



Introdução Inteligência Computacional

Profa. Dra. Ana Paula Abrantes de Castro e Shiguemori

anapaula.acs@ifsp.edu.br

Instituto Federal de Educação, Ciência e Tecnologia – IFSP

Dr. Elcio Hideiti Shiguemori

elcio@ieav.cta.br

Instituto de Estudos Avançados - IEAv

Curso: Análise e Desenvolvimento de Sistemas
4º. Semestre - IIC14 – 4 aulas Semanais

Representação do Conhecimento

Representação do Conhecimento

Conhecimento se refere à informação ou modelos armazenados utilizados por uma pessoa ou uma máquina para interpretar, prever e responder adequadamente ao ambiente.

► Características da representação do conhecimento:

1. Que informação representar
2. Como codificar fisicamente a informação
3. Uma boa solução de uma boa representação do conhecimento

Tarefas das Redes Neurais

- RNA deve aprender um modelo consistente do mundo real (ambiente) no qual ela está embutida
- Manter o modelo suficientemente consistente com o mundo real a fim de atender os objetivos especificados pela aplicação de interesse

Tipos de Conhecimento

1. **Fatos conhecidos** – representado pelos fatos sobre o que é e o que era conhecido (*informação prévia*)
2. **Observações do mundo** – conjunto de informações de onde são retirados os *exemplos* utilizados para treinar a rede neural (*obtidas por sensores, em geral ruidosas*)

Aprendizagem

Processos de Aprendizagem

- ▶ Habilidade importante de uma Rede Neural → Capacidade de aprender a partir de seu próprio ambiente
- ▶ A aprendizagem melhora o desempenho da RNA
- ▶ RNA conhece melhor o ambiente após cada iteração na aprendizagem

Processos de Aprendizagem

“Aprendizagem é um processo no qual os parâmetros livres (pesos) de uma rede neural são adaptados através de um processo contínuo de simulação pelo ambiente no qual a rede está embutida. O tipo do aprendizado é determinado pela maneira pela qual a modificação dos parâmetros ocorre”

(Mendel e McLaren, 1970)

Processos de Aprendizagem

Seqüência de eventos para aprendizagem:

- RNA é estimulada por um ambiente
- RNA sofre modificações nos seus parâmetros livres como resultado da estimulação
- RNA responde de uma maneira nova ao ambiente por causa das modificações ocorridas na sua estrutura interna

