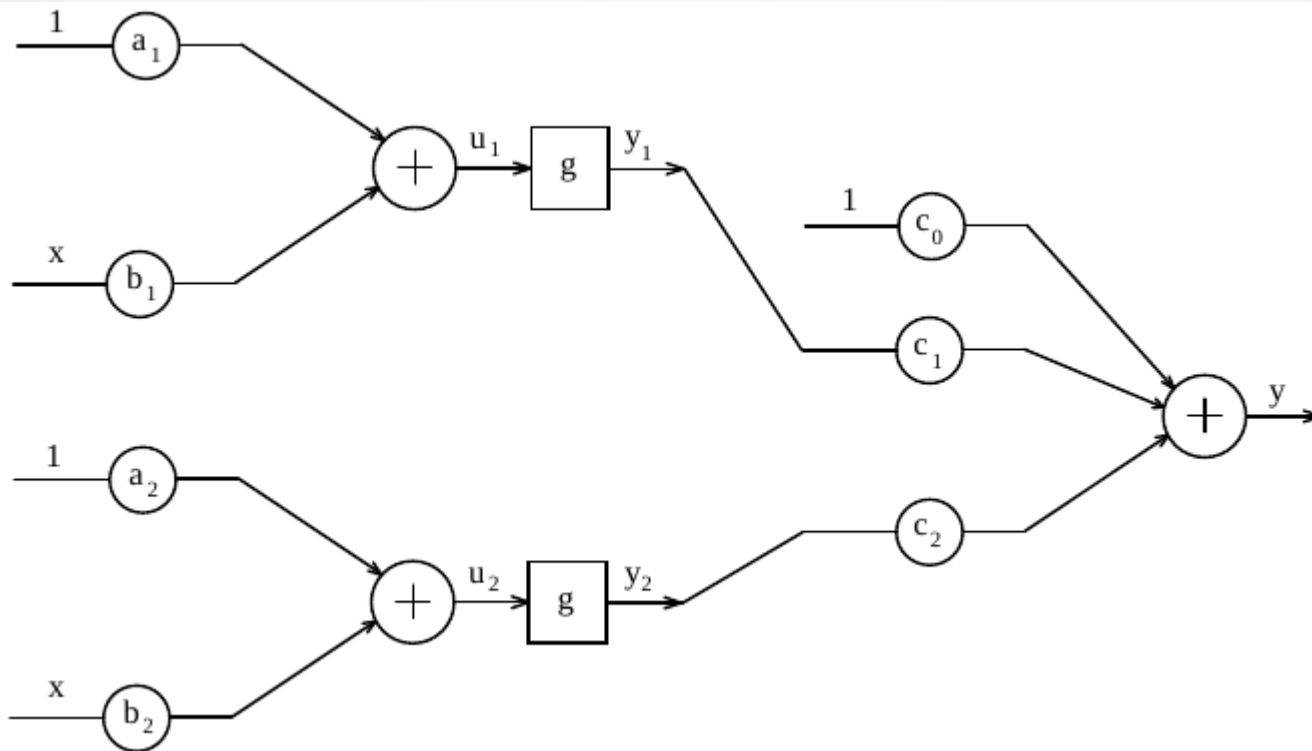
Retro-propagação das sensibilidades



Redes Neurais Artificiais

O papel dos pesos sinápticos

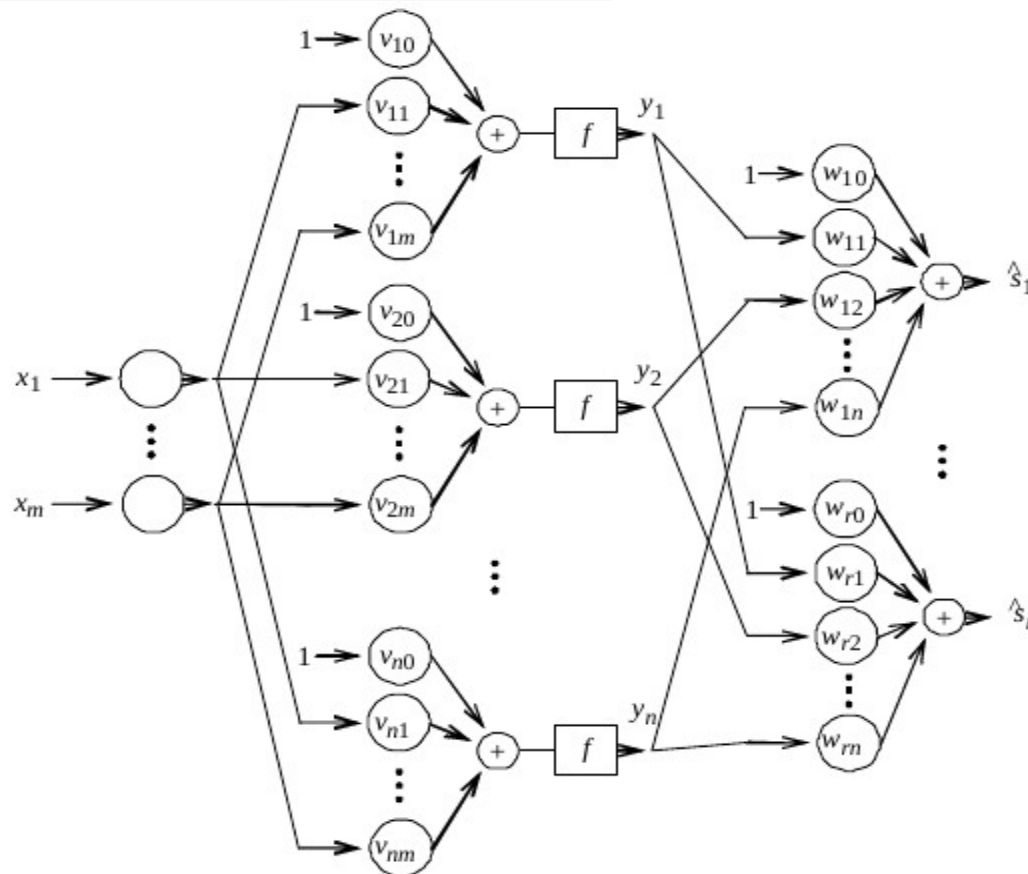
$$y = c_0 + \sum_{n=1}^p c_n g(b_n x + a_n)$$



