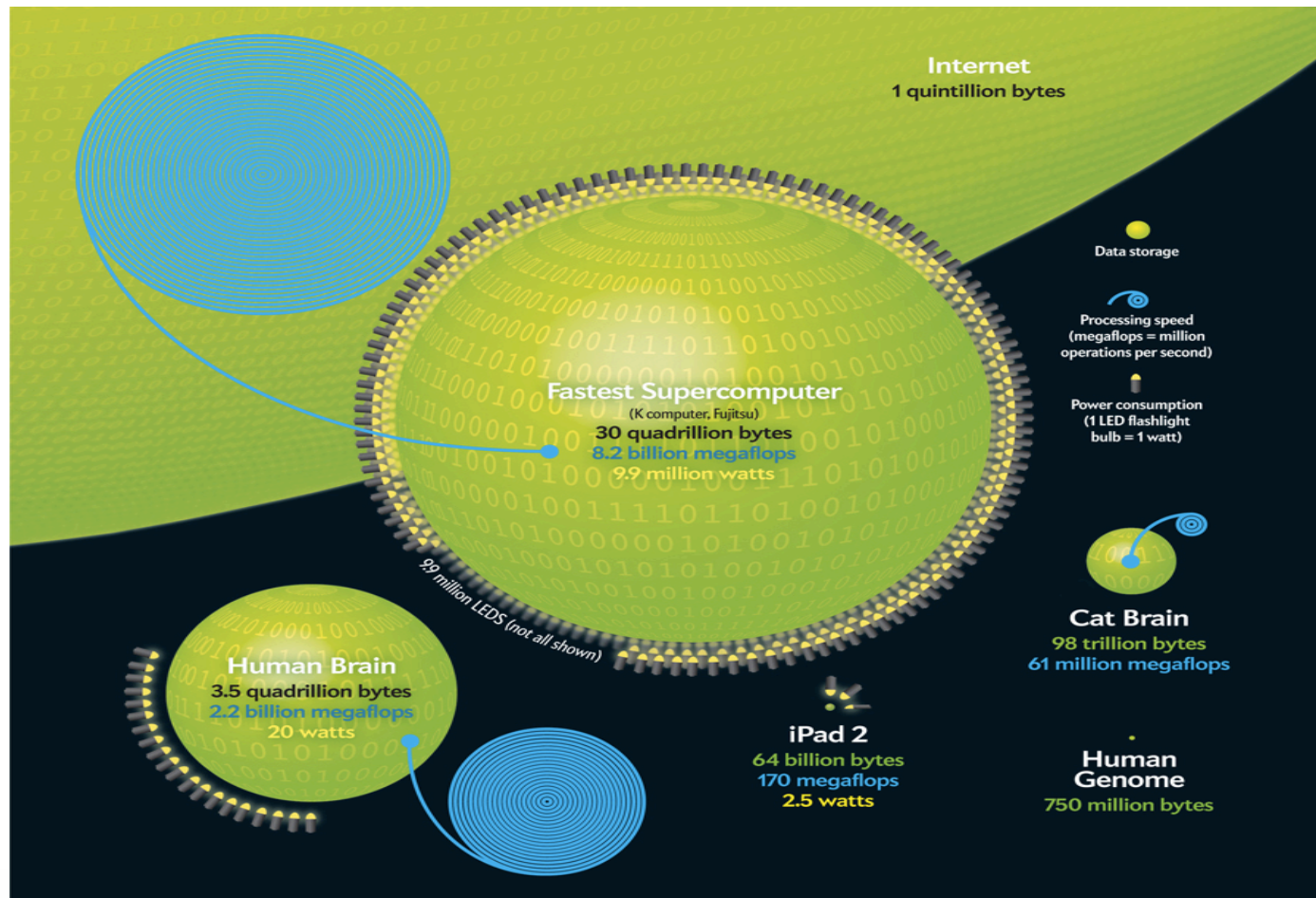
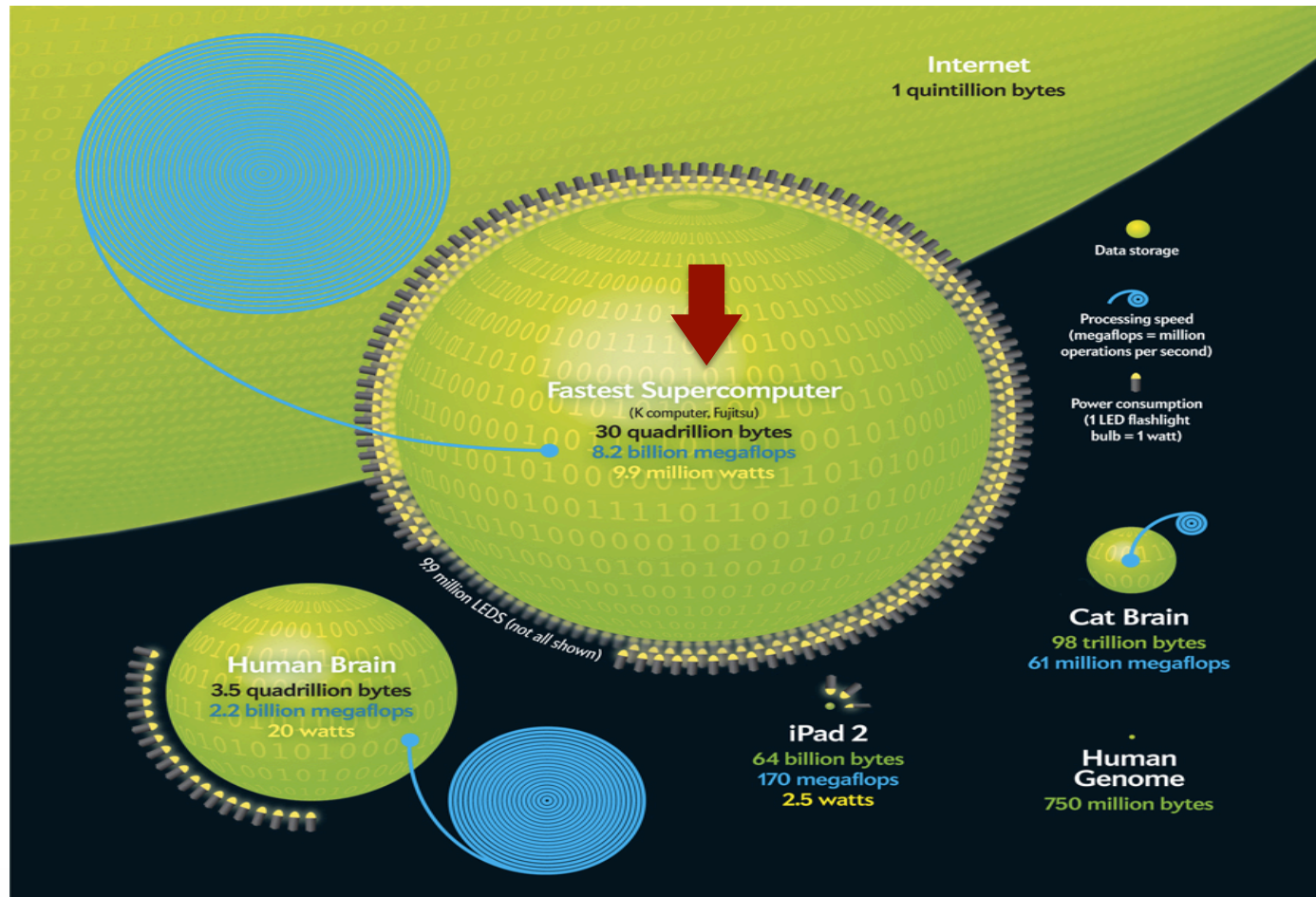
