

LECTURE 02 - LEARNING

NEURAL

LEARNING

1. Graph representation of Neural Networks

2. What is Learning in Neural Networks

3. Learning Techniques

1. Error-correction learning; Memory-based learning; Hebbian learning; Comp
Boltzmann learning

4. Supervised learning

5. Unsupervised learning

6. semi-supervised learning

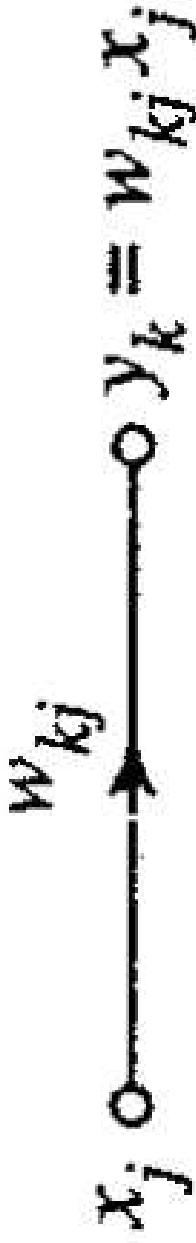
GRAPH-BASED REPRESENTATION

- Fornece uma descrição funcional dos vários que constituem o modelo de um neurônio a
- Vantagens:
 - Simplificação da representação
 - Pode representar sinais inter e intra neurônios

GRAPH-BASED REPRESENTATION

- O fluxo de sinal nas diversas partes do grafo é ditado por regras:
 - Regra 1: Um sinal flui ao longo de um elo somente se definido pela seta

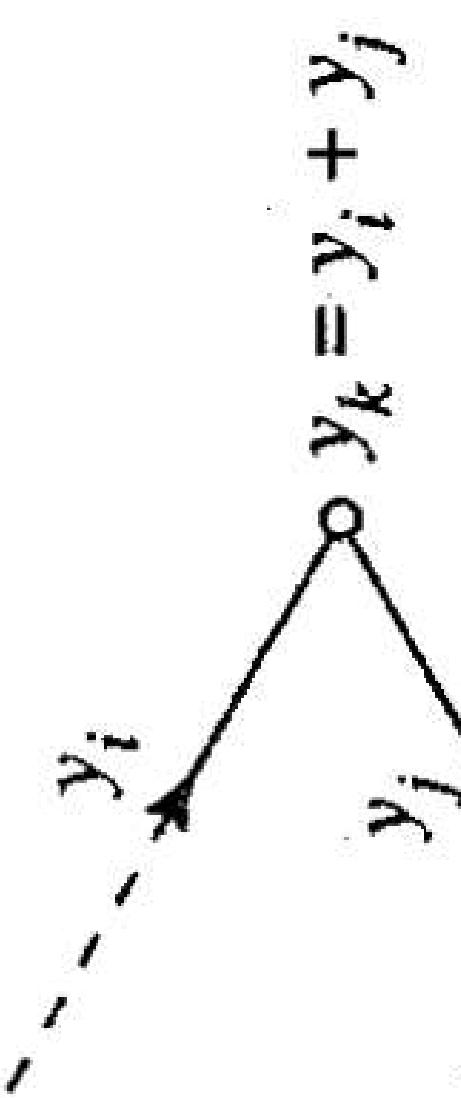
a) Elos sinápticos: relação entrada-saída linear



b) Elos de ativação: relação entre entrada-saída não linear

GRAPH-BASED REPRESENTATION

1. Regra 2: Um sinal nodal é igual à soma dos sinais que entram no nó via os elos in-



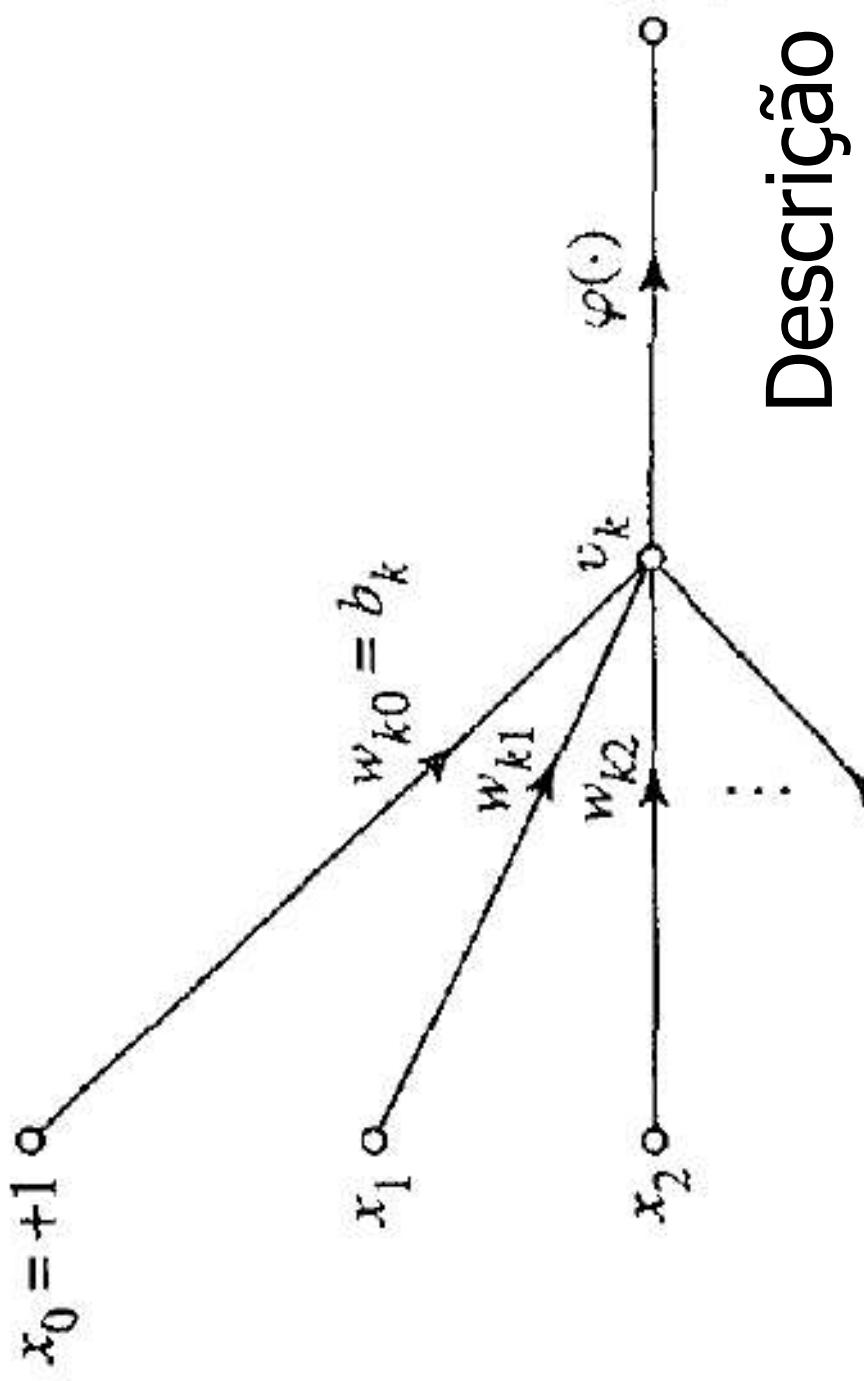
GRAPH-BASED REPRESENTATION

1. Regra 3: O sinal de um nó é transmitido pelo elo de saída originário deste nó, sendo transmissão independente das funções transferência dos elos de saída



GRAPH-BASED REPRESENTATION

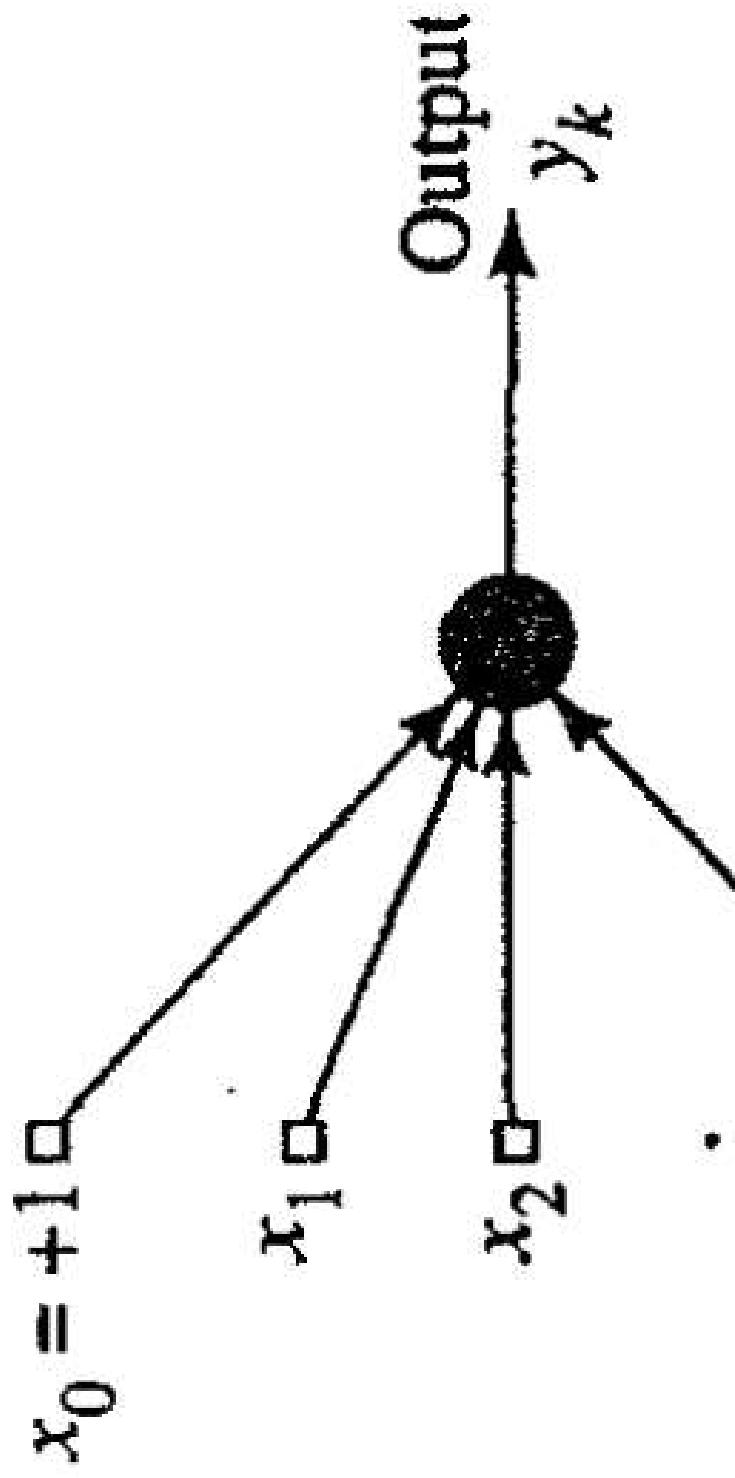
1. Exemplo de grafo de fluxo de um neurônio