Redes Neurais Artificiais

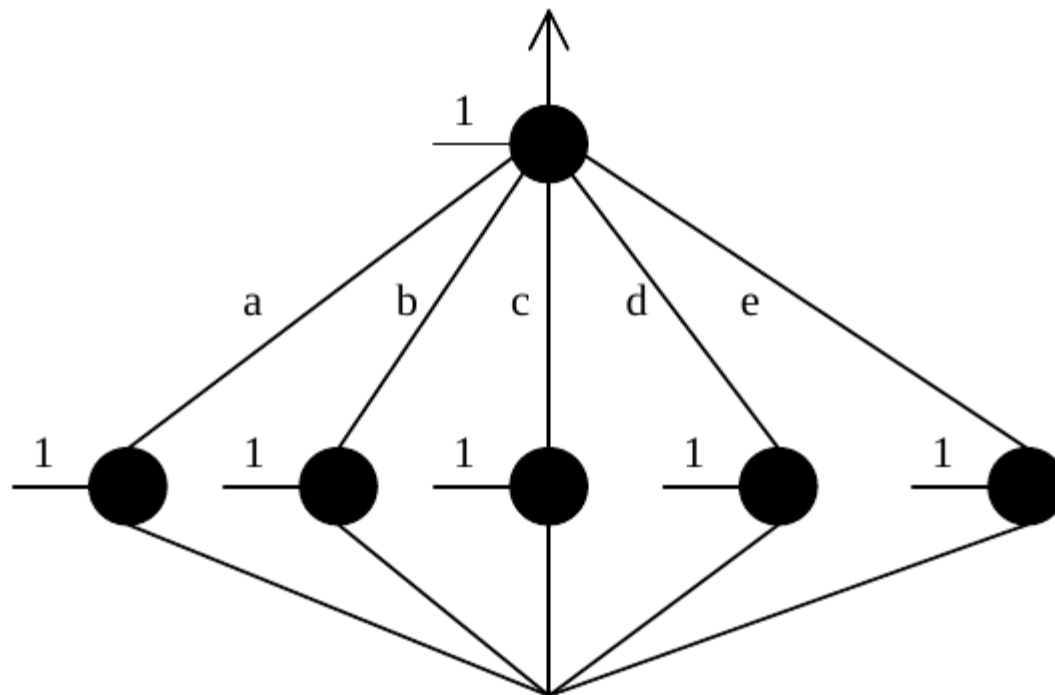
O papel dos pesos sinápticos

$$y = c_0 + \sum_{n=1}^p c_n g(b_n x + a_n)$$



Redes Neurais Artificiais

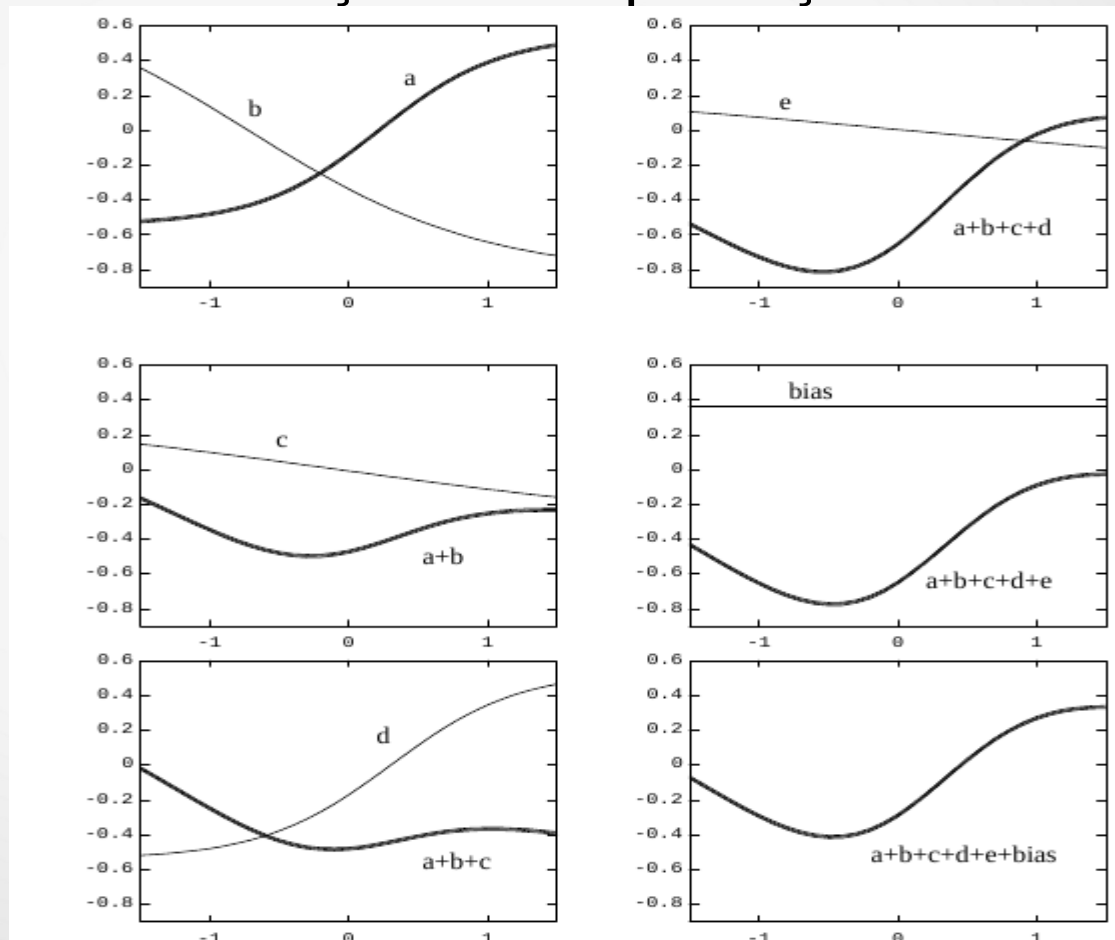
O papel dos pesos sinápticos



$$f(\mathbf{w}) = \underbrace{c_1 g(b_1 x + a_1)}_a + \underbrace{c_2 g(b_2 x + a_2)}_b + \underbrace{c_3 g(b_3 x + a_3)}_c + \underbrace{c_4 g(b_4 x + a_4)}_d + \underbrace{c_5 g(b_5 x + a_5)}_e + \underbrace{c_0}_{bias}$$