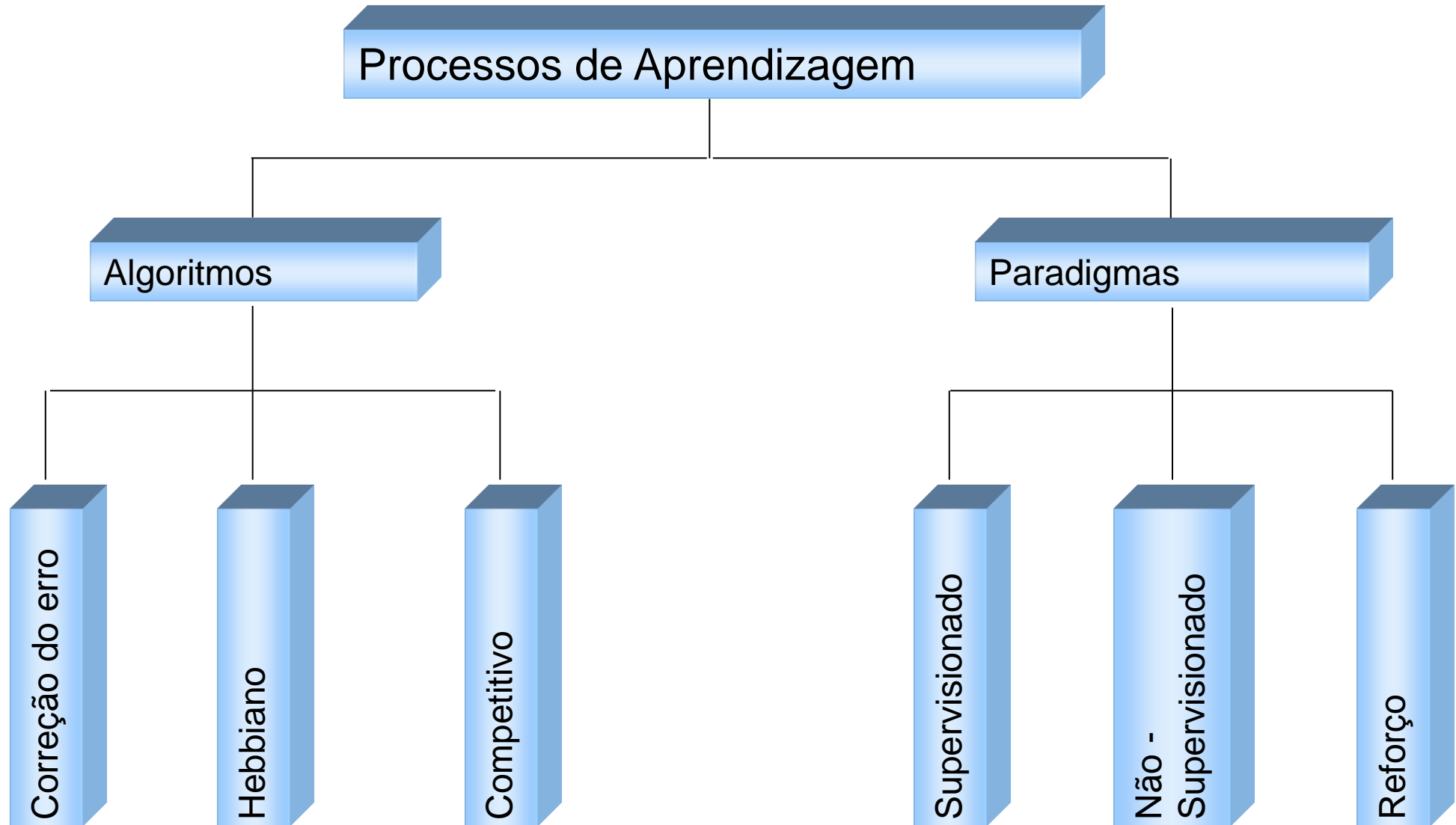
Algoritmo de Aprendizagem

- Conjunto de regras bem definidas para a solução de um problema de aprendizagem, ou seja, um conjunto de regras para a atualização dos pesos.
- Não existe um algoritmo único

Algoritmos diferem:

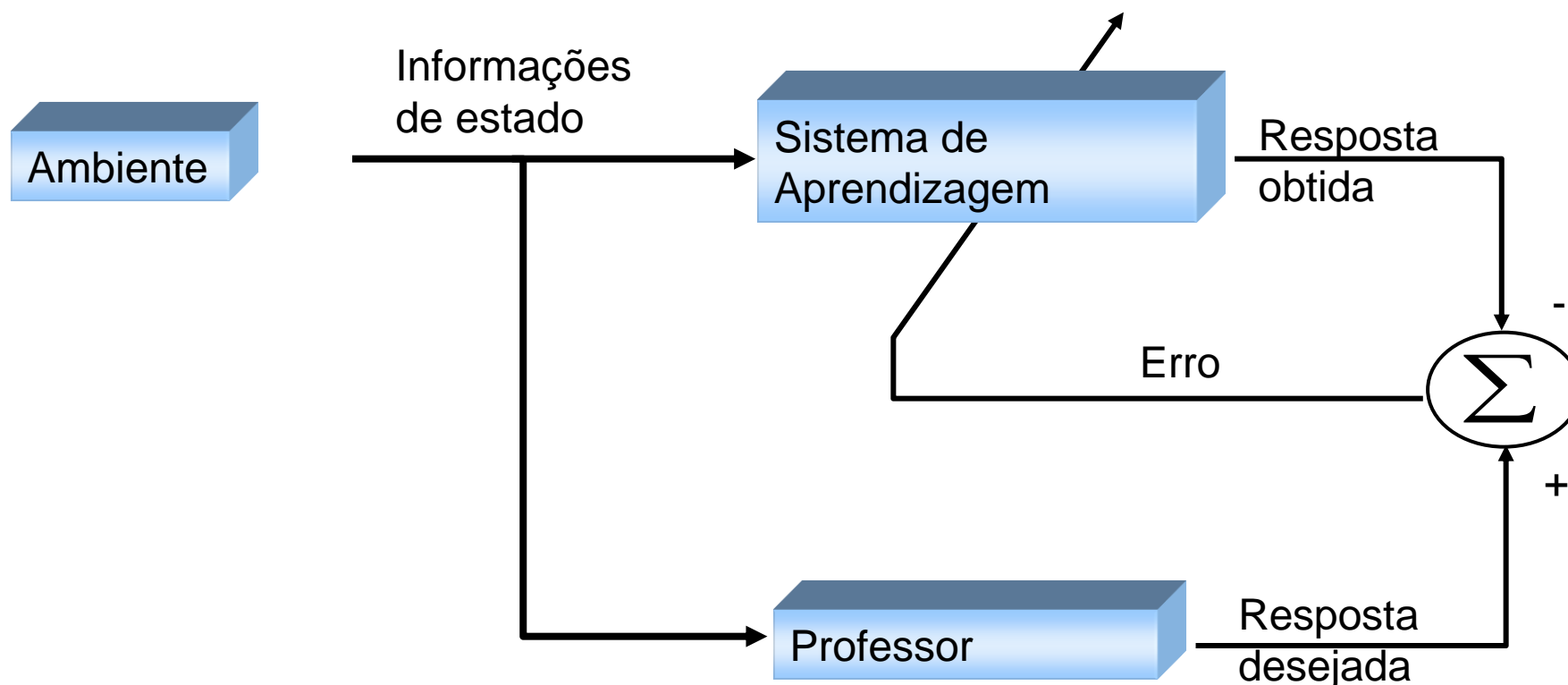
- Diferem na maneira como os pesos são modificados, ou seja na forma de calcular o
- Na maneira de se relacionar com o ambiente