Descrição

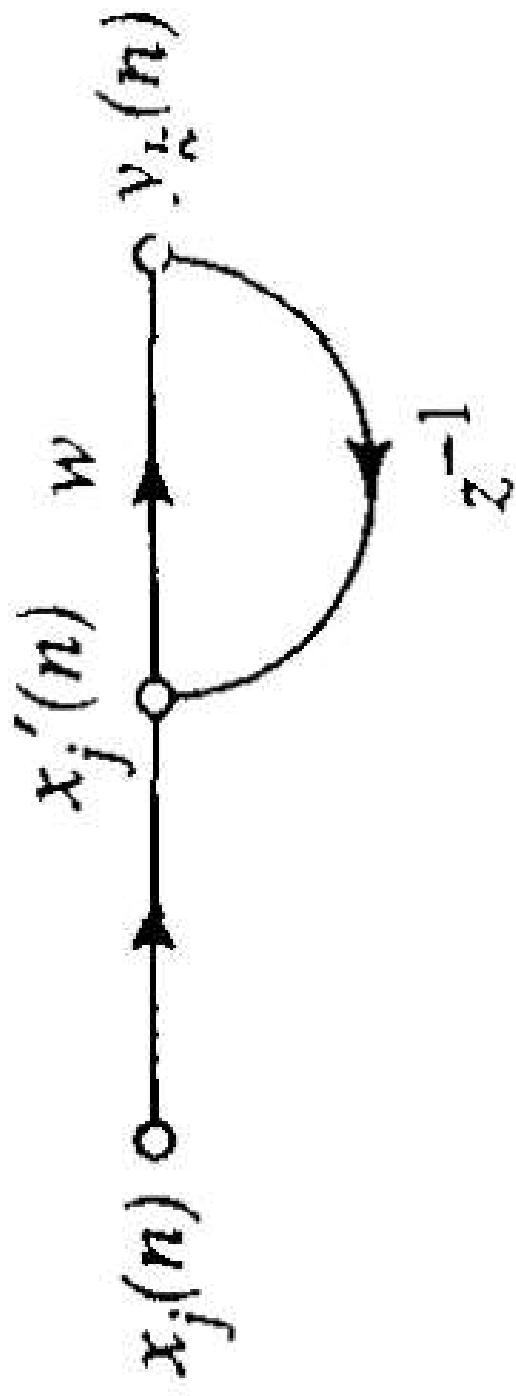
GRAPH-BASED REPRESENTATION

1. Exemplo de grafo arquitetural



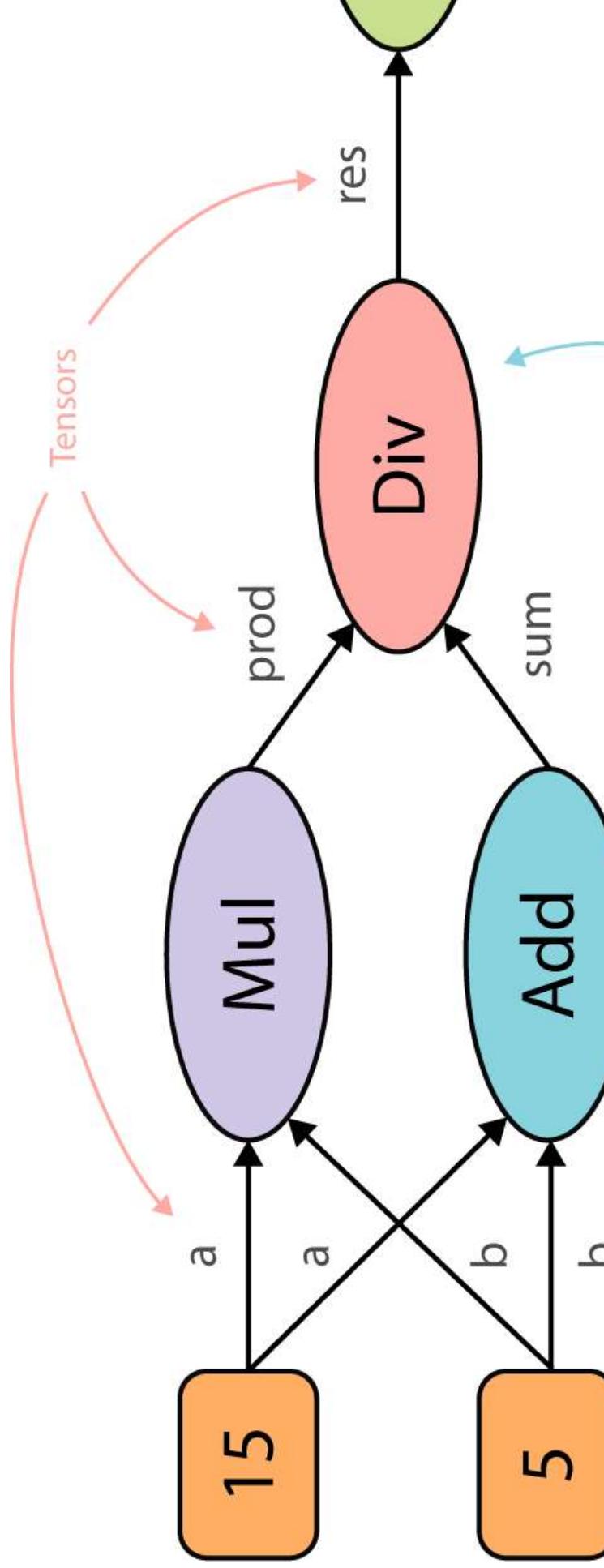
GRAPH-BASED REPRESENTATION

1. Reimplementação



GRAPH-BASED REPRESENTATION

1. TensorFlow (Graph)



TRAINING A NEURAL NETWORK

FEARLESSLY

APRENDIZAGEM

1. Aprendizado é o processo pelo qual os **parâmetros das conexões entre os neurônios** e **hiperparâmetros (taxas, topologia, etc.)** de um neural são ajustados através de uma forma de estímulo pelo ambiente no qual a rede está operando.
2. Aprendizado também é chamado como treinamento.

APRENDIZAGEM

1.O processo de aprendizado implica a seguinte sequência de evolução:

- 1.** A rede neural é estimulado por um ambiente.
- 2.** A rede neural sofre modificações nos seus parâmetros livres como resultado do estímulo.
- 3.** A rede neural responde de uma maneira nova ao ambiente, devido as mudanças ocorridas na sua estrutura interna.

2.Um conjunto de regras bem-definidas para a solução de um problema de aprendizagem é denominado um **algoritmo de aprendizagem** (de treinamento)

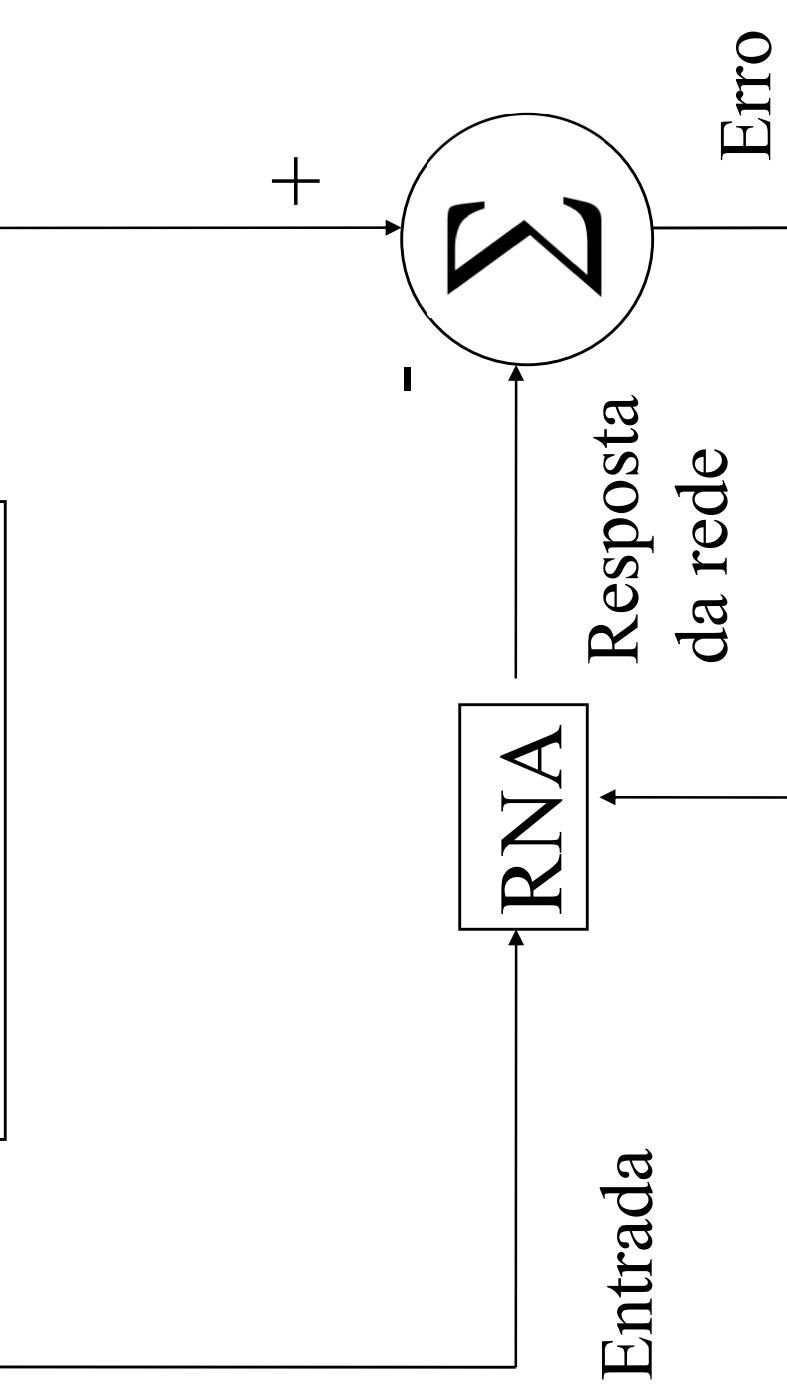
REGRAS DE APRENDIZAGEM

1. Error-correction learning
2. Memory-based learning
3. Hebbian learning
4. Competitive learning
5. Boltzmann learning

ERROR-CORRECTION LEARNING

Resposta
desejada

Professor



ERROR-CORRECTION LEARNING

1. Processo de minimização do erro quadrático p
realizado pelo método do Gradiente Descendente
2. Cada peso sináptico i do elemento processador
atualizado proporcionalmente ao negativo da c
parcial do erro deste processador com relação

ERROR-CORRECTION LEARNING

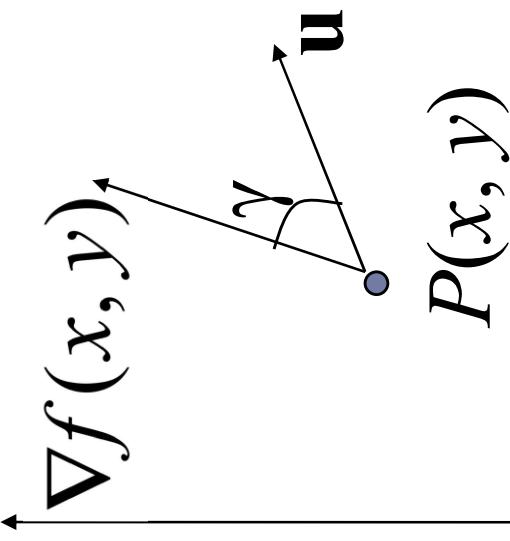
Gradiente:

$$\nabla f(x, y) = \left(\frac{\partial}{\partial x} f(x, y), \frac{\partial}{\partial y} f(x, y) \right)$$

Derivada direcional:

$$D_{\mathbf{u}} f(x, y) = \nabla f(x, y) \cdot$$

$$= \left\| \nabla f(x, y) \right\|$$
$$= \left\| \nabla f(x, y) \right\|$$



$D_{\mathbf{u}} f(x, y)$ é a taxa de varia-

ERROR-CORRECTION LEARNING

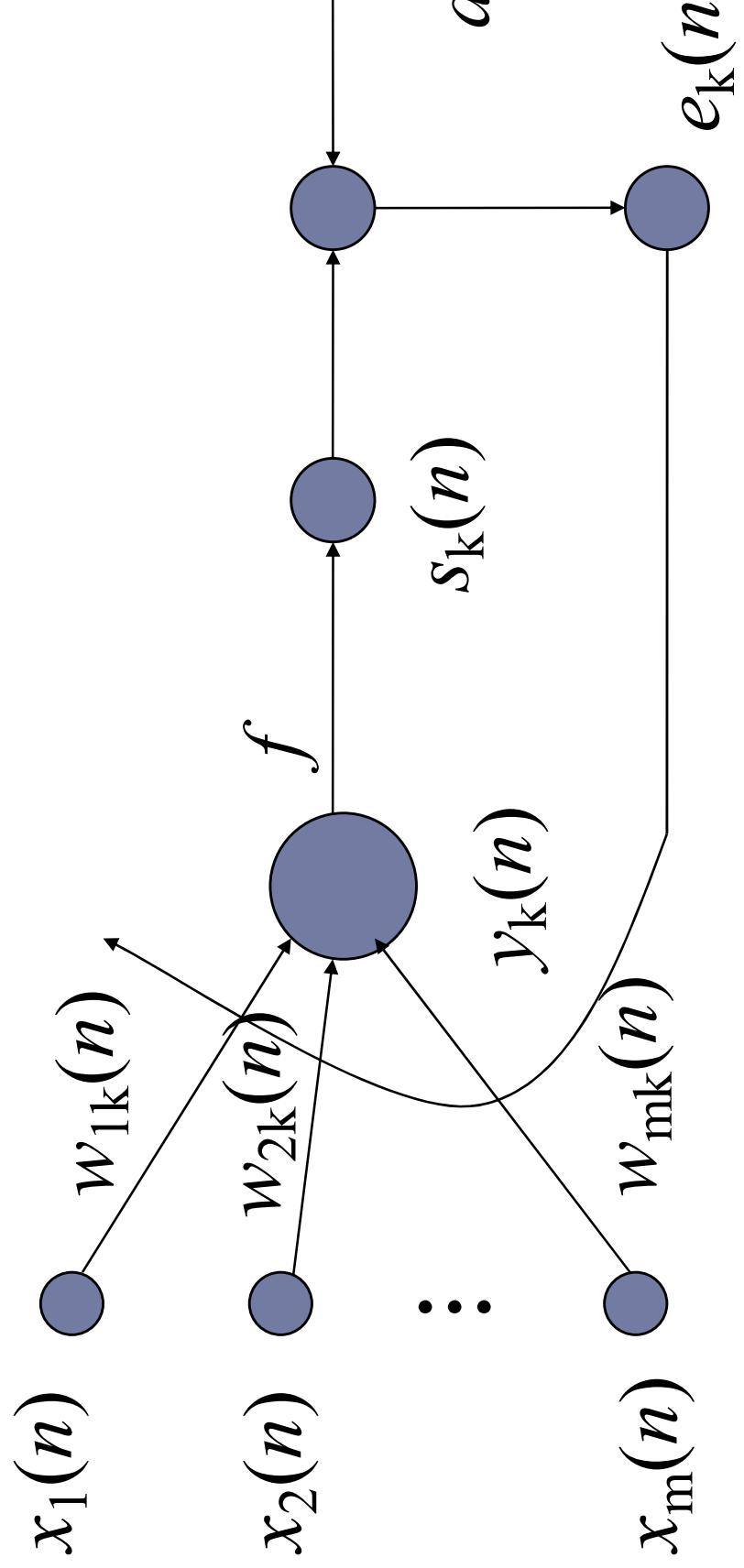
1. Teorema do gradiente: Seja f uma função de duas variáveis, diferenteável no ponto $P(x, y)$.