Redes Neurais Artificiais

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



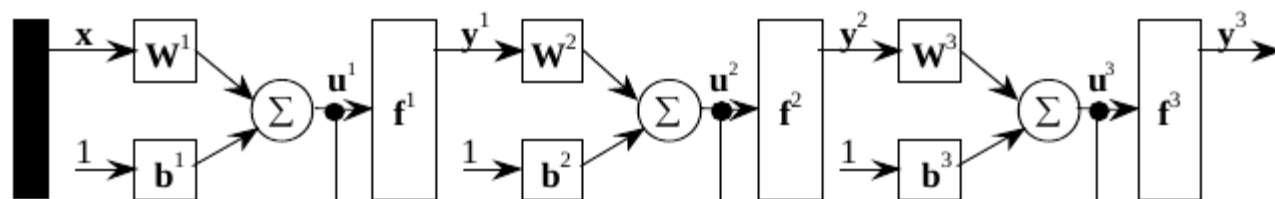
Redes Neurais Artificiais

Retro Propagação dos Erros

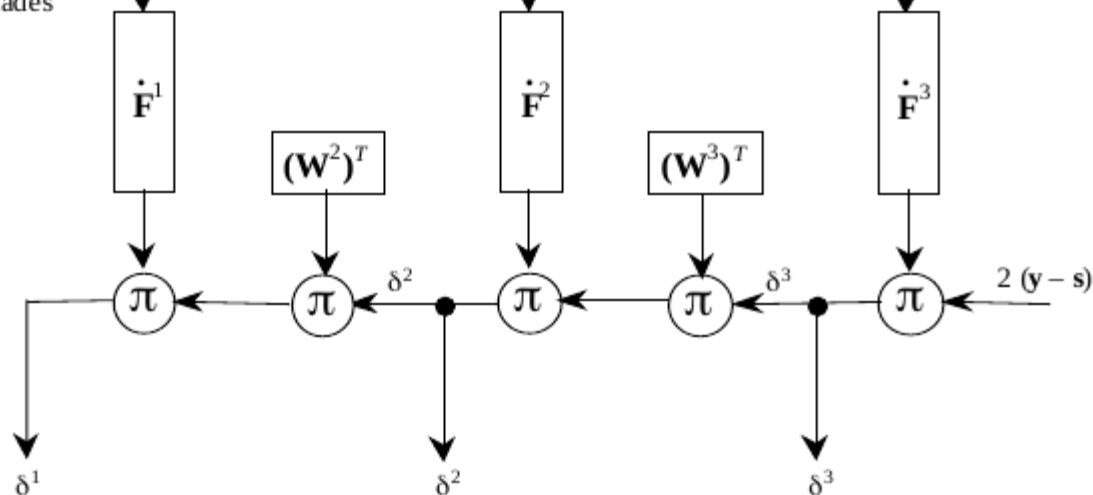
- Em aplicações práticas do algoritmo de retro-propagação, o aprendizado é resultado de apresentações repetidas de todas as amostras do conjunto de treinamento ao MLP;
- Cada apresentação de todo o conjunto de treinamento durante o processo de aprendizagem é chamada de época. O processo de aprendizagem é repetido época após época, até que o conjunto de pesos e limiares estabilize e o erro quadrático médio do conjunto de treinamento convirja para um valor mínimo;
- É uma boa prática fazer com que a ordem de apresentação das amostras seja feita aleatoriamente de uma época para a outra. Esta aleatoriedade tende a fazer com que a busca no espaço de pesos tenha um caráter estocástico ao longo dos ciclos de treinamento;