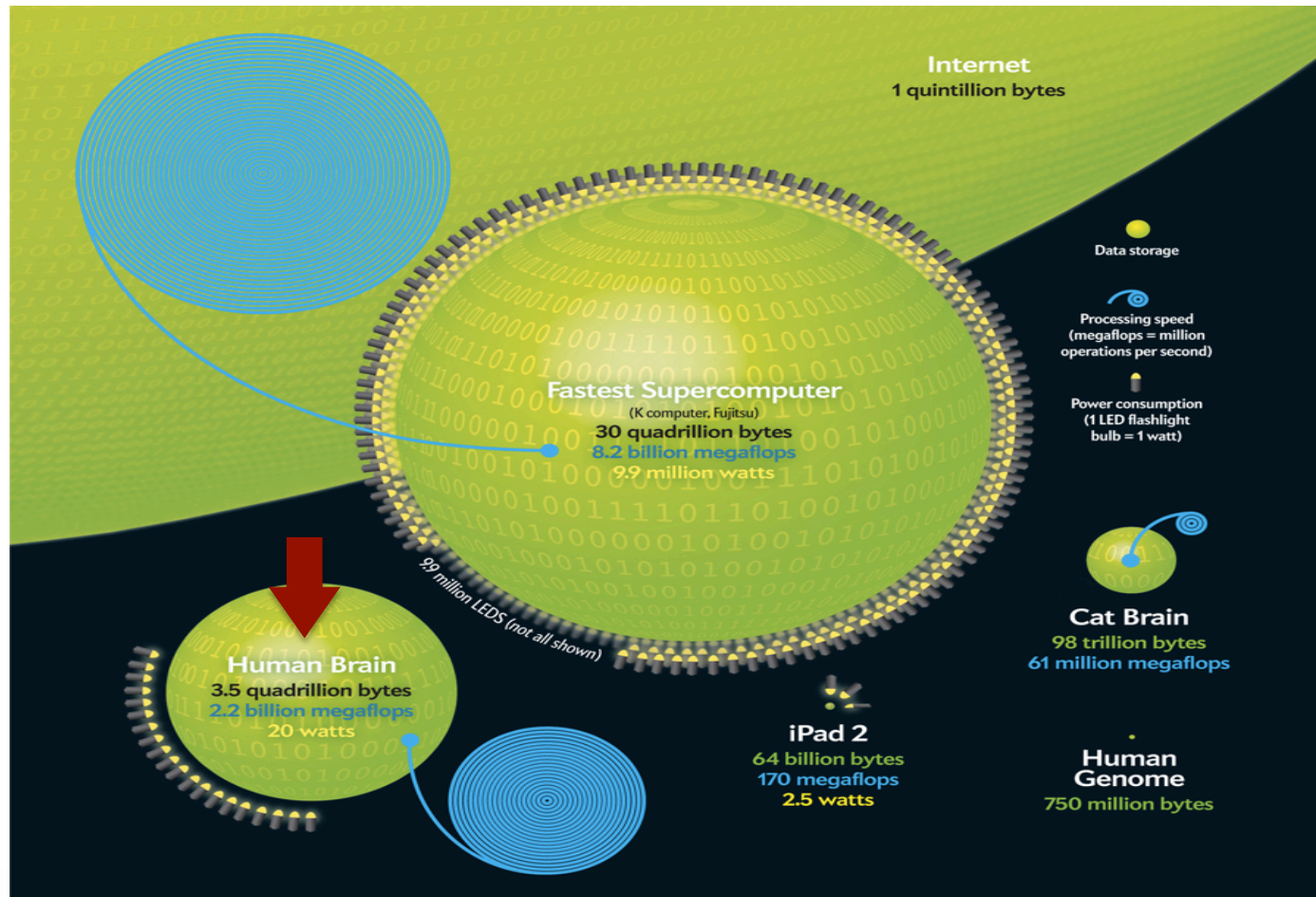