- 1.** i) O máximo de $D_u f(x, y)$ em $P(x, y)$ é $\|\nabla f(x, y)\|$.
- 2.** ii) O máximo da taxa de crescimento de $f(x, y)$ em $P(x, y)$ ocorre em $\nabla f(x, y)$.

2. Corolário: Seja f uma função de duas variáveis, diferenciáveis no ponto $P(x, y)$.

- 1.** i) O mínimo de $D_u f(x, y)$ em $P(x, y)$ é $-\|\nabla f(x, y)\|$.
- 2.** ii) O máximo da taxa de decrescimento de $f(x, y)$ em $P(x, y)$ ocorre em $-\nabla f(x, y)$.

ERROR-CORRECTION LEARNING



ERROR-CORRECTION LEARNING

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$$

$$E_p = \frac{1}{2} \sum_j (d_j - s_j)^2$$

ERROR-CORRECTION LEARNING

Cálculo do Δw_{ij}

$$\Delta w_{ij} = -\eta \frac{\partial E_p}{\partial w_{ij}} = -\eta$$

$$\begin{array}{|c|c|}\hline \frac{\partial E_p}{\partial s_j} & \frac{\partial s_j}{\partial w_{ij}} \\ \hline \end{array}$$

$$E_p = \frac{1}{2} \sum_j (d_j - s_j)^2$$

$$2 \times \frac{1}{2} \times (d_j - s_j) (-1)$$

$$s_j = \sum x_i$$

ERROR-CORRECTION LEARNING

1. Logo, a correção dos pesos via gradiente descendente reduz o erro total do sistema

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$$

MEMORY-BASED LEARNING

- Nesta forma de aprendizagem, todos ou a maioria dos exemplos armazenados explicitamente em uma grande memória
- Para se classificar um novo exemplo não presente no conjunto de treinamento, i.e.
 \mathbf{X}_{teste}

MEMORY-BASED LEARNING

1. Algoritmos baseados em memória envolvem dois principais:

1. O critério utilizado para definir a vizinhança local do vetor de teste
2. A regra de aprendizagem aplicada aos exemplos de treinamento vizinhança local de \mathbf{x}_{teste} ,

2. Exemplo: A regra do vizinho mais próximo (*k*-NN)

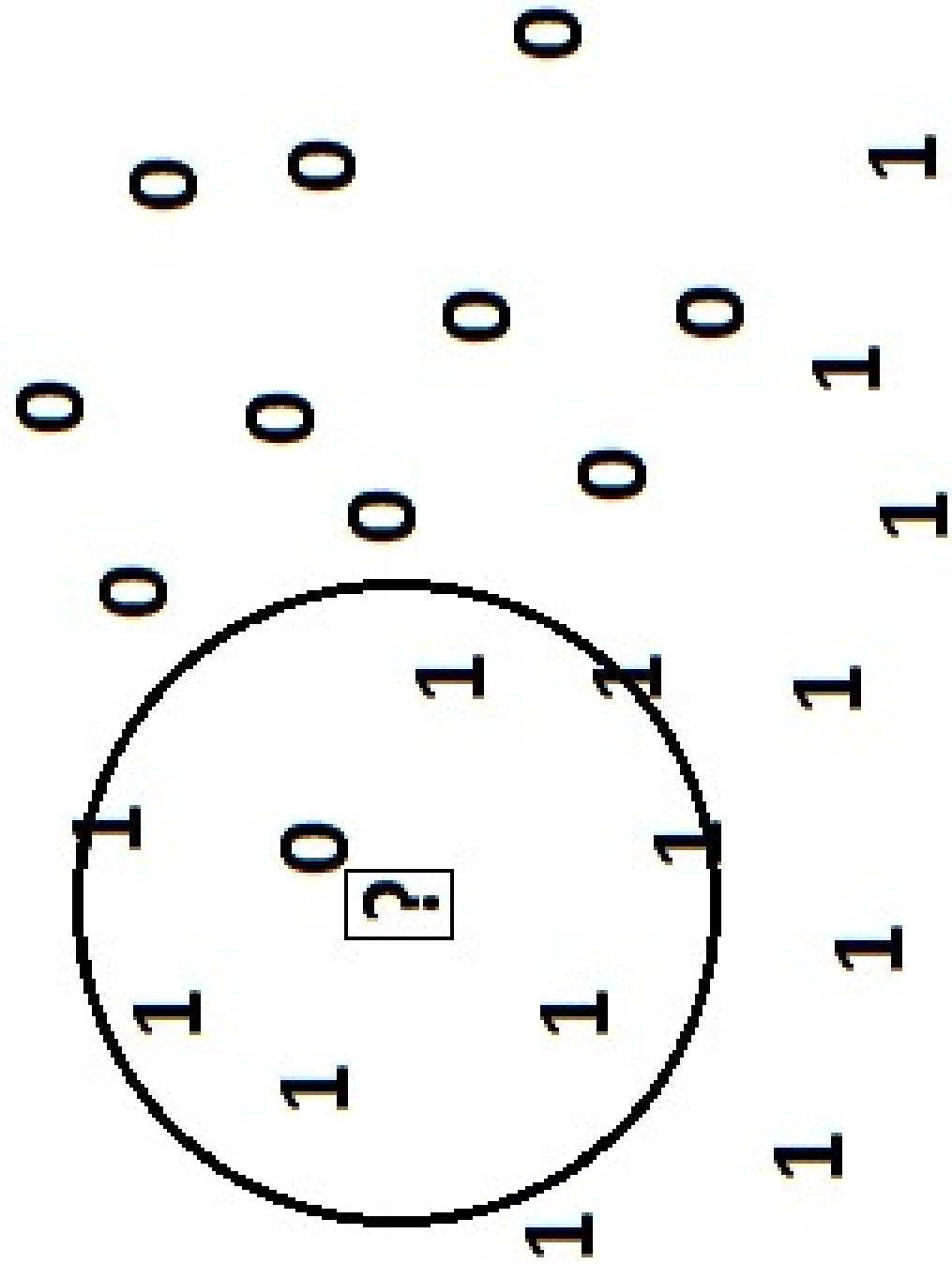
1. Busca-se pelos (*k*) exemplos do conjunto de treinamento \mathbf{X}_{teste}

MEMORY-BASED LEARNING

1. Onde $d(i,j)$ representa uma medida de distância quase sempre a distância Euclidiana
2. Identificado os k vizinhos mais próximos, verifica-se a classe predominante entre tais exemplos selecionados
3. Essa classe é associada ao exemplo \mathbf{x}_{teste} ,
4. O classificador k -NN atua como um dispositivo que “média” entre os vizinhos do exemplo a ser classificado

MEMORY-BASED LEARNING

MEMORY-BASED LEARNING



HEBBIAN LEARNING

1. Primeiro estudo sobre redes de neurônios
Postulado de Hebb – The Organization of Behavior

- 1.** Se dois neurônios em ambos os lados de uma sinapse são ativados simultaneamente (sincronamente), então a força daquela sinapse é seletivamente aumentada;

- 2.** Se dois neurônios em ambos os lados de uma sinapse ativados assincronamente, então a força daquela sinapse é seletivamente enfraquecida ou eliminada

HEBBIAN LEARNING

1. Propriedades da sinapse Hebbiana:

1. Mecanismo dependente do tempo: dependem exato de ocorrência dos sinais pré e pós-sinápticos

2. Mecanismo local: as modificações são locais (sinapses)

3. Mecanismo interativo: depende explicitamente entre os neurônios pré e pós-sinápticos de uma sinapse

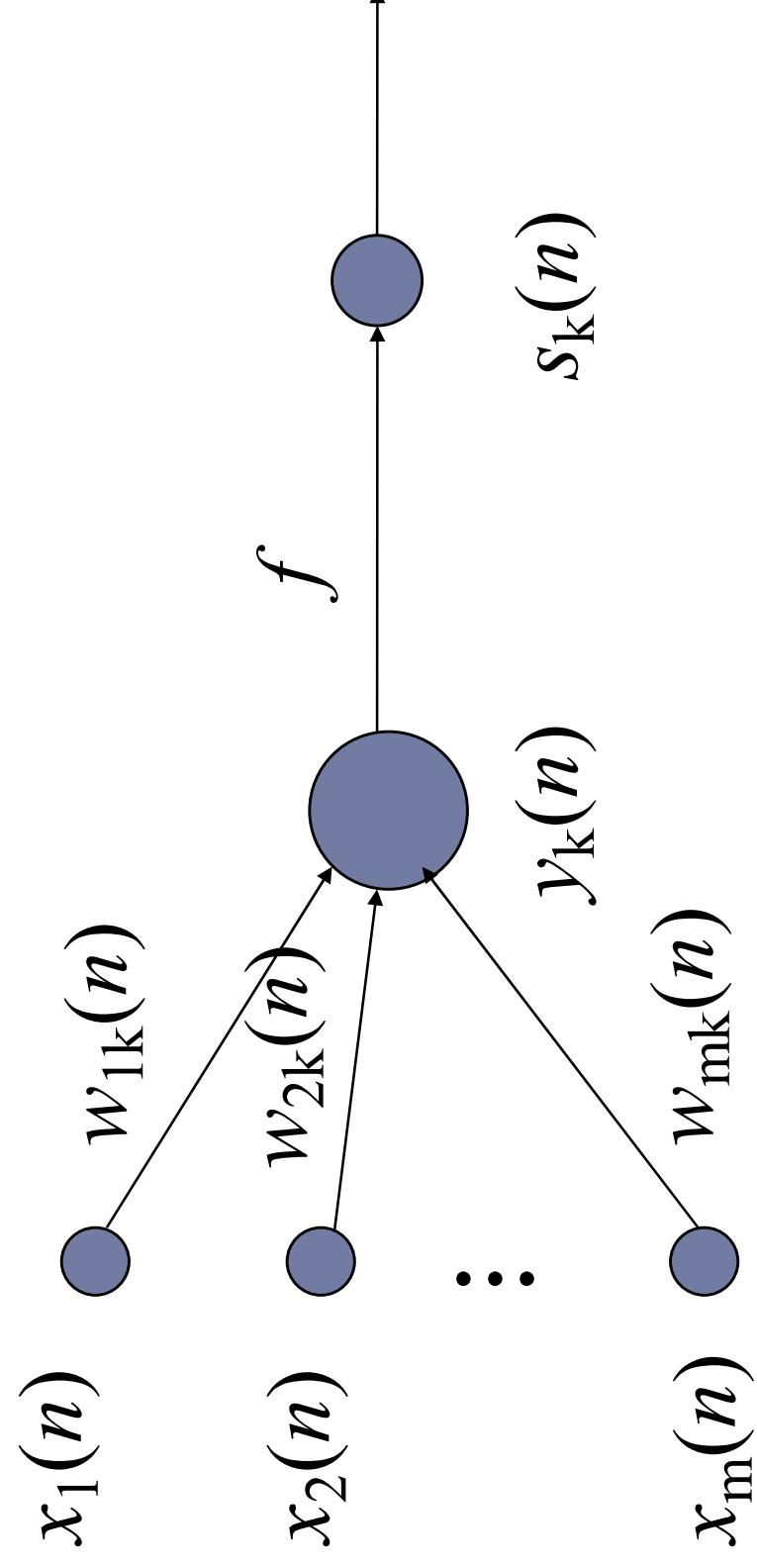
4. Mecanismo correlativo: a correlação temporal entre as atividades das sinapses as ligações