Classificação do processo de aprendizagem



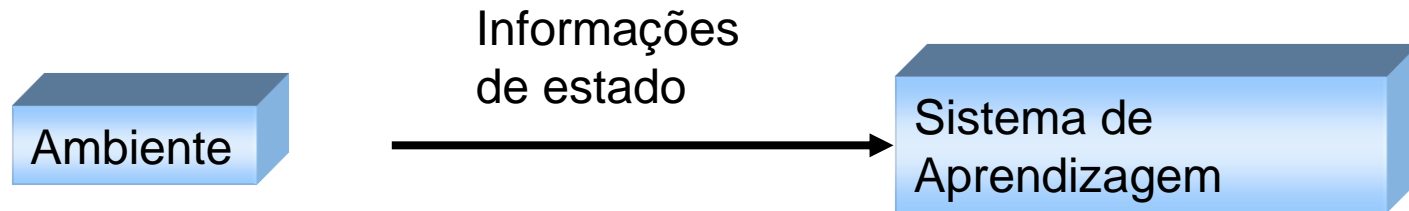
Aprendizagem Supervisionada

- O supervisor tem conhecimento sobre o ambiente, sendo o conhecimento representado por um conjunto de exemplos de entrada-saída



(Demisio, 2005)

Aprendizagem não-supervisionada



- ▶ Não há exemplos específicos
- ▶ Usa medida independente da tarefa para indicar a qualidade da representação que a rede deve aprender; os pesos são otimizados de acordo com essa medida.
- ▶ Aprendizagem pára quando a rede está sintonizada com as regularidades estatísticas dos dados de entrada.

Aprendizagem por reforço

- A única informação de realimentação fornecida para a rede é se uma determinada saída está correta ou não, isto é, não é fornecida para a rede a resposta correta para o padrão de entrada.
- Aprendizado on-line obtido por um mapeamento de entrada-saída é realizado através de um processo de triagem e erro desenvolvido para maximizar o índice de desempenho escalar chamado sinal de reforço

Objetivo: Minimizar uma função custo para avançar, definida como a expectativa do custo cumulativo de ações tomadas ao longo de uma seqüência de passos, em vez simplesmente do custo imediato.

Aprendizagem por reforço

► Termo “reforço” teve a sua origem em estudos experimentais sobre aprendizado dos animais, lembrando o contexto da *Lei de Efeito* de Thorndike

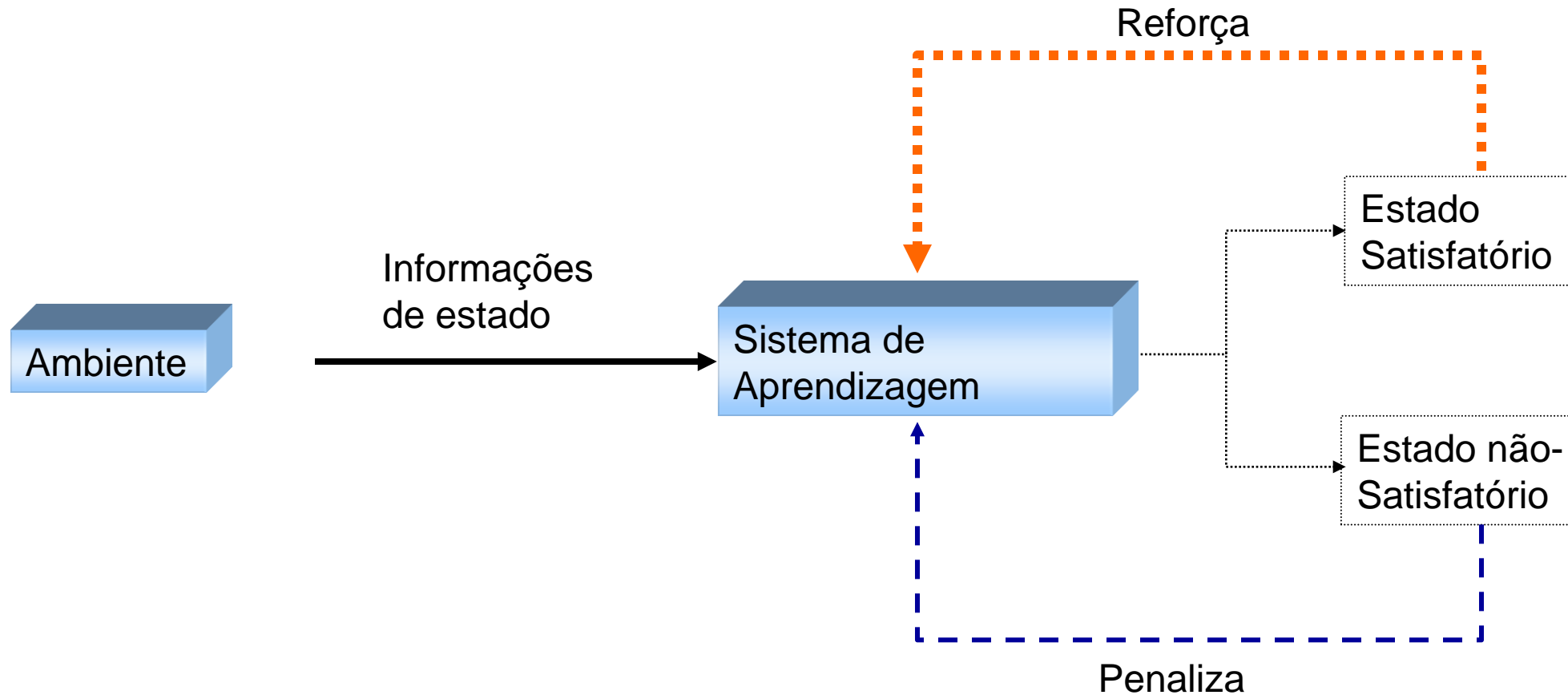
“Quanto maior a satisfação obtida com uma certa experiência em um animal, maiores as chances dele aprender”