Redes Neurais Artificiais

Gisele L. Pappa







Definição

“Redes Neurais Artificiais são dispositivos de computação paralela que consistem de muitos processadores interconectados. Esses processadores são simples... Cada processador está ciente apenas dos sinais que manda para e recebe de outros processadores periodicamente”

(Callan, “The Essence of Neural Networks”)

De onde veio a inspiração?

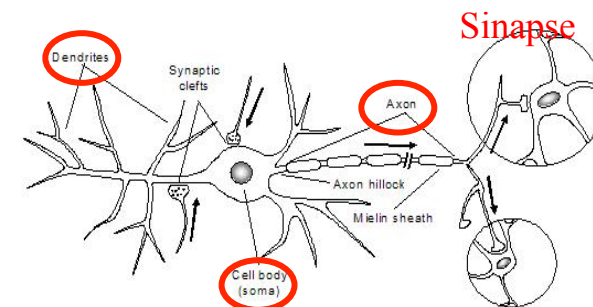
- O processamento que ocorre no cérebro é altamente **paralelo**
 - Paralelismo é uma das características de arquitetura mais difíceis de serem computacionalmente bem exploradas
 - Entender como o cérebro “computa” pode nos ajudar

De onde veio a inspiração?

- Os “processadores” do cérebro (neurônios) são muito lentos se comparados com dispositivos eletrônicos.
- A ideia é produzir neurônios artificiais que usem a mesma arquitetura de processamento do cérebro, mas que sejam mais rápidos que os neurônios naturais

Neurônios

- Processamento de sinal
- Dependendo das condições do ambiente, neurônios geram sinais (potenciais elétricos) que são utilizados para transmitir “informações” a outros neurônios aos quais ele está conectado.



De Neurônios a Redes Neurais

- Neurônios podem ter conexões “para frente” (*forward*) ou “por realimentação” (*feedback*) com outros neurônios
- Essas interconexões dão origem às **redes neurais** (neuronaes)
- Representação da informação é feita de maneira distribuída, e seu processamento é paralelo

Redes Neurais Artificiais

- Processamento da informação ocorre nos neurônios
- Neurônios recebem e enviam estímulos do/para outros neurônios e de/para o ambiente
- Neurônios podem ser conectados formando redes neurais