HEBBIAN LEARNING

1. Variações:

- 1.** Sinapse hebbiana
- 2.** Sinapse anti-hebbiana
- 3.** Sinapse não-hebbiana

HEBBIAN LEARNING



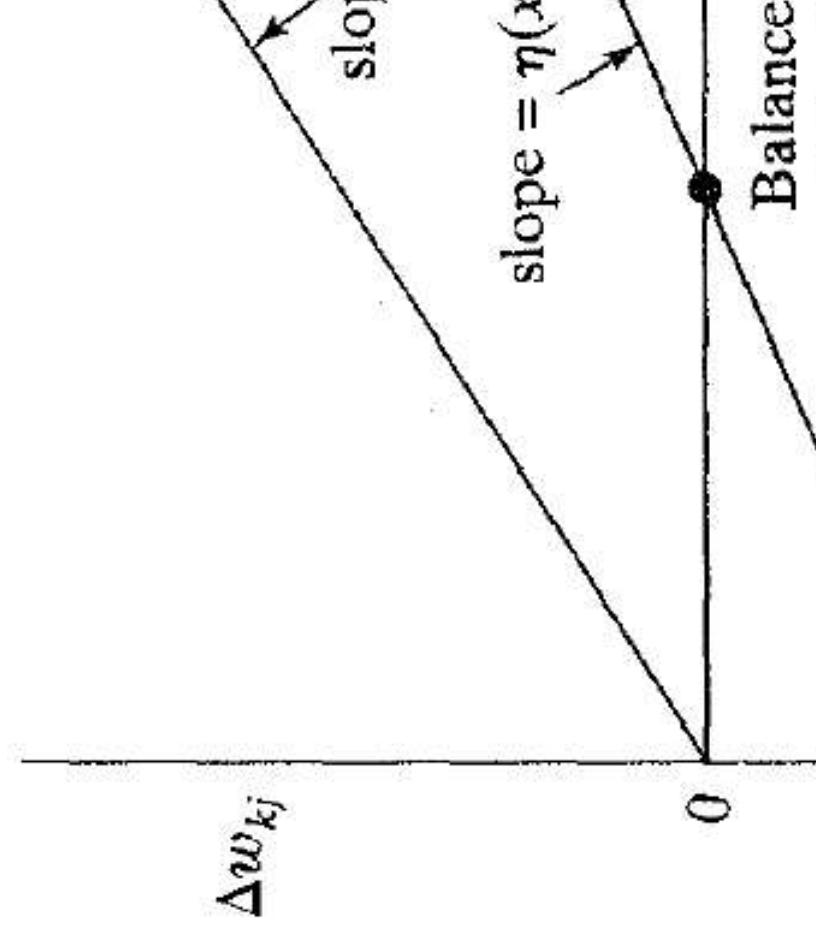
HEBBIAN LEARNING

1. Forma geral:

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

3. Forma correlativa

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



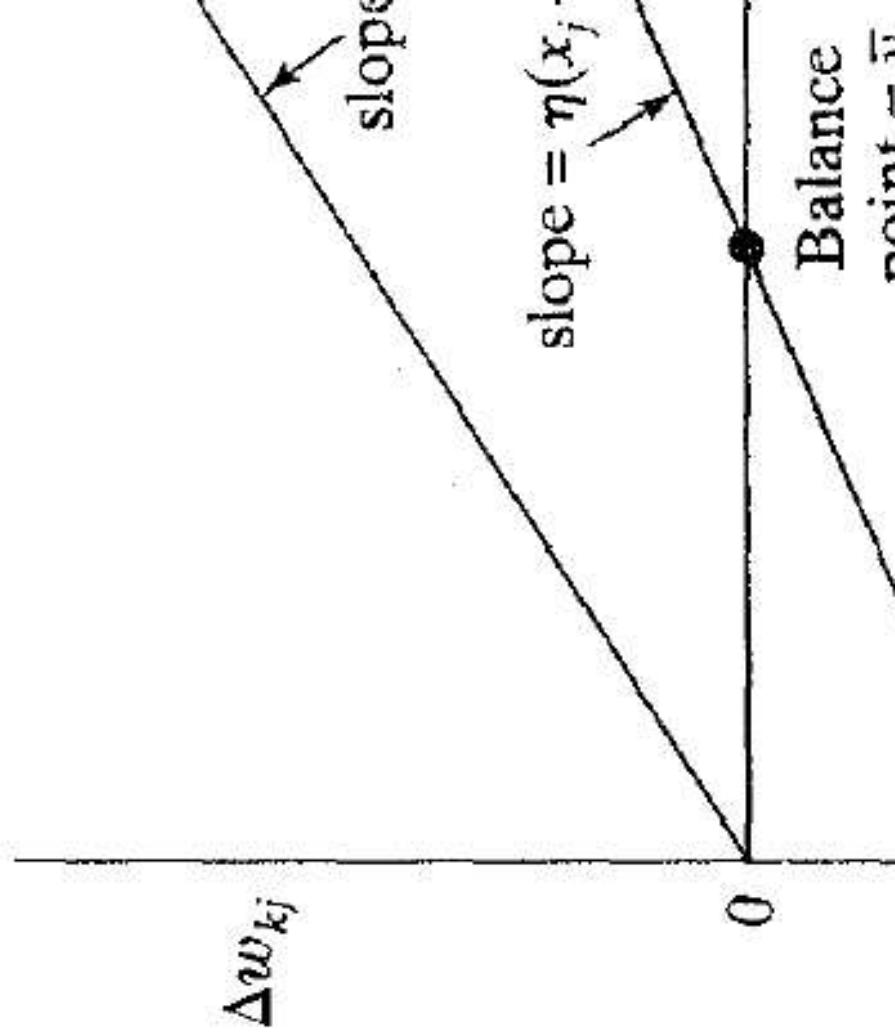
HEBBIAN LEARNING

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

1. Forma correlativa

2. Saturação sináptica
com aumento
exponencial de y

3. Nenhuma
informação pode ser
armazenada



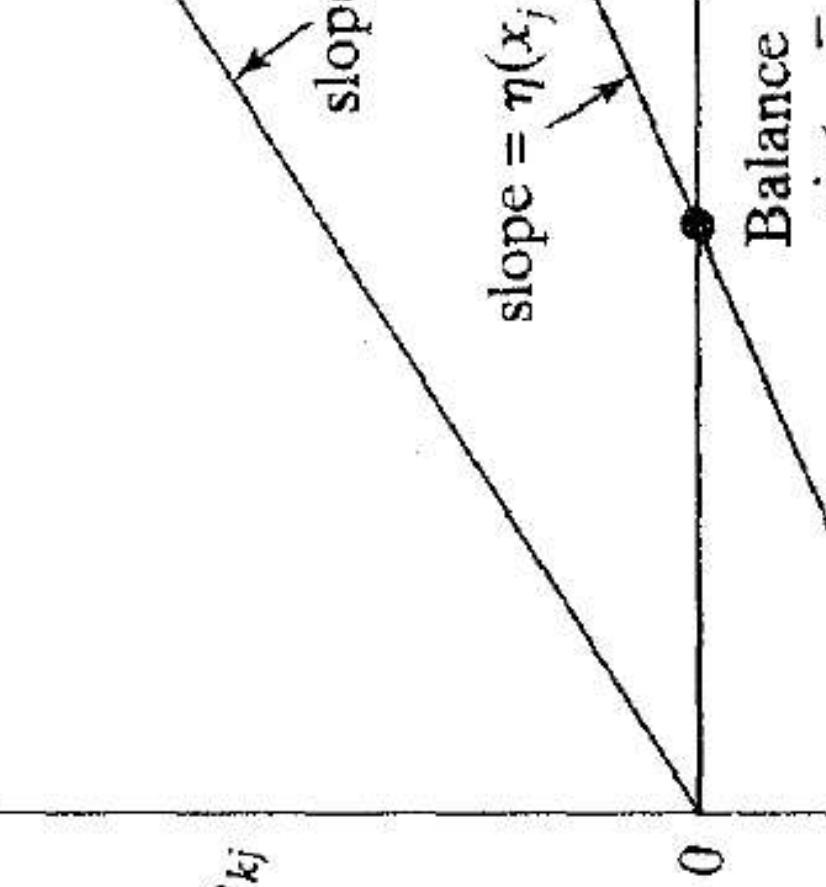
HEBBIAN LEARNING

1. Forma da Covariância

$$\Delta w_{kj} = \eta \left(x_j - \bar{x} \right) \left(s_k - \bar{s} \right)$$

X, S

Δw_{kj}



5. valor médios, que constituem os limiares pré-sináptico e pós-

HEBBIAN LEARNING

1. Forma da Covariância

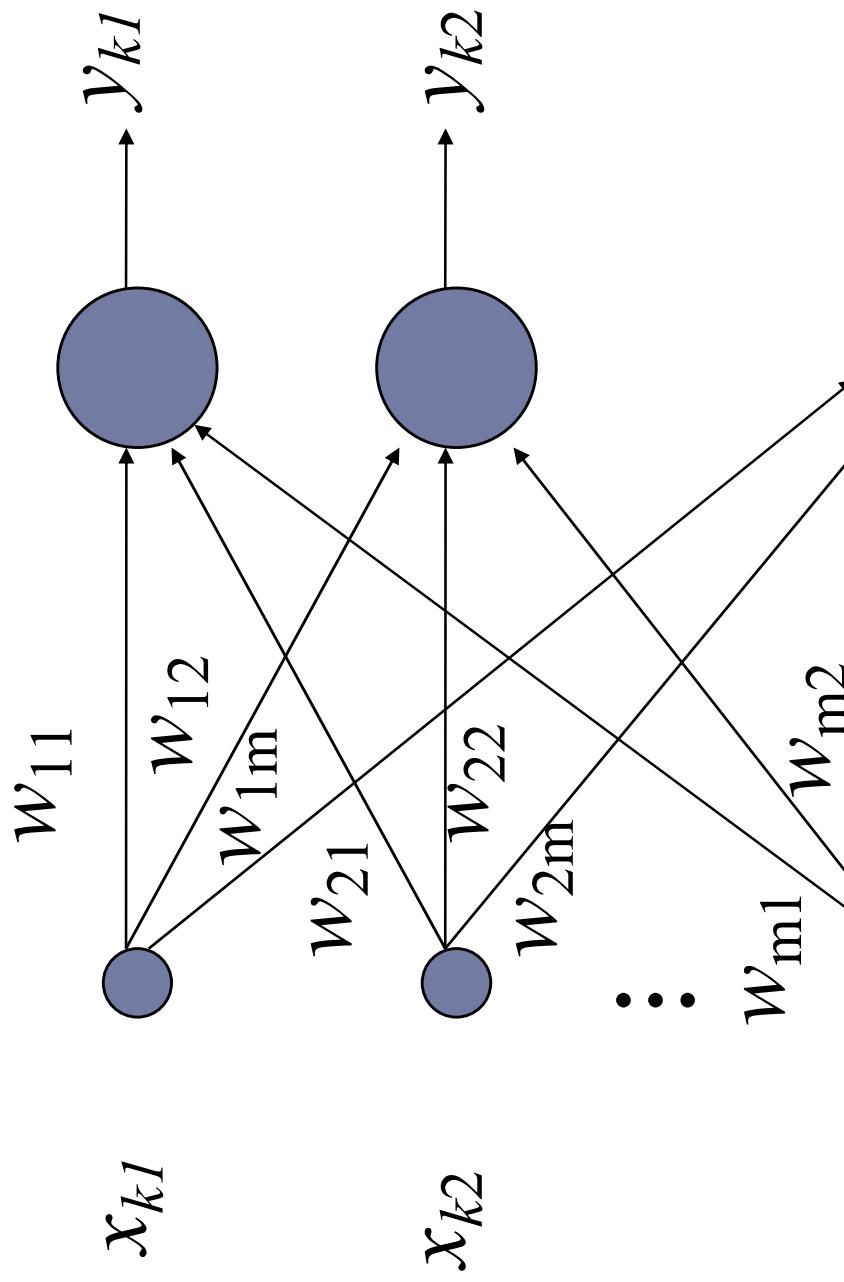
$$\Delta w_{kj} = \eta (x_j - \bar{x})$$