(Thorndike, 1911)

“Se uma ação tomada pelo sistema de aprendizagem é seguida de estados satisfatórios, então a tendência do sistema de produzir esta ação particular é reforçada. Se não for seguida de estados satisfatórios, a tendência do sistema de produzir esta ação é enfraquecida”

(Sutton, 1991)

Aprendizagem por reforço



Aprendizagem por reforço

Paradigma pode ser do tipo:

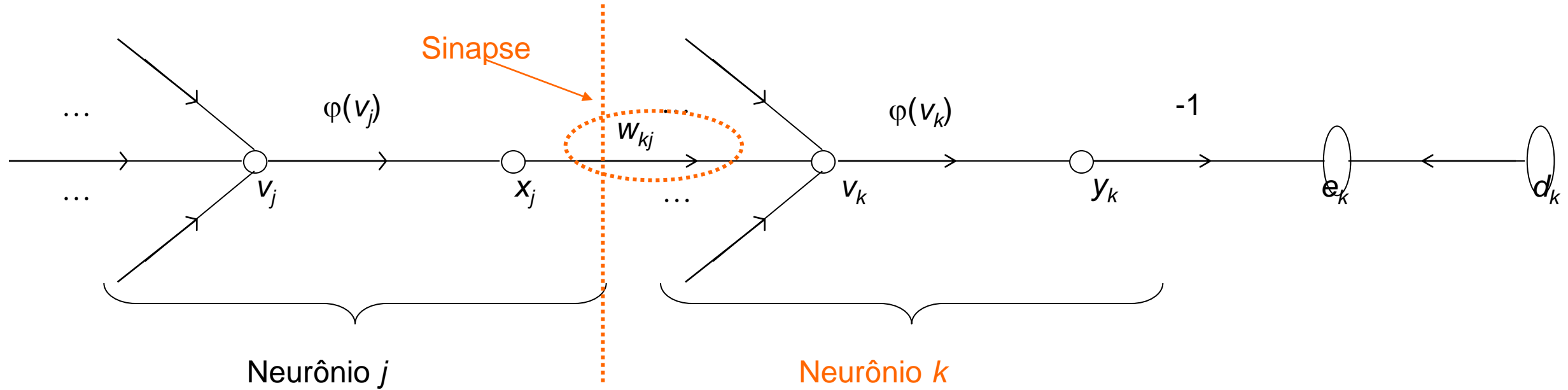
Não-associativo

- Tarefa de aprendizagem é selecionar uma ação ótima
- A única informação recebida pelo sistema de aprendizagem é o reforço

Associativo

- Tarefa de aprendizagem é aprender um mapeamento na forma da associação *estímulo-ação*

Aprendizagem por correção do erro



$$e_k(n) = d_k(n) - y_k(n)$$

Aprendizagem por correção do erro

- Minimiza uma função de custo baseada no sinal de erro $e_k(n)$ (Critério mais usado, é o erro quadrático médio)

$$J = E \left[\frac{1}{2} \sum_k e_k^2(n) \right]$$

- Para uma função de custo escolhida => problema de otimização
→ A minimização de J pelo método do gradiente descendente

Aprendizagem por correção do erro

- Problema: requer conhecimento das características estatísticas do processo
- Solução: buscar por uma solução aproximada
- Critério de interesse: **valor instantâneo da soma dos erros quadráticos**

$$\varepsilon(n) = \frac{1}{2} \sum_k e_k^2(n)$$

→ otimização pela minimização de $\varepsilon(n)$ em função dos pesos

Aprendizagem por correção do erro

$$\varepsilon(n) = \frac{1}{2} \sum_k e_k^2(n) \quad \text{Método do gradiente descendente}$$

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n) \qquad \Delta w_{kj}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial w_{kj}(n)}$$

$$\Delta w_{kj}(n) = -\eta \left(d_k(n) - \sum_{j=0}^p w_{kj}(n) x_j(n) \right) (-x_j(n))$$

$$\Delta w_{kj}(n) = \eta \left(d_k(n) - \sum_{j=0}^p w_{kj}(n) x_j(n) \right) x_j(n)$$

$$w_{kj}(n+1) = w_{kj}(n) + \eta \left(d_k(n) - \sum_{j=0}^p w_{kj}(n) x_j(n) \right) x_j(n)$$

Aprendizagem por correção do erro

- ▶ A regra de aprendizagem por correção do erro (ou Regra Delta):

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n) , \quad \eta > 0 \rightarrow \text{taxa de aprendizagem}$$