Redes Neurais Artificiais

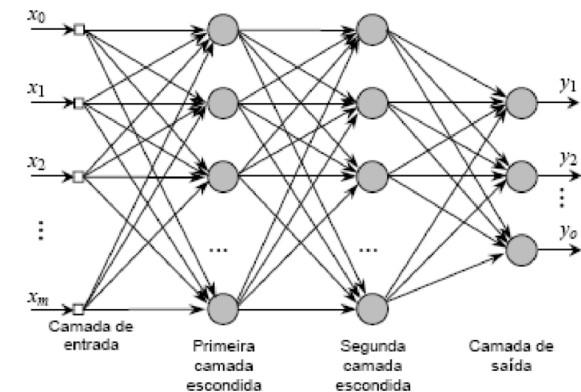
- Informação é transmitida através de sinapses
- A eficiência de uma sinapse é representada por um peso, que corresponde a informação armazenada no neurônio
- Conhecimento é adquirido através de aprendizado
 - Adaptação dos pesos das sinapses de acordo com a informação do ambiente

Redes Neurais Artificiais

- Caracterizadas por 3 elementos:
 - Conjunto de neurônios artificiais
 - Um padrão de conexão entre os neurônios
 - Arquitetura ou estrutura da rede
 - Método que determina o valor dos pesos
 - Algoritmo de treinamento ou aprendizagem

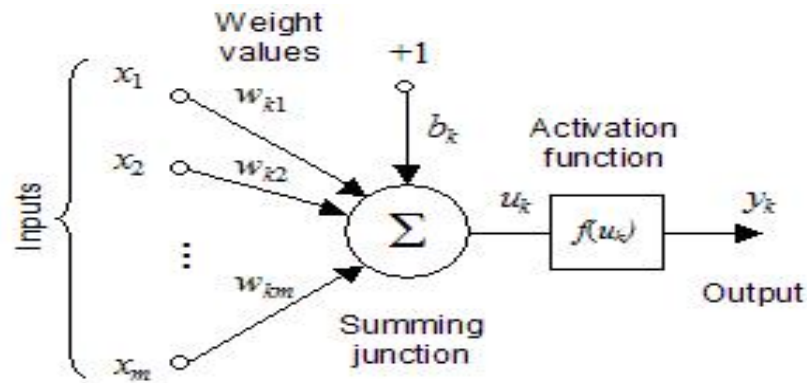
Redes Neurais Artificiais

- Caracterizadas por 3 elementos:
 - Conjunto de neurônios artificiais
 - Um padrão de conexão entre os neurônios
 - Arquitetura ou estrutura da rede
 - Método que determina o valor dos pesos
 - Algoritmo de treinamento ou aprendizagem



Neurônios Artificiais

- Alto nível de abstração em relação a neurônios biológicos



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

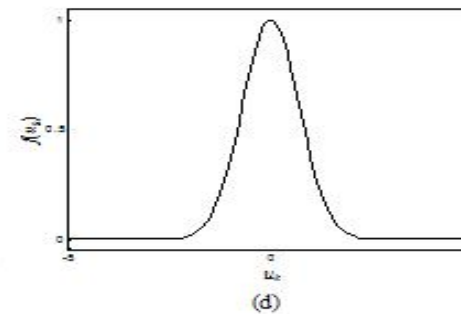
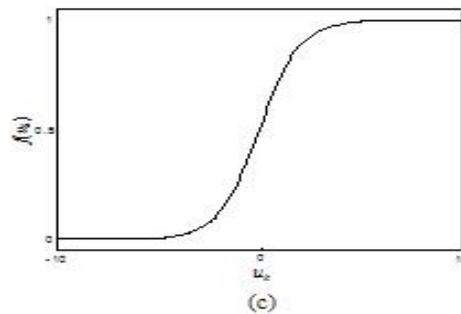
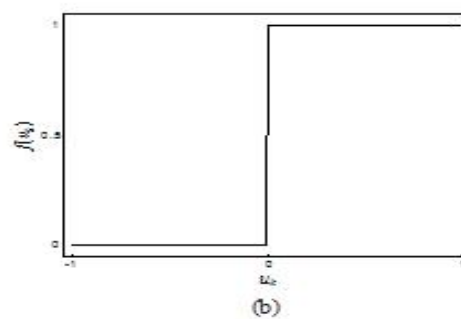
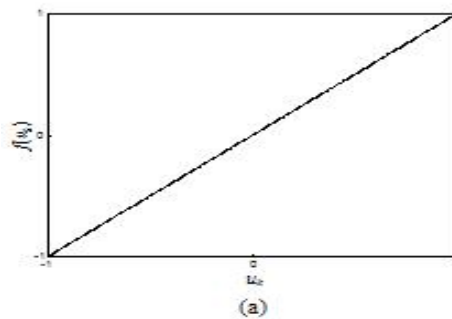
Ou, assumindo que $w_{k0} = b_k$

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

- Bias do neurônio: Aumenta ou diminui a entrada da rede para a função de ativação (provoca um deslocamento da função de ativação)