2. Observações sobre a hipótese da covariância:

1. O peso sináptico w_{kj} é reforçado se houver níveis de atividades pré e pós-sinápticas
2. O peso é deprimido se ocorrer uma das situações:

HEBBIAN LEARNING - ILLUSTRATION

1. Aprendizado Hebbiano em Associador Linear



MEMÓRIA (ASSOCIADOR LINEAR)

$$\begin{array}{ll} \text{Entrada} & \mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{km}]^T \\ \text{Saída} & \mathbf{y}_k = [y_{k1}, y_{k2}, \dots, y_{km}]^T \end{array}$$

Transformação linear $\mathbf{y}_k = \mathbf{W}(k)\mathbf{x}_k$

$$\mathbf{W}(k) = \begin{bmatrix} w_{11}(k) & w_{12}(k) & \dots & w_{1m}(k) \\ w_{21}(k) & w_{22}(k) & \dots & w_{2m}(k) \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}$$

MEMÓRIA (ASSOCIADOR LINEAR)

1. Cada par de padrões (k) é associado por sua respectiva matriz

$$\mathbf{y}_k = \mathbf{w}(k) \mathbf{x}_k$$

3. Podemos definir uma matriz de memória m -por- m que descreve matrizes para o conjunto inteiro de associações de padrões:

$$\mathbf{M} = \sum_{k=1}^q \mathbf{w}(k)$$

6. Essa matriz define a conectividade global entre as camadas de da memória associativa

MEMÓRIA (ASSOCIADOR LINEAR)

1. Essa matriz contém uma parte de cada par entrada-saída apresentados a memória

$$M = \sum_{k=1}^q w(k)$$

3. Reestruturando de forma recursiva:

$$M_k = M_{k-1} + w(k)$$

5. $w(k)$ é “diluída” na em M_k , contudo a informação sobre armazenados pode ser preservada.

MEMÓRIA (ASSOCIADOR LINEAR)

1. Suponha que $\hat{\mathbf{M}}$ apreendeu todos os k padrões
2. Vamos supor que a matriz $\hat{\mathbf{M}}$ representa uma matriz memória \mathbf{M}

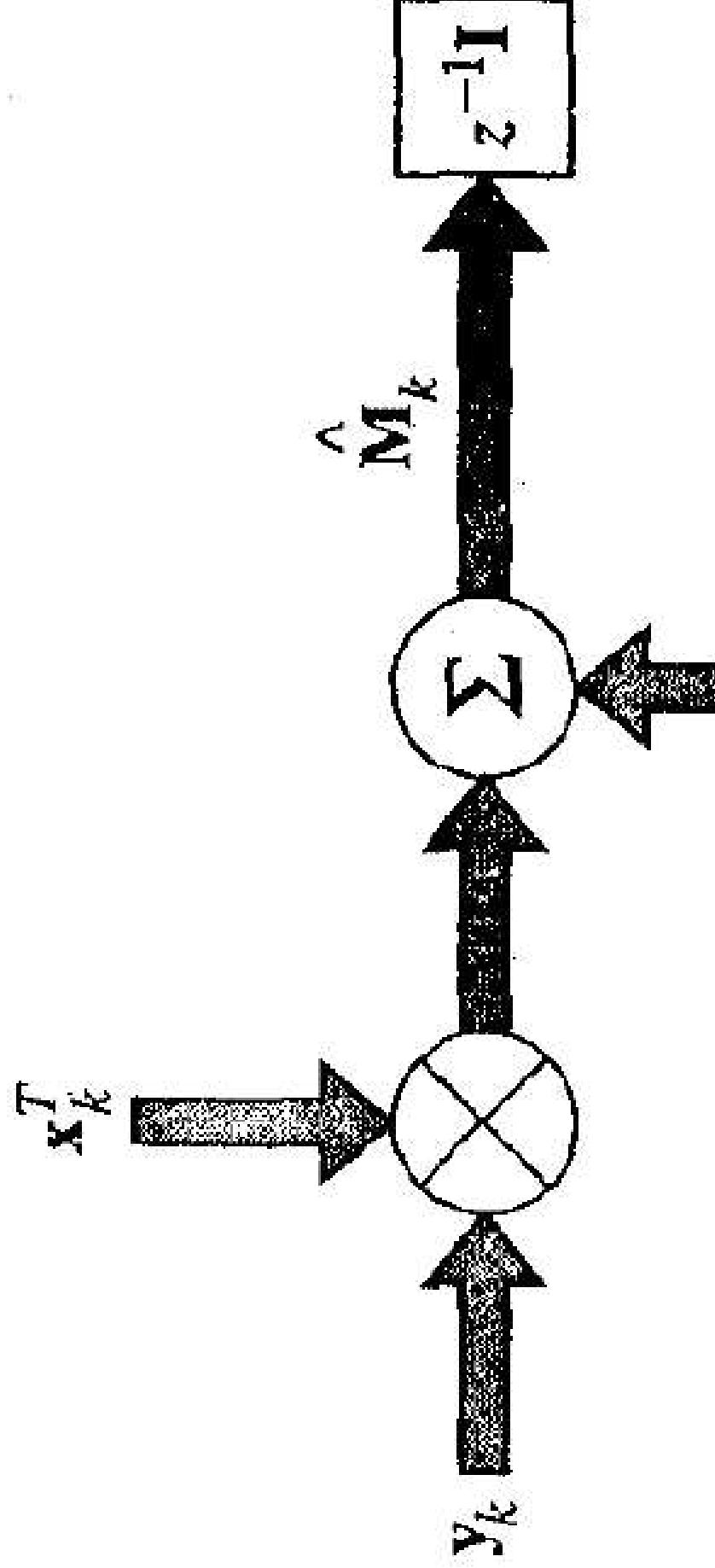
Regra de Aprendizado:

$$\hat{\mathbf{M}} = \sum_{k=1}^q \mathbf{y}_k \mathbf{x}_k^T \quad \begin{array}{l} \text{Estimativa} \\ \text{de memória} \end{array}$$

$\hat{\mathbf{M}} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_q^T \end{bmatrix}$

MEMÓRIA (ASSOCIADOR LINEAR)

1. Representação por grafo de fluxo



MEMÓRIA (ASSOCIADOR LINEAR)

Recordação para entrada \mathbf{x}_j :

$$\mathbf{y} = \mathbf{M}^{\wedge} \mathbf{x}_j = \sum_{k=1}^m \mathbf{y}_k \mathbf{x}_k^T \mathbf{x}_j = \sum_{k=1}^m \mathbf{y}_k (\mathbf{x}_k^T \mathbf{x}_j)$$

MEMÓRIA (ASSOCIADOR LINEAR)

$$\mathbf{y} = \mathbf{y}_j + \mathbf{v}_j$$

$$\mathbf{v}_j = \sum_{\substack{k=1 \\ k \neq j}}^m (\mathbf{x}_k^T \mathbf{x}_j) y_k$$

Interferência

Assume-se $\mathbf{x}_i^T \mathbf{x}_i = 1$

MEMÓRIA (ASSOCIADOR LINEAR)

$$\begin{aligned}\cos(\mathbf{x}_k, \mathbf{x}_j) &= \frac{\mathbf{x}_k^T \mathbf{x}_j}{\|\mathbf{x}_k\| \|\mathbf{x}_j\|} \\ &= \mathbf{x}_k^T \mathbf{x}_j \\ \mathbf{v}_j &= \sum_{\substack{k=1 \\ k \neq j}}^m \cos(\mathbf{x}_k, \mathbf{x}_j) \mathbf{y}_k\end{aligned}$$

A memória é perfeita (a resposta é iguala \mathbf{y}_j) se os vetores-chave pertencerem a um conjunto ortogonal.

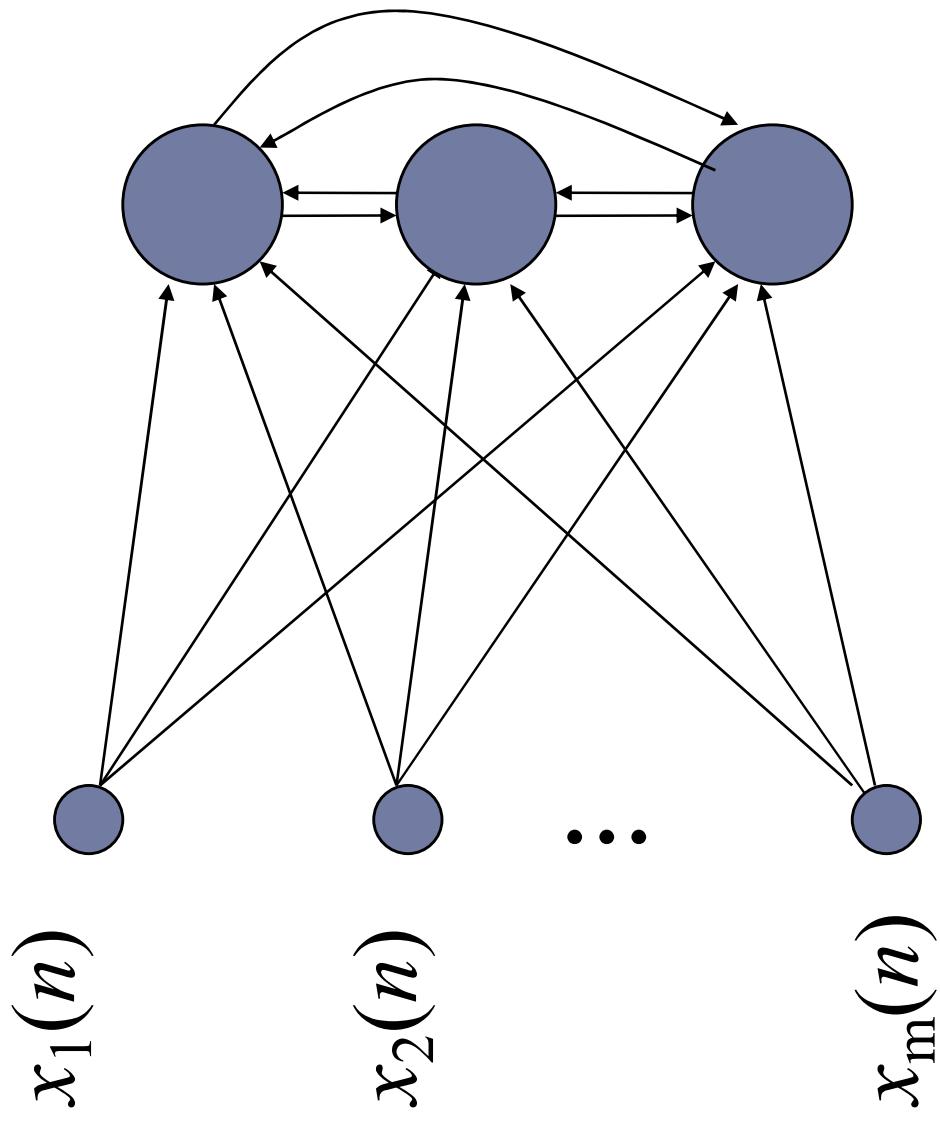
COMPETITIVE LEARNING

- Neurônios de saída competem entre si para se tornar ativos (
 - Um neurônio (*vencedor*) ganha o direito de disparar enquanto permanecem em repouso
 - Existe três elementos básicos em uma regra de aprendizagem
- 1.** Um conjunto de neurônios que são todos iguais entre si, e pesos sinápticos distribuídos aleatoriamente, e por isso respondem diferentemente a um conjunto de padrões de entrada;

COMPETITIVE LEARNING

2. Um limite imposto sobre a “força” de cada neurônio
3. Um mecanismo que permita que somente um neurônio por grupo, esteja saída, ou somente um neurônio que vence a competição é denominado vencedor

COMPETITIVE LEARNING



Camada de
Camada de

COMPETITIVE LEARNING

$$y_k = \begin{cases} 1 & se v_k > v_j \text{ para todos } j, j \neq k \\ 0 & caso contrário \end{cases}$$