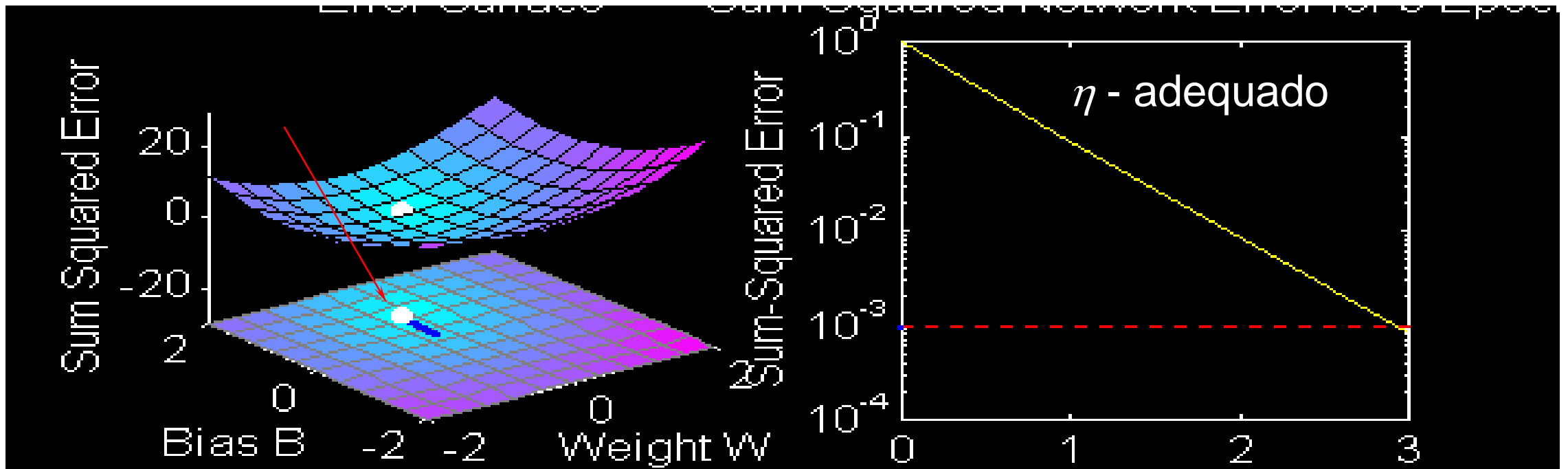
- ▶ Aprendizagem por correção do erro → um sistema realimentado fechado
- ▶ η deve ser escolhido para garantir estabilidade no processo