Neurônio

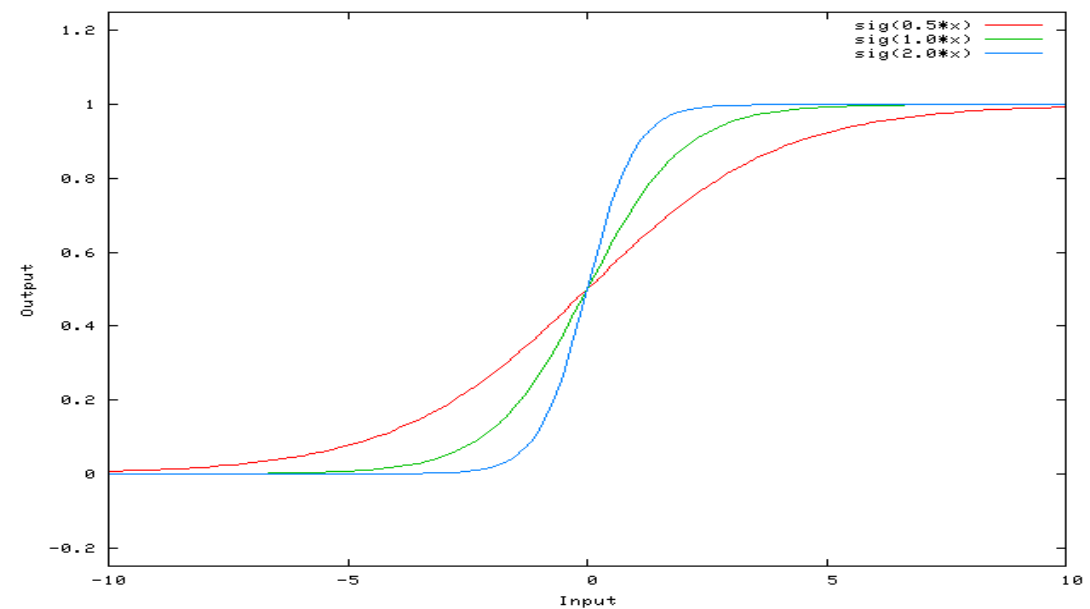
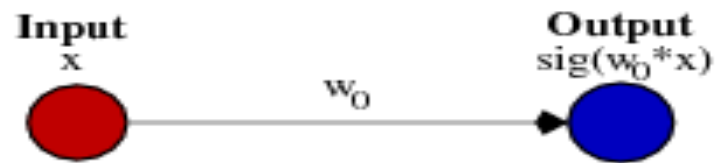
- Funções de ativação



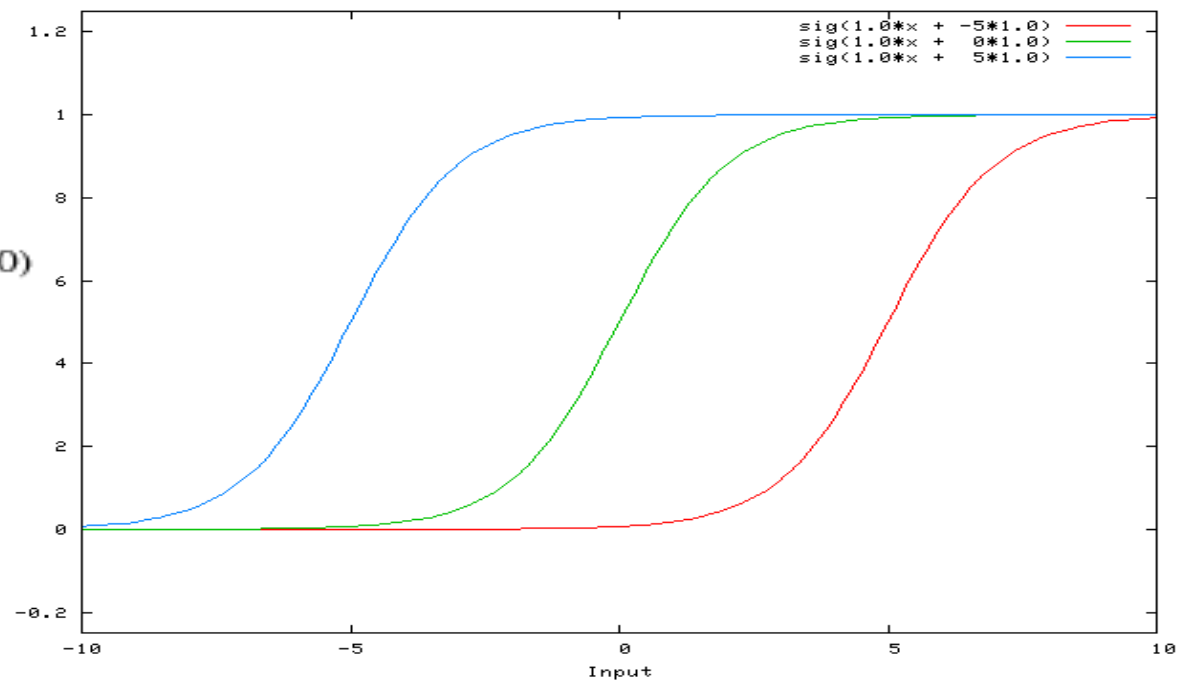
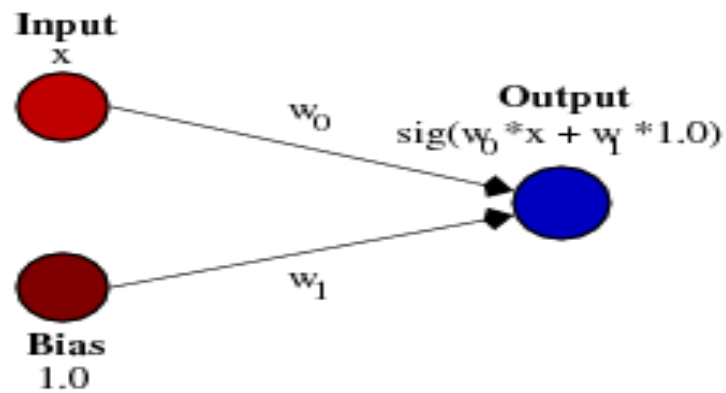
- (a) Linear
- (b) Limiar
- (c) Sigmoidal (logística)
- (d) Gaussiana