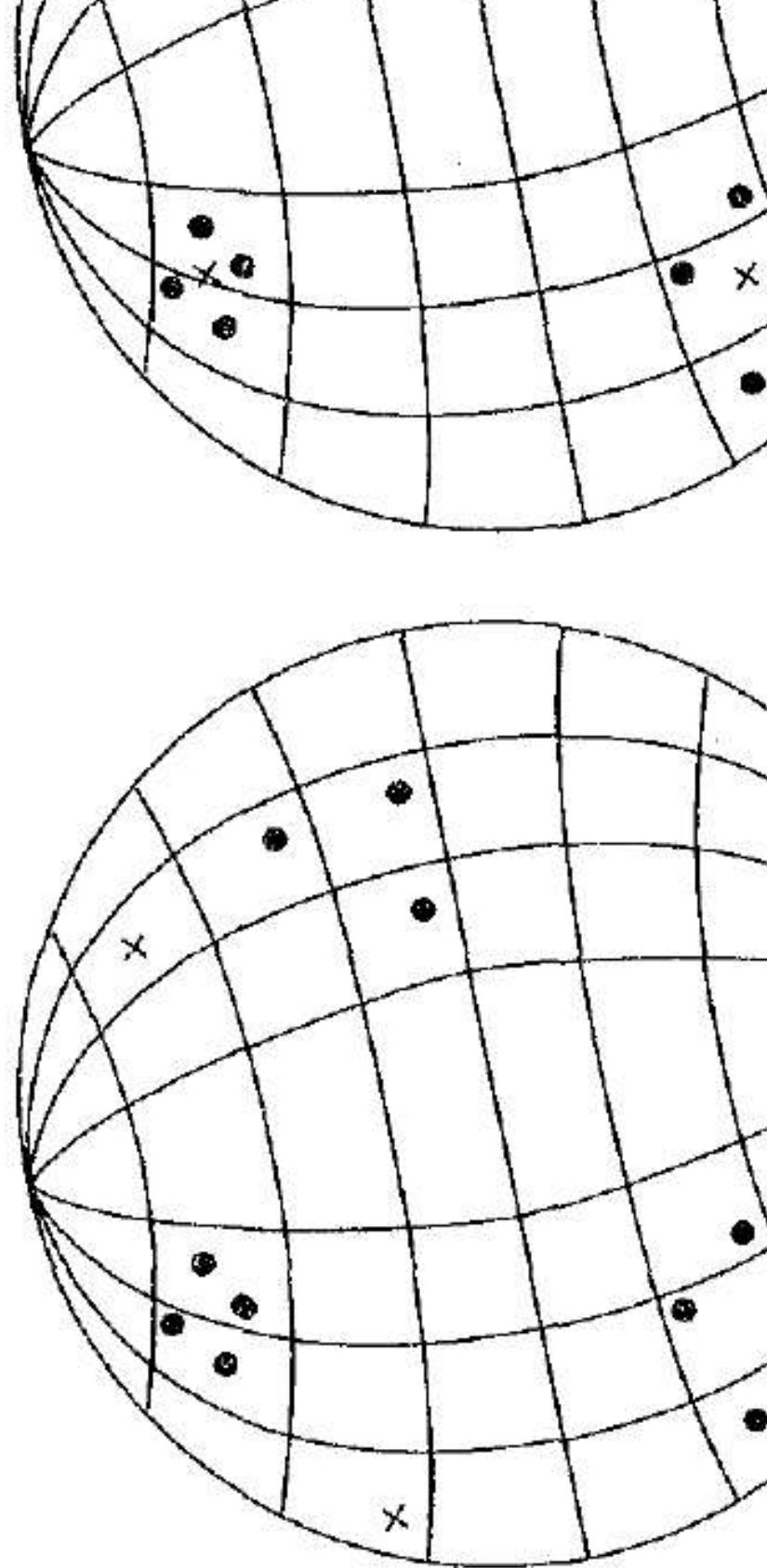
O neurônio k é denominado como neurônio *vencedor*:

Regra de aprendizagem competitiva:

$$\Delta_{v_k} = \eta(x_j - w_{kj}) \quad y_k = 1$$

COMPETITIVE LEARNING

1. Analogia geométrica:



COMPETITIVE LEARNING

1. Os neurônios se especializam nos padrões de dados
2. Tornam-se “protótipos” do agrupamento que resulta
3. Ou ainda, detectores de características de um conjunto de padrões

BOLTZMANN LEARNING

1. Regra de aprendizagem enraizada nas ideias da mecânica estatística. Homenagem a Ludwig Boltzmann
2. Máquina de Boltzmann – Proposta por Geoffrey Hinton e Sejnowski em 1985
3. Neurônios representam células binárias (estádo ligado ou desligado)
4. A máquina é caracterizada por uma função de energia

BOLTZMANN LEARNING

5. A máquina opera de forma estocástica, alterando um dado neurônio com probabilidade:

$$P(X_k \rightarrow -X_k) = \frac{1}{1 + \exp(-\Delta E_k / T)}$$

6. A aplicação contínua dessa regra irá levar a máquina ao equilíbrio “térmico”

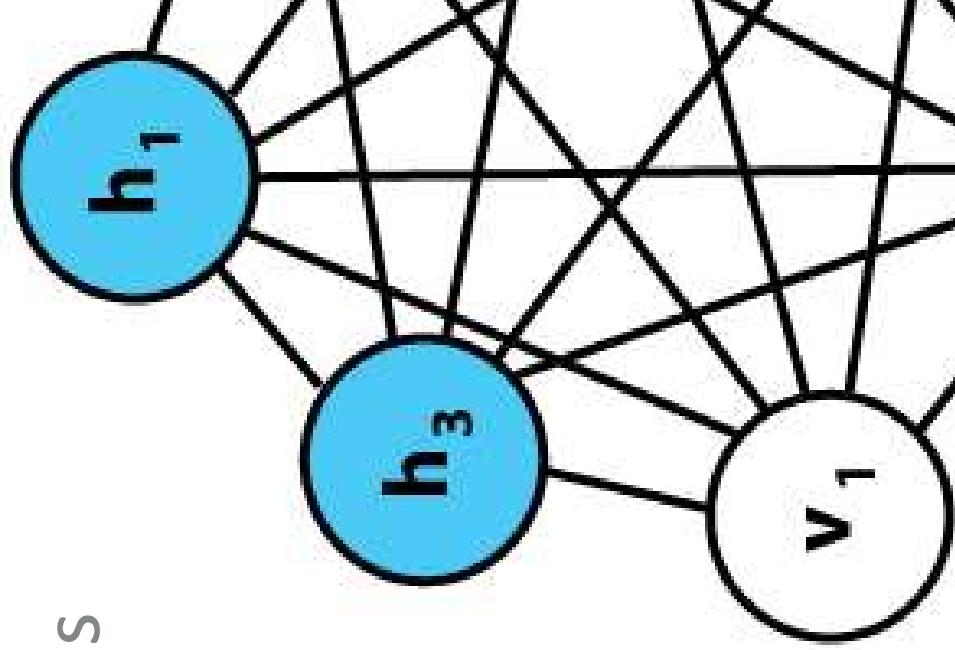
BOLTZMANN LEARNING

1. A máquina de Boltzmann contém dois tipos de neurônios: Visíveis e Ocultos

2. Duas formas de operação:

1. Presa: neurônios visíveis com estados fixos

2. Livre: neurônios operam livremente



APRENDIZAGEM DE BOLTZMANN

1. Aprendizagem:

$$\Delta W_{kj} = \eta (\rho_{kj}^+ - \rho_{kj}^-)$$

4. ρ_{kj}^+ Representa a correlação entre os estados dos neurônios condicão presa
6. ρ_{kj}^- Representa a correlação entre os estados dos neurônios

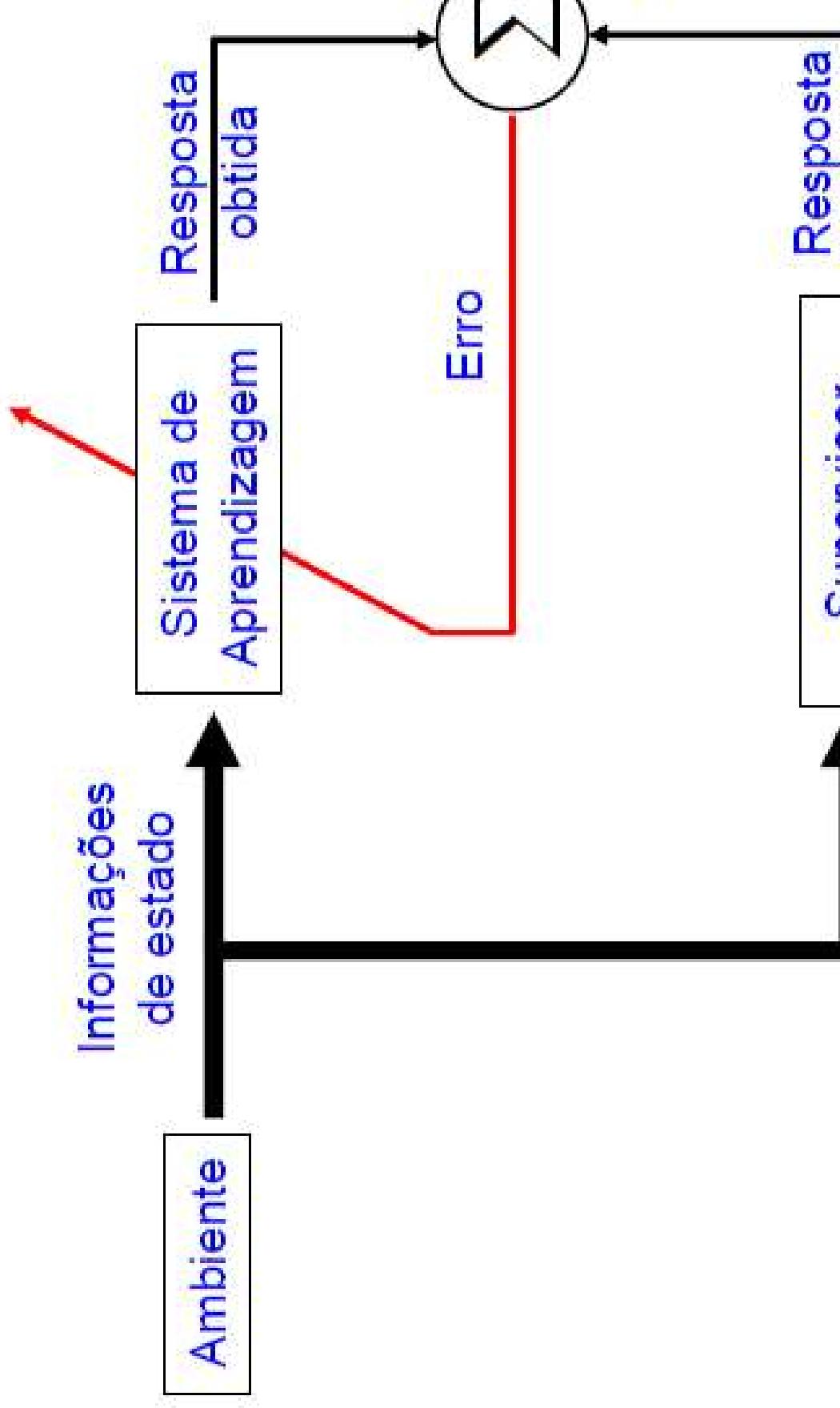
LEARNING PARADIGMS

- 1.** Supervised Learning
- 2.** Unsupervised Learning
- 3.** Reinforcement Learning
- 4.** Outros...

SUPERVISED LEARNING

1. A entrada e saída desejadas para a rede são fornecidas supervisor
2. Objetivo é ajustar os parâmetros da rede, de forma a **ligação** entre os pares de entrada e saídas fornecidas
3. O professor indica explicitamente um comportamento para a rede, visando direcionar o processo de treinamento
 - *Forma mais comum: aprendizagem por correção*

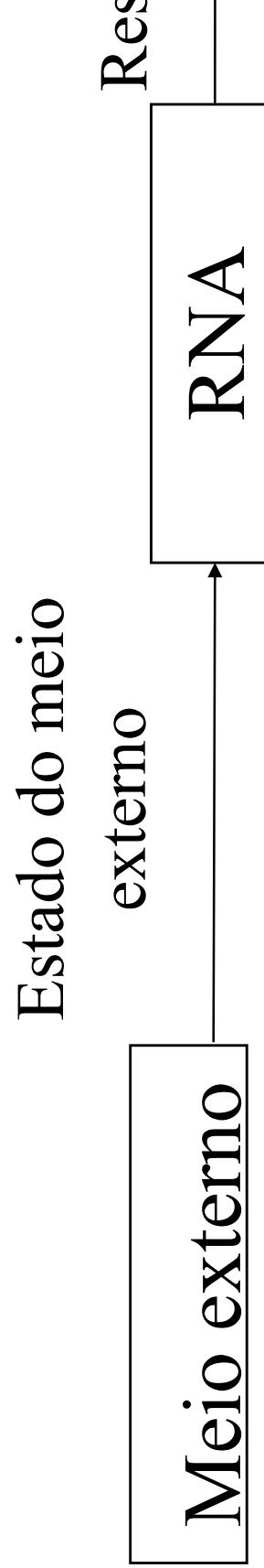
SUPERVISED LEARNING



UNSUPERVISED LEARNING

- 1.** Não há um professor ou supervisor para acompanhar de aprendizado;
- 2.** A partir do momento em que a rede estabelece uma com as regularidades estatísticas da entrada de dado se nela uma habilidade de formar representações int codificar características de entrada;
- 3.** Este tipo de aprendizado só se torna possível quando redundância nos dados de entrada.