Redes Neurais Artificiais

Propagação dos sinais



Retro-propagação das sensibilidades



Redes Neurais Artificiais

RetroPropagação dos Erros

A função g que associa a cada vetor de entrada $x \in X$ uma saída escalar $s \in R$ pode ser aproximada com base no conjunto de dados de aproximação $\{(x, s) \in R^m \times R^n\}_{i=1}^n$ por uma composição aditiva de funções de expansão ortogonal na forma:

$$\hat{s}_l = \hat{g}(\mathbf{x}_l, \theta) = \sum_{j=0}^n w_j f\left(\sum_{i=0}^m v_{ji} x_{li}\right) = \sum_{j=0}^n w_j f(\mathbf{v}_j^T \mathbf{x}_l)$$

onde θ é o vetor contendo todos os pesos da rede neural;

Redes Neurais Artificiais

- Logo, o erro quadrático médio produzido na saída da rede neural, considerando as N amostras, assume a forma:

$$\begin{aligned} J(\theta) &= \frac{1}{N} \sum_{l=1}^N (\hat{s}_l - s_l)^2 = \frac{1}{N} \sum_{l=1}^N (\hat{g}(\mathbf{x}_l, \theta) - s_l)^2 = \\ &= \frac{1}{N} \sum_{l=1}^N \left(\sum_{j=0}^n w_j f \left(\sum_{i=0}^m v_{ji} x_{li} \right) - s_l \right)^2 = \frac{1}{N} \sum_{l=1}^N \left(\sum_{j=0}^n w_j f(\mathbf{v}_j^T \mathbf{x}_l) - s_l \right)^2 \end{aligned}$$

- Sendo P a dimensão do vetor θ , então tem-se que: $J : \mathbb{R}^P \rightarrow \mathbb{R}^1$.
- A superfície de erro definida por $J(\theta)$ reside no espaço \mathbb{R}^{P+1} , sendo que deve-se buscar em \mathbb{R}^P um ponto que minimiza $J(\theta)$, supondo que se queira minimizar o erro entre a saída produzida pelo rede neural e a saída desejada.

Redes Neurais Artificiais

- Aprendizado a partir de dados amostrados ;
- O aprendizado supervisionado visto como um problema de otimização não-linear ;
- A função objetivo (critério de desempenho a ser otimizado) e os parâmetros ajustáveis:

$$\min_{\theta} J(\theta)$$

Redes Neurais Artificiais

- Expansão em série de Taylor em torno do ponto $\mathbf{x}^* \in \mathbb{R}^n$:

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \nabla^2 f(\mathbf{x}^*) (\mathbf{x} - \mathbf{x}^*) + O(3)$$

$$\nabla f(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial f(\mathbf{x}^*)}{\partial x_1} \\ \frac{\partial f(\mathbf{x}^*)}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x}^*)}{\partial x_n} \end{bmatrix} \quad \nabla^2 f(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_2^2} & & \vdots \\ & & \ddots & \\ \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_n \partial x_1} & \dots & & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_n^2} \end{bmatrix}$$

Vetor gradiente

Matriz hessiana

Redes Neurais Artificiais

RetroPropagação dos Erros

- O algoritmo de retropopagação do erro (do inglês Backpropagation) é empregado para obter o vetor gradiente, onde cada elemento do vetor gradiente está associado a um peso da rede neural e indica o quanto a saída é influenciada por uma variação incremental neste peso;
- Existem várias técnicas para obter exatamente ou aproximadamente a informação de 2a. ordem em Redes Neurais Artificiais MLP;

Redes Neurais Artificiais

RetroPropagação dos Erros

→ O processo de otimização não-linear envolvido no ajuste de pesos de uma rede neural vai realizar aproximações locais de primeira ordem ou de primeira e segunda ordem junto à superfície de erro e realizar ajustes incrementais e recursivos na forma:

$$\theta_{k+1} = \theta_k + \text{passo} * \text{direção}$$

→ Parte-se de uma condição inicial θ_0 e aplica-se iterativamente a fórmula acima, sendo que a direção depende da informação local de primeira e segunda ordem. Cada proposta de algoritmo de otimização vai diferir na forma de computar o passo e a direção de ajuste, a cada iteração.