- A sigmoide é a função mais utilizada na literatura

Qual o papel do peso e do *bias*?



Qual o papel do peso e do *bias*?



Redes Neurais Artificiais

- Caracterizadas por 3 elementos:
 - Conjunto de neurônios artificiais
 - Um padrão de conexão entre os neurônios
 - Arquitetura ou estrutura da rede
 - Método que determina o valor dos pesos
 - Algoritmo de treinamento ou aprendizagem

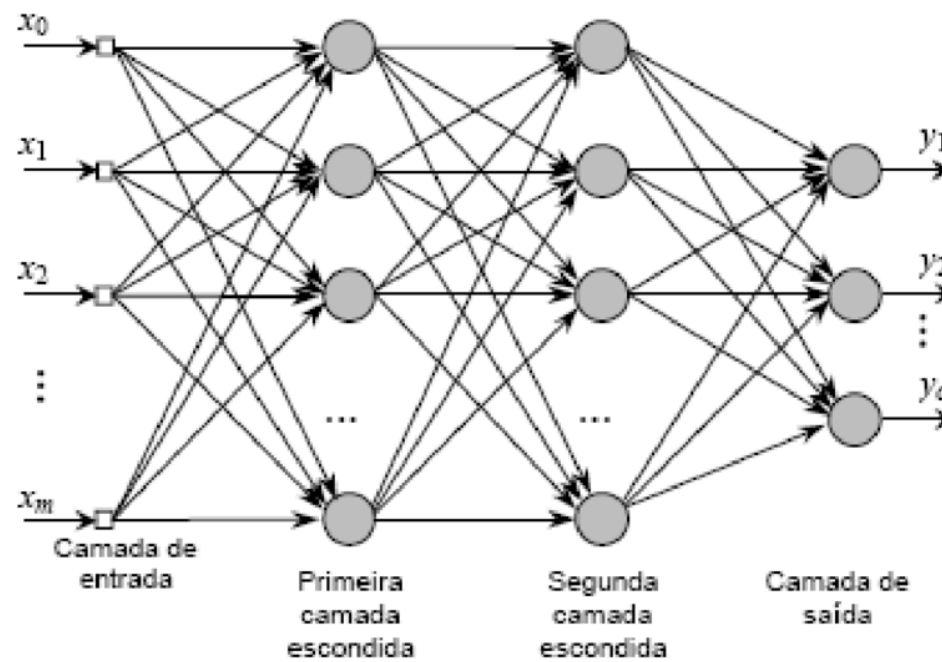
Arquiteturas de Rede

- Sistemas naturais
 - Não se conhece muito sobre como os neurônios estão conectados no cérebro
 - Comportamento de um neurônio pode afetar o comportamento de outros
 - Individualmente, o comportamento de um neurônio não leva a nenhuma conclusão
 - Interação entre eles leva a um comportamento emergente

Arquiteturas de Rede

- Existe uma arquitetura padrão para redes
 - Uma camada de entrada
 - Uma ou mais camadas intermediárias (escondidas)
 - Uma camada de saída
- A maneira como os neurônios estão conectados é fortemente dependente do algoritmo de aprendizado utilizado para treinar a rede