$\eta \rightarrow$ muito pequeno → convergência demorada

$\eta \rightarrow$ muito grande → aprendizagem acelerada, mas pode divergir

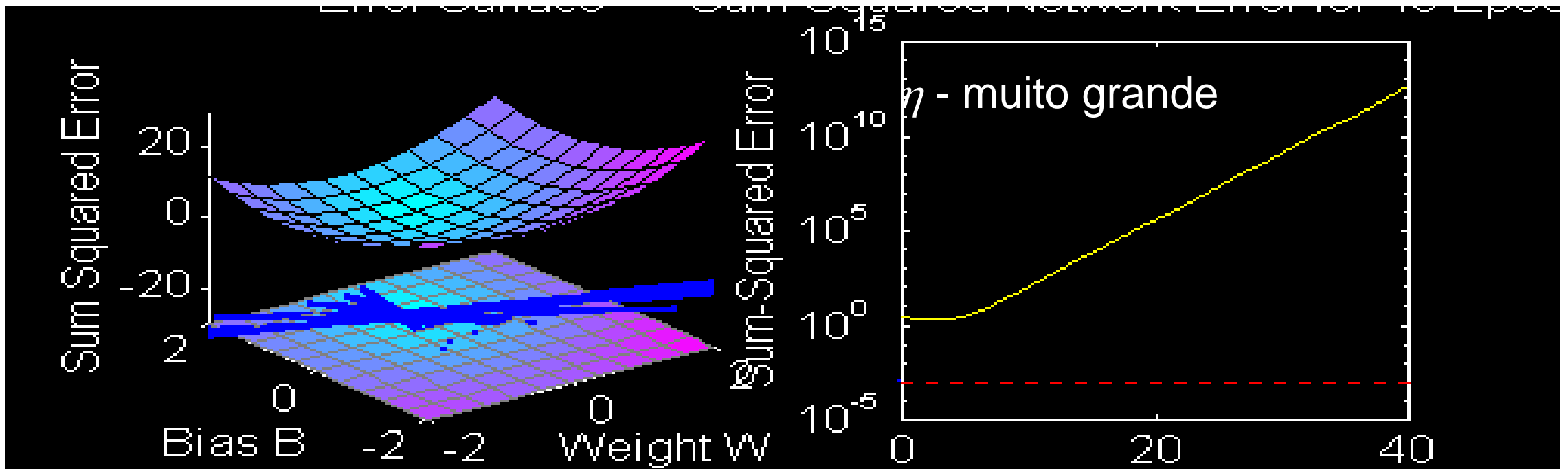
Aprendizagem por correção do erro

Superfície do erro



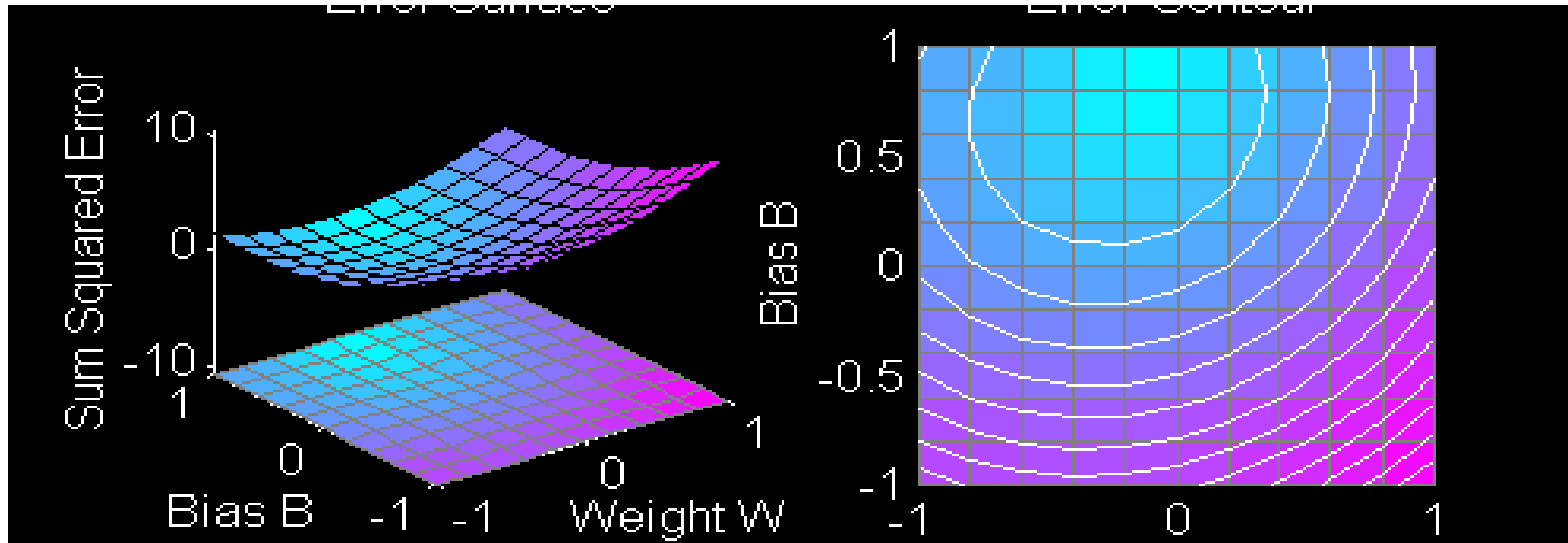
Aprendizagem por correção do erro

Superfície do erro



Aprendizagem por correção do erro

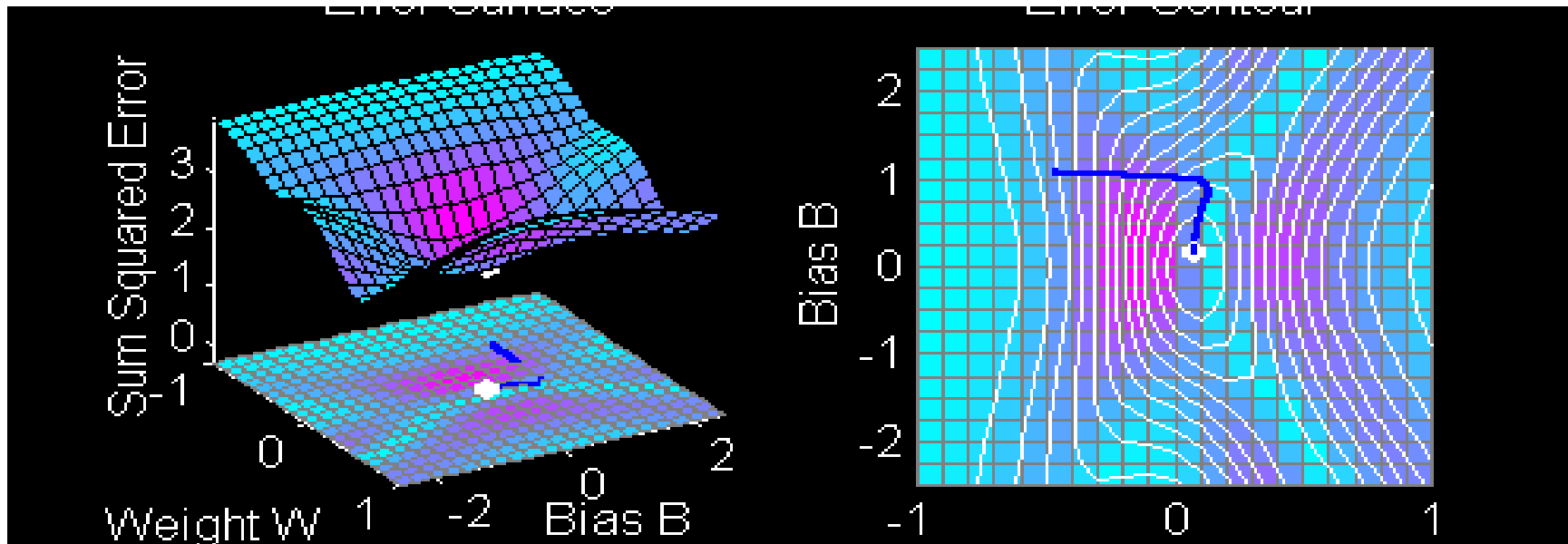
- ▶ RNA composta por neurônios lineares → superfície de erro é uma função quadrática dos pesos com um ponto de mínimo → objetivo atingido



- ▶ O algoritmo inicia em um ponto arbitrário da superfície (pesos iniciais) e move-se para um mínimo global, passo a passo.