Redes Neurais Artificiais

Treinamento Local ou em Lote

→ **Atualização local:** neste método, a atualização dos pesos é feita imediatamente após a apresentação de cada amostra de treinamento

→ **Atualização em lote:** no método em lote, a atualização dos pesos só é feita após a apresentação de todas as amostras de treinamento que constituem uma época. O ajuste relativo a cada apresentação de um par de entrada é acumulado. Este método também é conhecido como método de atualização off-line ou batch.

Redes Neurais Artificiais

Treinamento Local

- Defina uma condição inicial para o vetor de pesos \mathbf{w} e escolha um passo α pequeno;
- Faça $k = 0$ e calcule $J(\mathbf{w}(k))$;
- Enquanto o critério de parada não for atendido, faça:
 - ◆ Ordene aleatoriamente os padrões de entrada-saída;
 - ◆ Para l variando de 1 até N , faça:
 - Apresente o padrão l de entrada à rede;
 - Calcule $J_l(\mathbf{w}(k))$ e $\nabla J_l(\mathbf{w}(k))$;
 - $\mathbf{w}(k+1) = \mathbf{w}(k) - \alpha \nabla J_l(\mathbf{w}(k))$;
 - ◆ $k = k + 1$;
 - ◆ Calcule $J(\mathbf{w}(k))$;

Redes Neurais Artificiais

Treinamento em Lote

- Defina uma condição inicial para o vetor de pesos \mathbf{w} e escolha um passo α pequeno;
- Faça $k = 0$ e calcule $J(\mathbf{w}(k))$;
- Enquanto o critério de parada não for atendido, faça:
 - ◆ Para l variando de 1 até N , faça:
 - Apresente o padrão l de entrada à rede;
 - Calcule $J_l(\mathbf{w}(k))$ e $\nabla J_l(\mathbf{w}(k))$;
 - ◆ $\mathbf{w}(k+1) = \mathbf{w}(k) - \frac{\alpha}{N} \sum_{l=1}^N \nabla J_l(\mathbf{w}(k))$;
 - ◆ $k = k + 1$;
 - ◆ Calcule $J(\mathbf{w}(k))$;

Redes Neurais Artificiais

Treinamento Local ou em Lote

- Do ponto de vista operacional, o método local de treinamento é preferível ;
- Requer um menor armazenamento local para cada conexão. Como as amostras são apresentadas aleatoriamente, o uso da atualização padrão por padrão torna a busca no espaço de conexões estocástica por natureza, reduzindo a possibilidade do algoritmo ficar preso em um mínimo local.
- Por outro lado, a utilização do método em lote fornece uma estimativa mais precisa do vetor gradiente. Numa análise final, entretanto, a eficiência dos dois métodos depende do problema a ser tratado ;

Redes Neurais Artificiais

Cr terios de Parada

- O processo de minimiza  o do MSE (ou da fun  o custo), em geral, n o tem converg ncia garantida.
- Um crit rio de parada n o recomend vel, por n o levar em conta o estado do processo iterativo de treinamento,   interromper o treinamento ap s um n mero fixo de itera  es.

Seja θ^* o vetor de par metros (pesos) que denota um ponto de m nimo, seja ele local ou global. Uma condi  o necess ria para que θ^* seja um m nimo,   que o vetor gradiente $\nabla J(\theta)$ (ou seja, a derivada parcial de primeira ordem) da superf cie de erro em rela  o ao vetor de pesos θ seja zero em $\theta = \theta^*$. Sendo assim,   poss vel formular crit rios de converg ncia definindo que θ seja um valor bem pequeno.

Redes Neurais Artificiais

A Capacidade de Aproximação Universal

- Uma arquitetura do tipo MLP pode ser vista como uma ferramenta prática geral para fazer um mapeamento não-linear de entrada-saída;
- A relação entrada-saída da rede define um mapeamento de um espaço euclidiano de entrada k -dimensional para um espaço euclidiano de saída m -dimensional, que é infinitamente continuamente diferenciável;