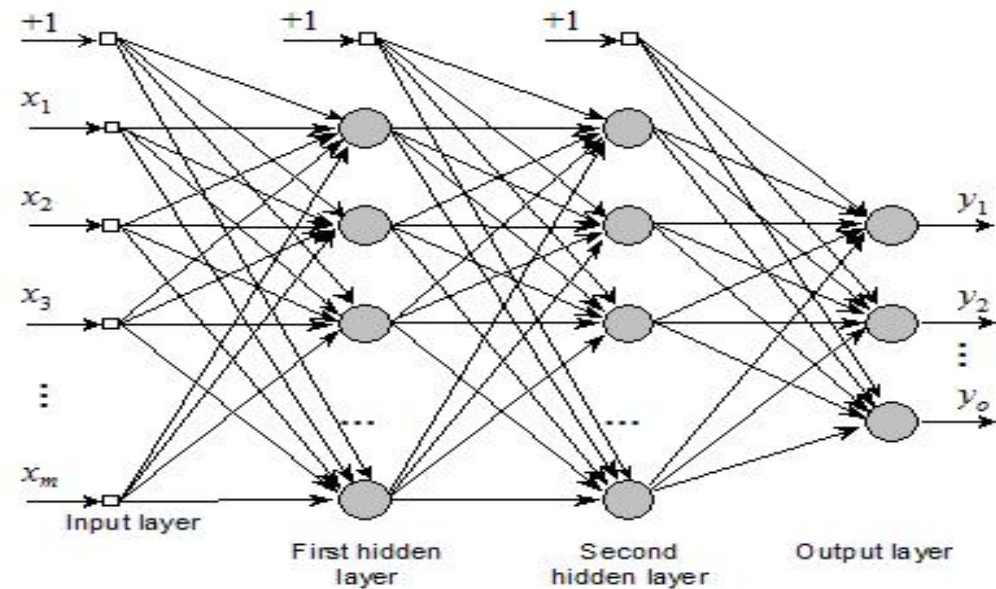
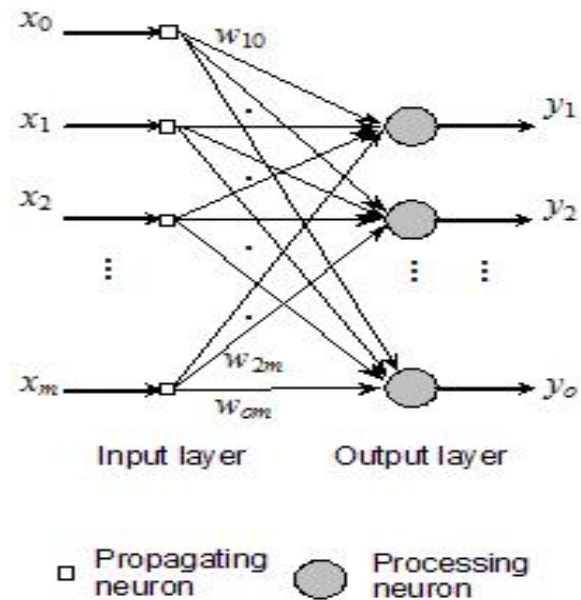
Arquiteturas de Rede



Arquiteturas de Rede

- Existem 3 tipos principais de arquitetura:
 - Redes *feedforward* de uma camada
 - Redes *feedforward* de multi-camadas
 - Redes recorrentes


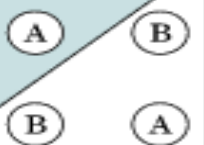


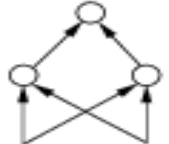
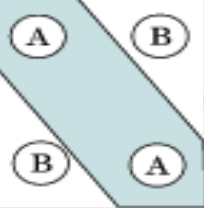
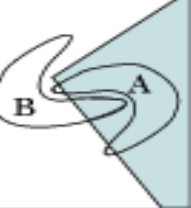