Aprendizagem por correção do erro

- ▶ RNA composta por neurônios não lineares → superfície de erro possui um mínimo global e vários mínimos locais → nem sempre atinge o objetivo, pode cair em um mínimo local.



- ▶ O algoritmo inicia em um ponto arbitrário da superfície (pesos iniciais) e move-se para um mínimo global, passo a passo.

Aprendizagem Hebbiana

- ▶ Regra mais antiga e famosa; Donald Hebb (1949) que postulou a base para a aprendizagem associativo:

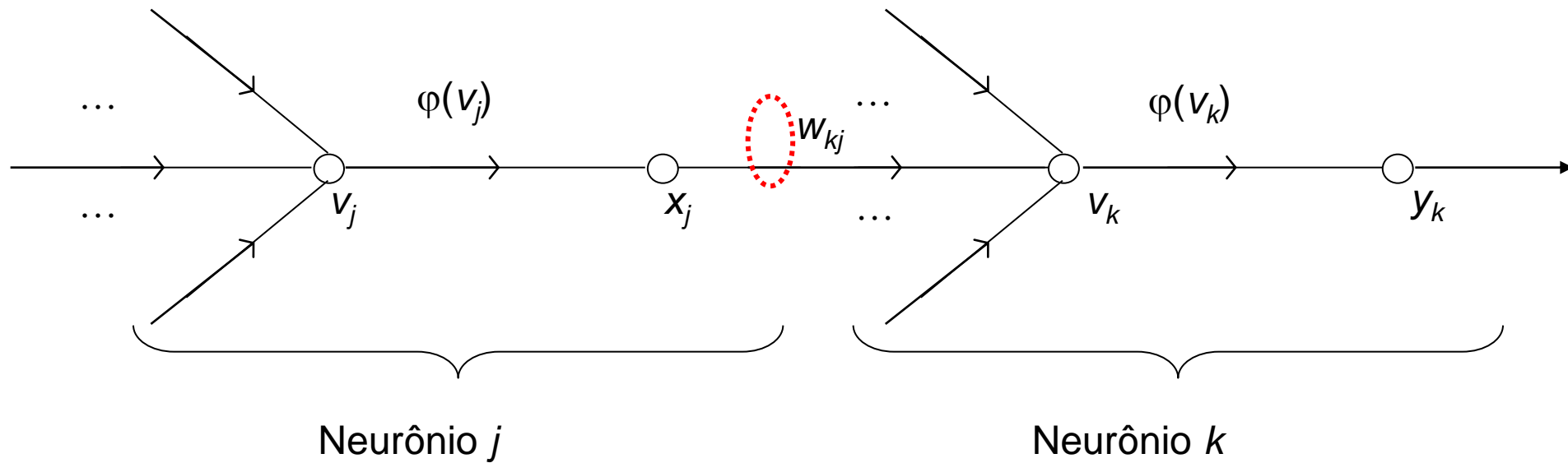
“Quando o axônio de uma célula A está perto o suficiente para excitar uma célula B e repetida ou persistentemente participa de sua ativação, algum processo de crescimento ou mudança metabólica acontece em uma ou em ambas as células tal que, a eficiência de A como uma das células que disparam B, é aumentada”

Aprendizagem Hebbiana

No contexto de RNA (Haykin, 1994) :

- Se dois neurônios em ambos os lados de uma sinapse (conexão) são ativados simultaneamente (i.e., em sincronismo) então a sinapse deve ser fortalecida
- Se dois neurônios em ambos os lados de uma sinapse são ativados assincronamente, então a sinapse deve ser enfraquecida ou eliminada

Aprendizagem Hebbiana



$$\Delta w_{kj}(n) = F(y_k(n), x_j(n))$$

$F(*, *)$ → função das atividades pré e pós-sinápticas

$y_k(n), x_j(n)$ → variáveis adimensionais

Aprendizagem Hebbiana

Hipótese de Hebb

- A forma mais simples de aprendizagem Hebbiana, referida como a regra do produto ativo:

$$\Delta w_{kj}(n) = \eta y_k(n) x_j(n)$$

Enfatiza a correlação inerente de uma sinapse Hebbiana

Aprendizagem Competitiva

- ▶ Neurônios da camada de saída da RNA *competem entre si para ser o neurônio ativado* (um único neurônio é ativado)
- ▶ Aprendizado competitivo é adequado para descobrir características estatísticas salientes, que podem ser usadas para classificar um conjunto de dados de entrada.