UNSUPERVISED LEARNING



O aprendizado competitivo é o mais comum nessa forma de aprendizagem

SEMI-SUPERVISED LEARNING

1. O aprendizado semi-supervisionado utiliza dados não rotulados e normalmente utiliza tanto conceito aprendizado supervisionado quanto do aprendizado supervisionado.
2. Geralmente, para se construir um classificador utilizando o aprendizado semi-supervisionado, uma base de dados composta por uma grande quantidade de dados não rotulados e apenas alguns itens rotulados é utilizada.

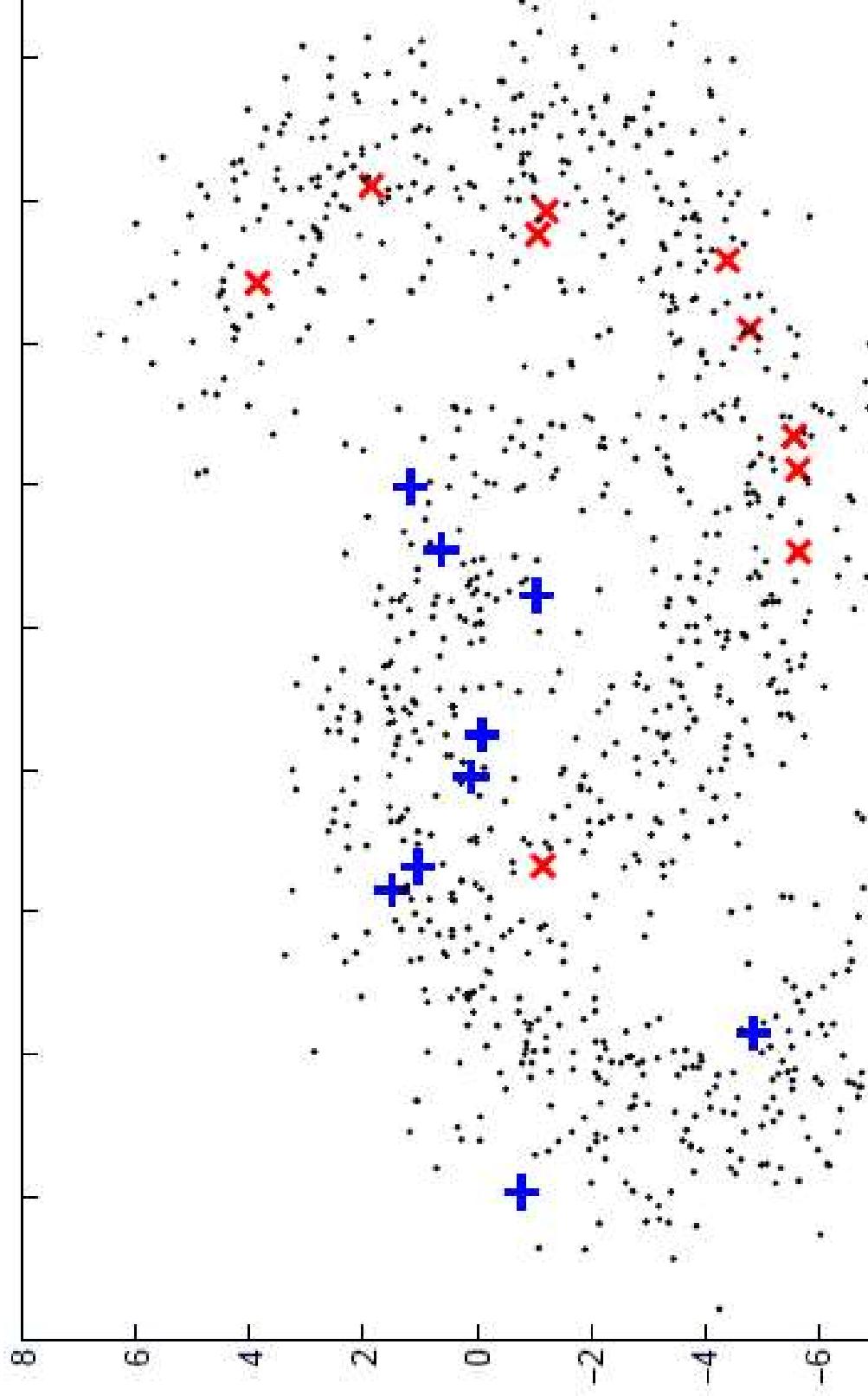
SEMI-SUPERVISED LEARNING

1. O aprendizado semi-supervisionado pode ter como fatores de classificação de dados quanto tarefas de a
2. No caso da classificação, os exemplos pré-rotulados no processo de rotulação de exemplos não rotulados
3. No agrupamento de dados, os exemplos rotulados podem guiar o processo de formação de *clusters*