Redes Neurais Artificiais

A Capacidade de Aproximação Universal

Teorema: *Seja $f(\cdot)$ uma função contínua não-constante, limitada, e monotonicamente crescente. Seja I_p um hipercubo unitário p -dimensional $(0,1)^p$. O espaço das funções contínuas em I_p é denominado $C(I_p)$. Então, dada qualquer função $g \in C(I_p)$ e $\varepsilon > 0$, existe um inteiro M e conjuntos de constantes reais α_i e w_{ij} , onde $i = 1, \dots, M$ e $j = 1, \dots, p$, tais que pode-se definir*

$$F(x_1, x_2, \dots, x_p) = \sum_{i=1}^M \alpha_i f \left(\sum_{j=1}^p w_{ij} x_j - w_{0i} \right), \quad (2.35)$$

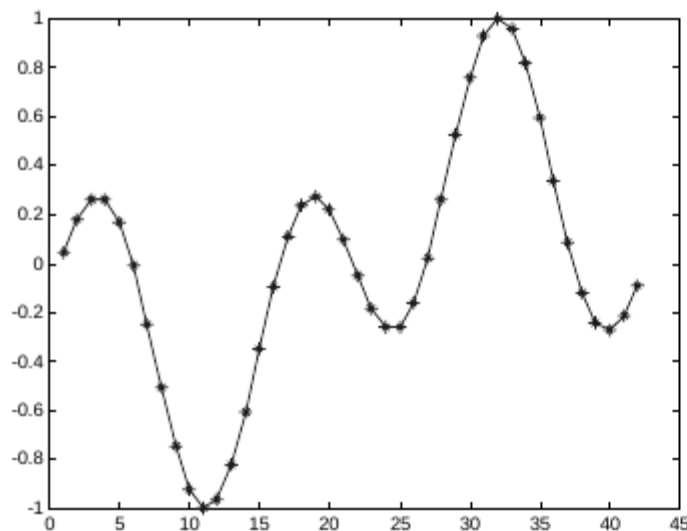
como uma aproximação da função $g(\cdot)$ tal que,

$$\left| F(x_1, x_2, \dots, x_p) - g(x_1, x_2, \dots, x_p) \right| < \varepsilon$$

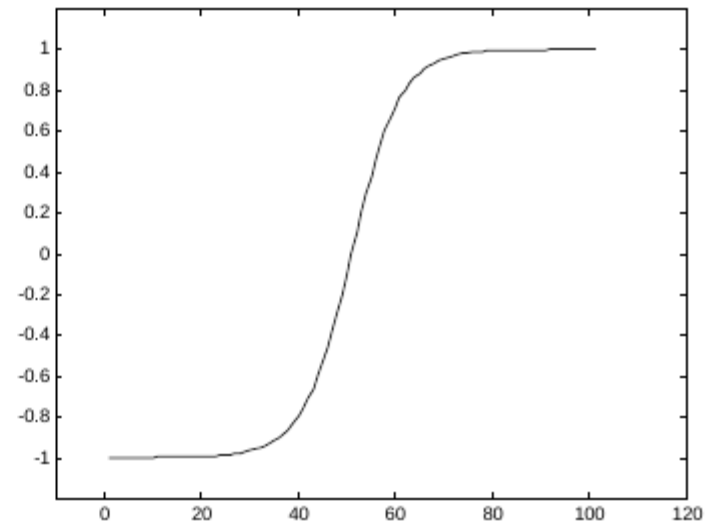
Para todo $\{x_1, \dots, x_p\} \in I_p$.

Redes Neurais Artificiais

Exemplo:



(a)



(b)

Figura 2.15: Aproximação universal. (a) Função a ser aproximada (informação disponível: 42 amostras do valor da função em pontos igualmente espaçados do domínio da função). (b) Função básica utilizada na aproximação. Tangentes hiperbólicas, satisfazendo as condições estabelecidas pelo teorema.

Redes Neurais Artificiais

Referencias Bibliograficas

<http://www.inf.ufg.br/node/101> - Prof. Cedric Luiz de
Carvalho - Instituto de Informática - UFG 2006

Prof. Fernando J. Von Zuben – DCA/FEEEC/Unicamp -
<http://www.dca.fee.unicamp.br/~vonzuben/>

Braga, Antonio De Padua; Ludermir, André Ponce De “
Redes Neurais Artificiais - Teoria e Prática – LTC. 2ª Ed.
2011