Aprendizagem Competitiva

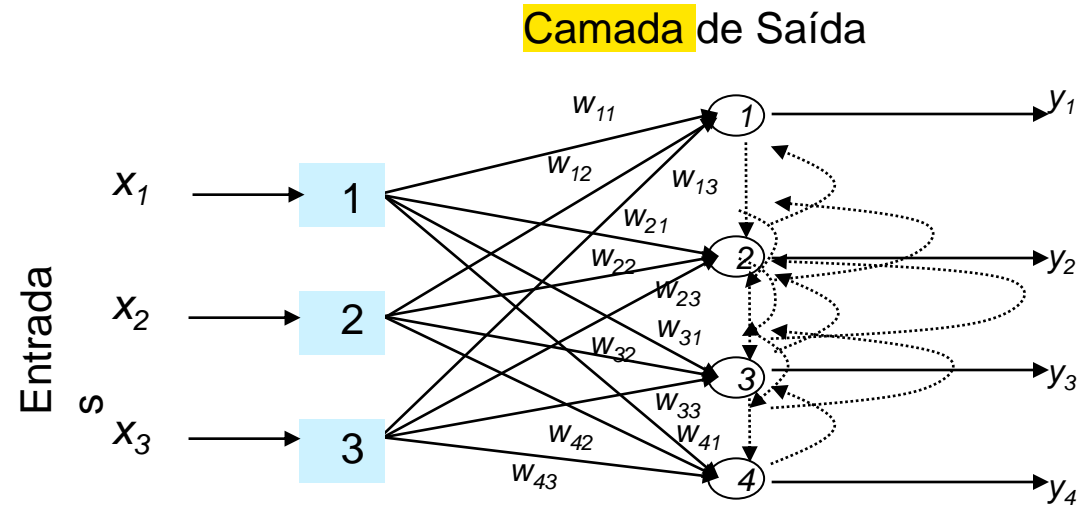
► Neurônios Elementos básicos para uma regra de aprendizado competitivo:

- Um conjunto de neurônios do mesmo tipo exceto por alguns pesos sinápticos distribuídos aleatoriamente, respondem diferentemente a um dado conjunto de padrões de entrada.
- Um limite imposto à robustez de cada neurônio.
- Um mecanismo que permite os neurônios competirem pelo direito de responder à uma certa entrada (subconjunto). O neurônio ganhador é o neurônio que leva tudo (winner-takes-all).

Aprendizagem Competitiva

- Os neurônios individuais aprendem e se especializam nos conjuntos de padrões similares → tornam-se **detetores de características**
- Forma mais simples de aprendizagem competitiva consiste de uma RNA de uma camada única com todos os neurônios totalmente conectados às entradas

Aprendizagem Competitiva



- O neurônio j é o vencedor se possuir o maior nível de atividade interna v_j para um vetor de entrada. A saída y_j do neurônio vencedor é feita igual a 1. As saídas dos outros neurônios são zero.

Aprendizagem Competitiva

- ▶ O peso entre o neurônio j e a entrada i .

$$w_{ji} > 0$$

- ▶ Alocação de pesos em quantidades fixas para cada neurônio:

$$\sum_i w_{ji} = 1 \quad \forall j$$

Aprendizagem Competitiva

- ▶ Aprendizagem é o deslocamento dos pesos sinápticos dos nós inativos para os ativos.
- ▶ Cada peso do neurônio vencedor de cada entrada é adaptado pela regra de aprendizagem abaixo competitiva padrão:

$$\Delta w_{ji} = \begin{cases} \eta(x_i - w_{ji}), & \text{se } j \text{ ganha} \\ 0, & \text{se } j \text{ perde} \end{cases}$$

- ▶ A regra move o vetor de pesos sinápticos w , do neurônio j , na direção do padrão de entrada.

Tarefas de Aprendizagem

Aproximação de função - aprendizagem supervisionada

Associação

- Autoassociação - aprendizagem não-supervisionada
- Heteroassociação - aprendizagem supervisionada

Classificação de Padrões

- Padrões conhecidos - aprendizagem supervisionada
- Padrões desconhecidos - aprendizagem não-supervisionada (clusterização)

Predição (problema de processamento de sinais temporais) - aprendizagem supervisionada

Controle - aprendizagem supervisionada

Bibliografia Utilizada

- Notas de aula do curso de Redes Neurais Artificiais do INPE, Prof. Dr. Demisio Simões da Silva, 2001 e 2005.
-
- Haykin, S. Redes Neurais: Princípios e Prática, New York: Mcmillan, 2001.
- Braga A. P.; Carvalho A. P. L. F.; Ludermir T. B. Redes Neurais Artificiais Teorias e aplicações, Rio de Janeiro: LTC, 2000.
- Azevedo, F.; Brasil, L.M., Oliveira, R.C.L., Redes Neurais com Aplicações em Controle e em Sistemas Especialistas. Visual Books, 2000.



Dúvidas?



Referências Bibliográficas

- José Demísio Simões da Silva – Notas de Aula
- Ana Paula A. C. Shiguemori – Notas de Aula
- Elcio Hideiti Shiguemori – Notas de Aula