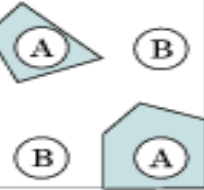


Redes *feedforward*



Redes *feedforward*

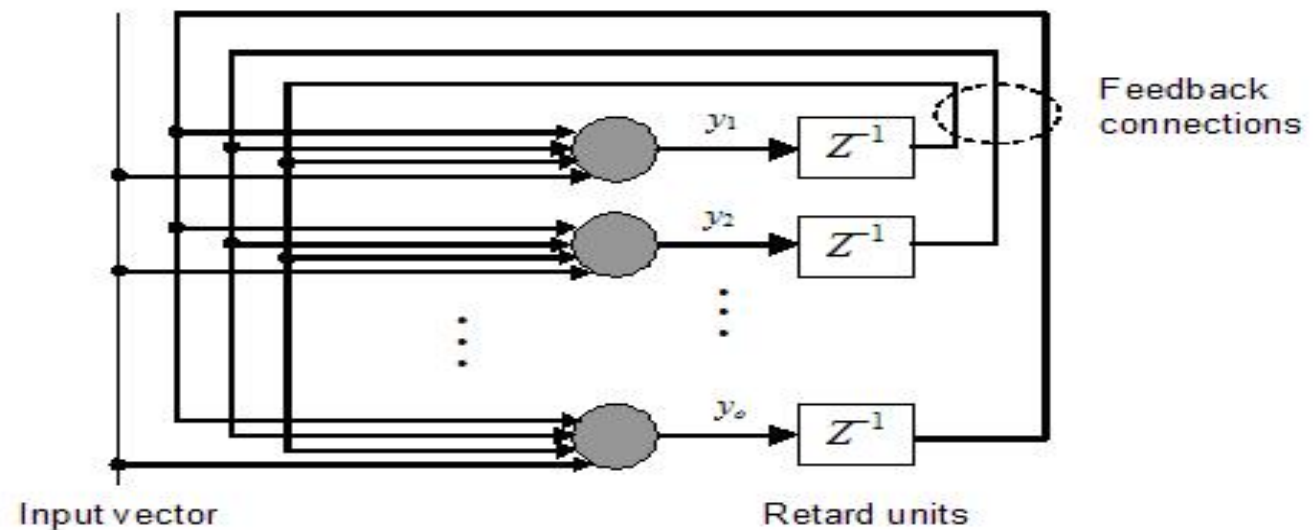
- Redes de uma camada
 - Normalmente as funções de entrada são lineares
 - Simplesmente propagam sinal para próxima camada
- Redes multi-camada
 - Inserem-se funções não lineares na camada oculta

O que uma camada a mais faz?

Structure	Types of decision regions	Exclusive OR problem	Classes with meshed regions	Most general region shapes	
Single-layer 	Half Plane (Bounded by hyperplane)				(a)
Two-layer 	Convex (Open or closed regions)				(b)
Three-layer 	Arbitrary (Complexity limited by number of neurons)				(c)

Redes Recorrentes

- Possuem pelo menos um laço realimentando a saída de neurônios para outros neurônios da rede.

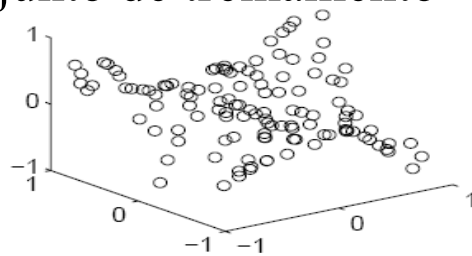


Redes Neurais Artificiais

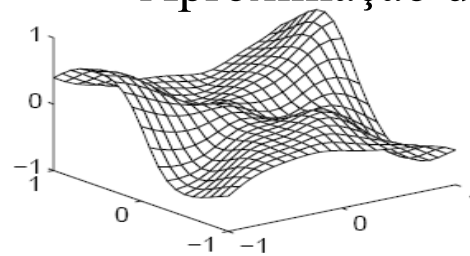
- Caracterizadas por 3 elementos:
 - Conjunto de neurônios artificiais
 - Um padrão de conexão entre os neurônios
 - Arquitetura ou estrutura da rede
 - Método que determina o valor dos pesos
 - Algoritmo de treinamento ou aprendizagem

O que uma rede faz?

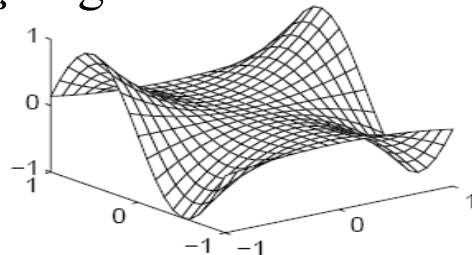
Conjunto de treinamento



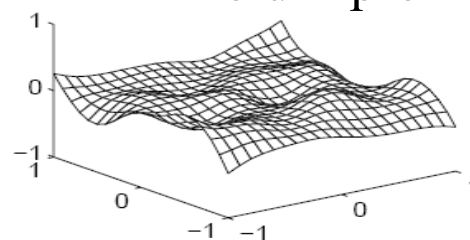
Aproximação da rede



Função geradora dos dados



Erro da aproximação



Aprendizado

- **Aprendizado** (treinamento da rede) corresponde ao processo de ajuste dos parâmetros livres da rede através de padrões (ou dados) de entrada ou de treinamento

Aprendizado

- Seja $w(t)$ um peso sináptico de um dado neurônio, no instante de tempo t .
 - O ajuste $\Delta w(t)$ é aplicado ao peso sináptico $w(t)$ no instante t , gerando o valor corrigido $w(t+1)$, da forma:

$$w(t+1) = w(t) + \Delta w(t)$$

- Várias maneira de obter $\Delta w(t)$:
 - regra de Hebb, regra Delta, algoritmo de *backpropagation*, estratégias de competição, máquina de Boltzmann

Leitura Recomendada

- A Brief Introduction to Neural Networks,
http://www.dkriesel.com/en/science/neural_networks, Parte 1