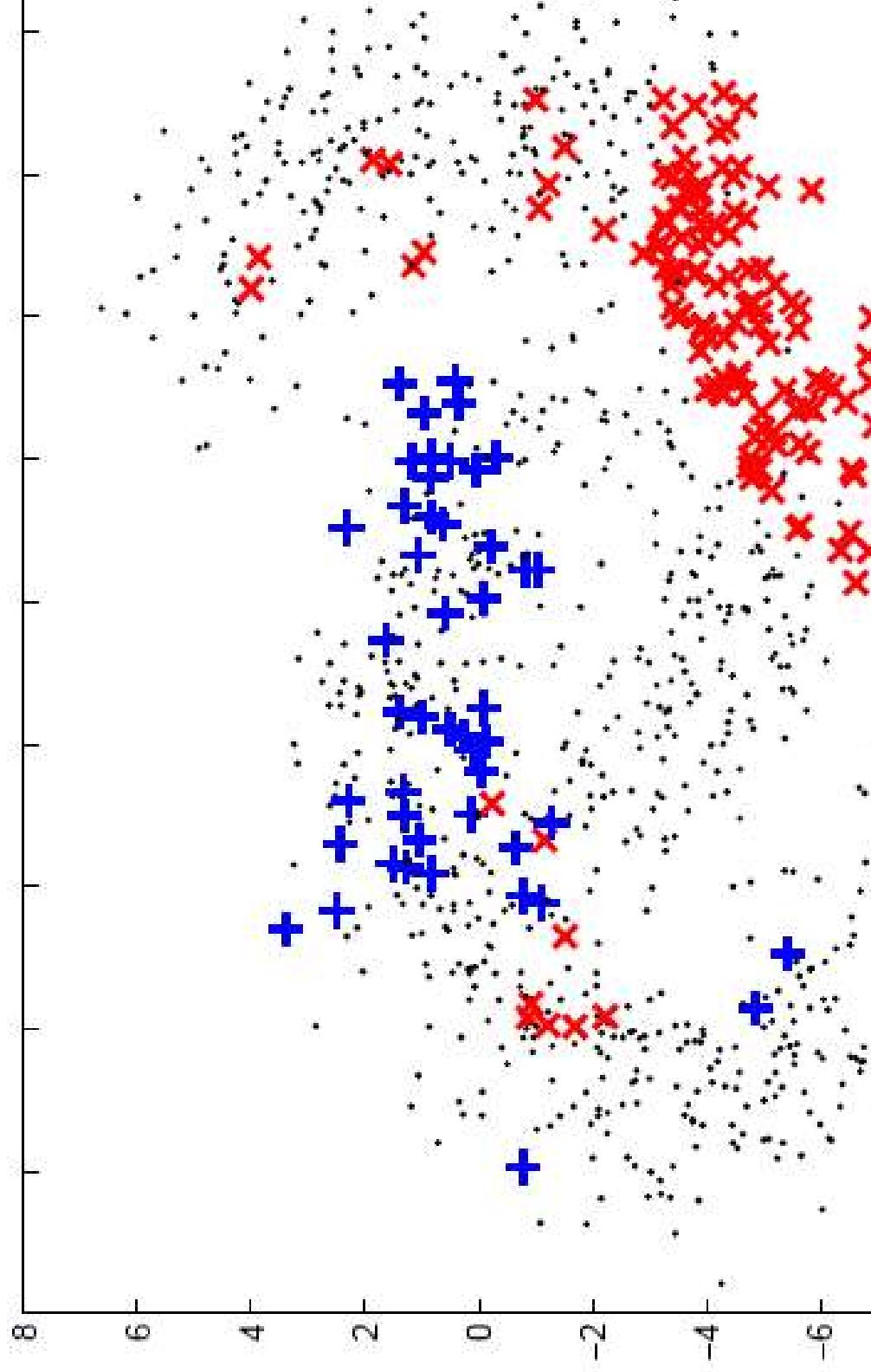
SEMI-SUPERVISED LEARNING

1. Aprendizado Transdutivo
2. Aprendizado Indutivo

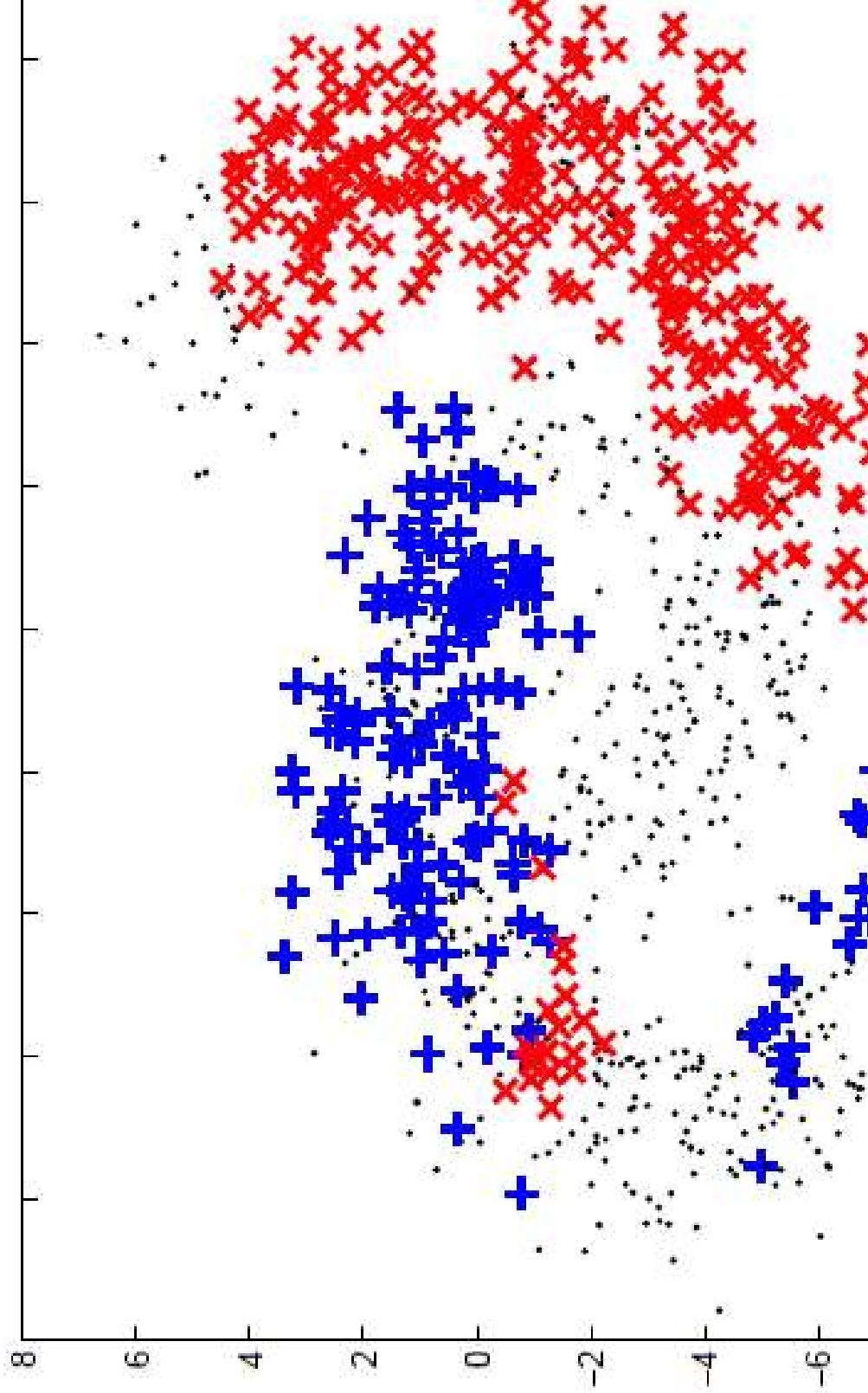
SEMI-SUPERVISED LEARNING



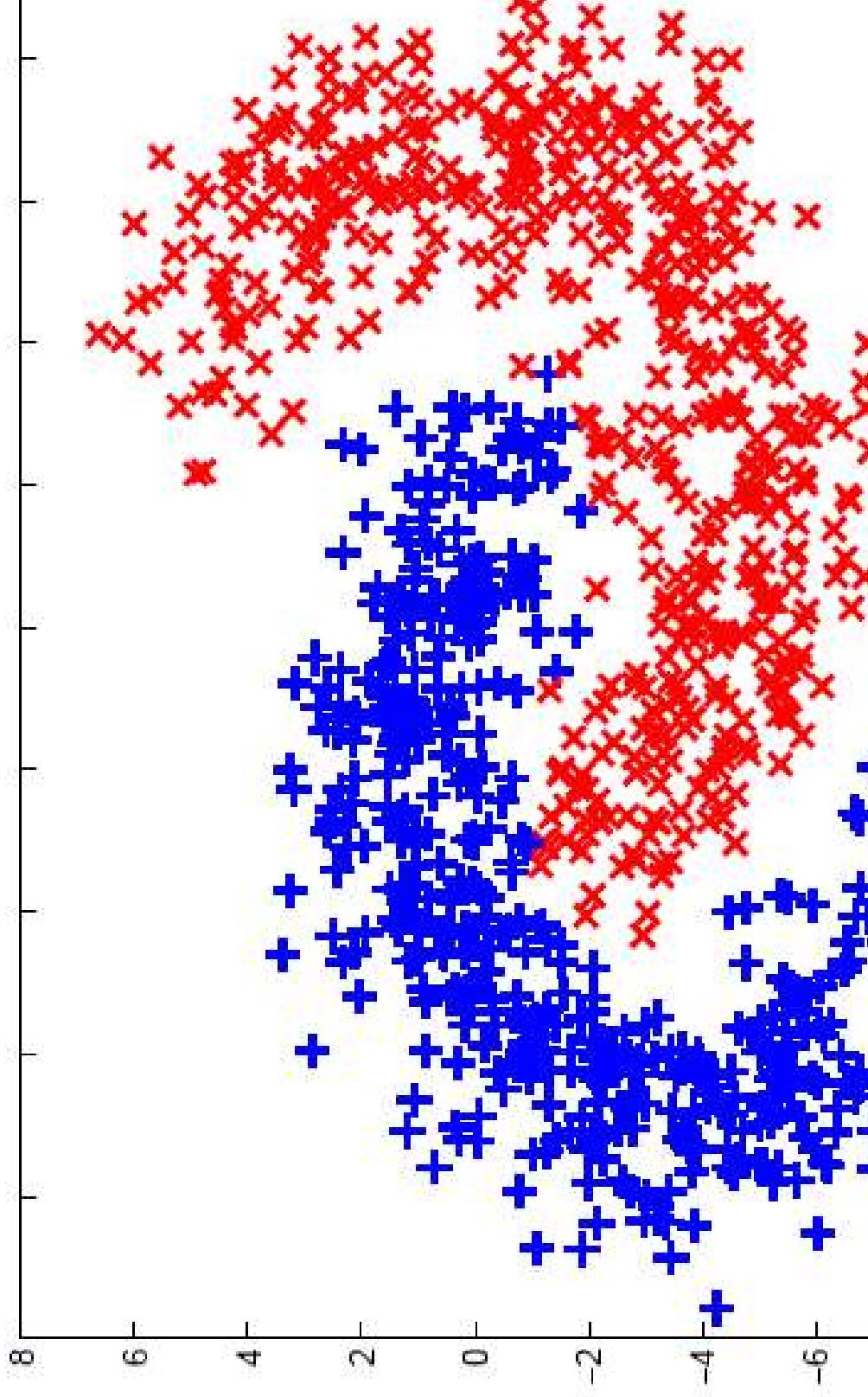
SEMI-SUPERVISED LEARNING



SEMI-SUPERVISED LEARNING



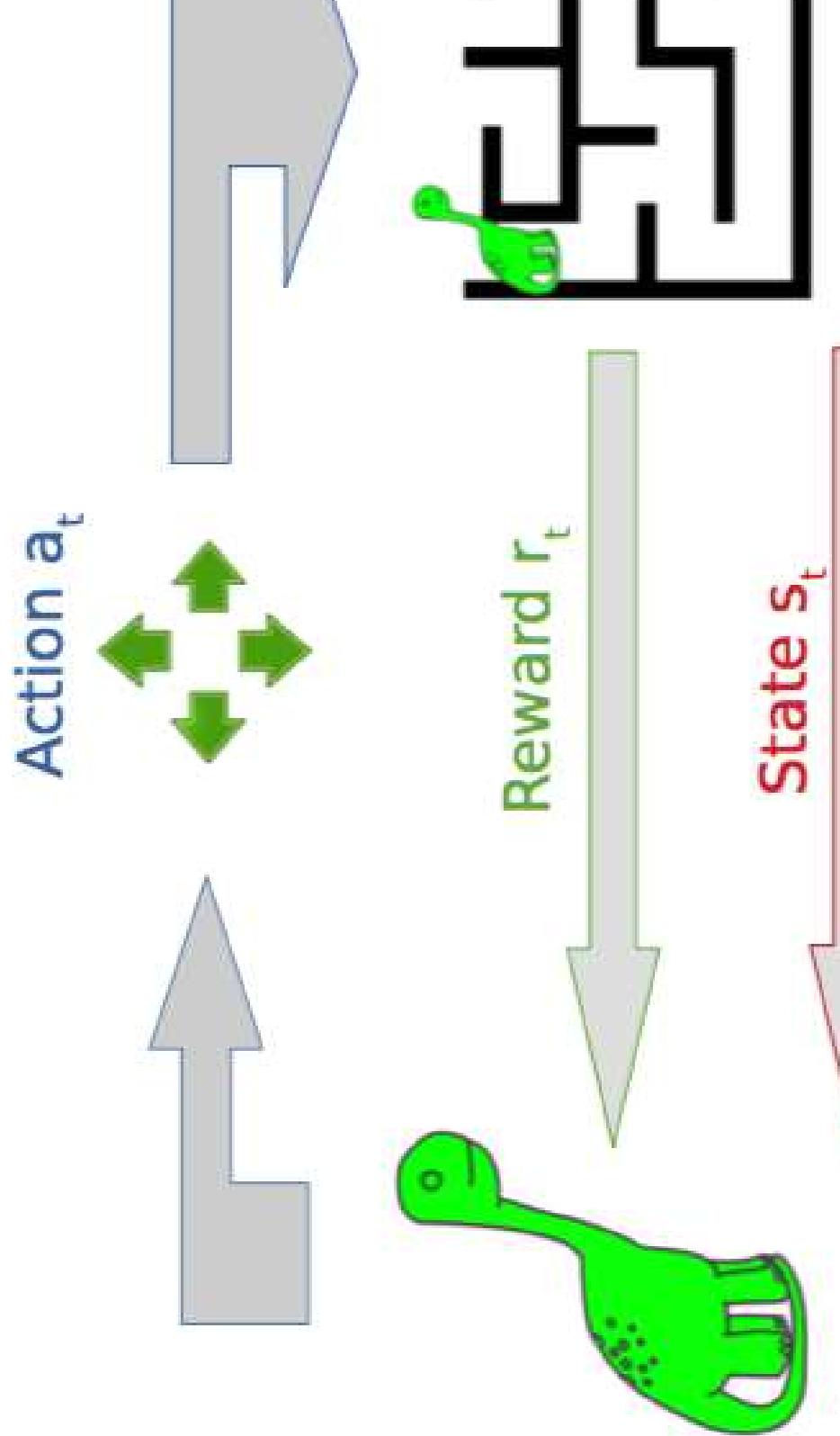
SEMI-SUPERVISED LEARNING



WHAT ABOUT REINFORCEMENT LEARNING?

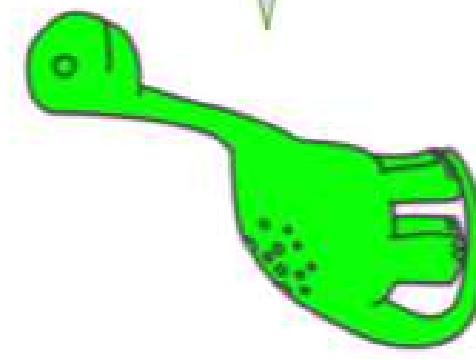
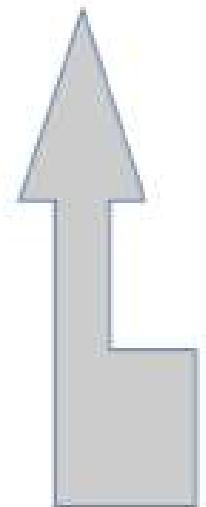
1. Dados são coletados a partir da interação do ambiente
 - a. Agente
 - b. Estado
 - c. Ações

WHAT ABOUT REINFORCEMENT LEARNING?



WHAT ABOUT REINFORCEMENT LEARNING?

Action a_t



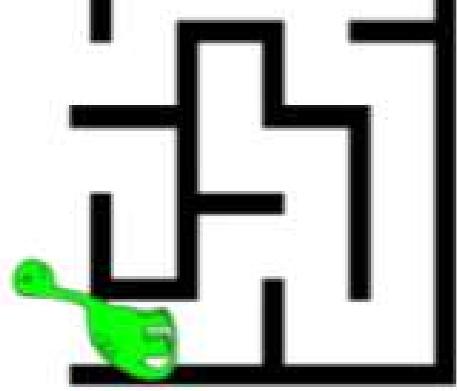
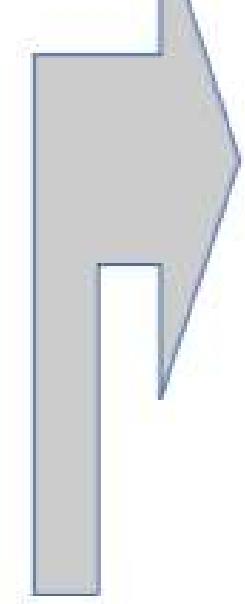
Reward r_t



State s_t

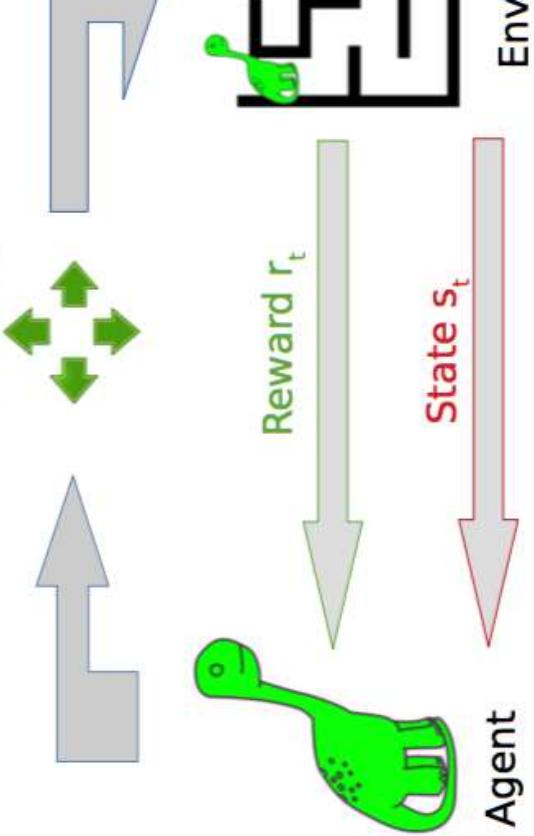


Agent



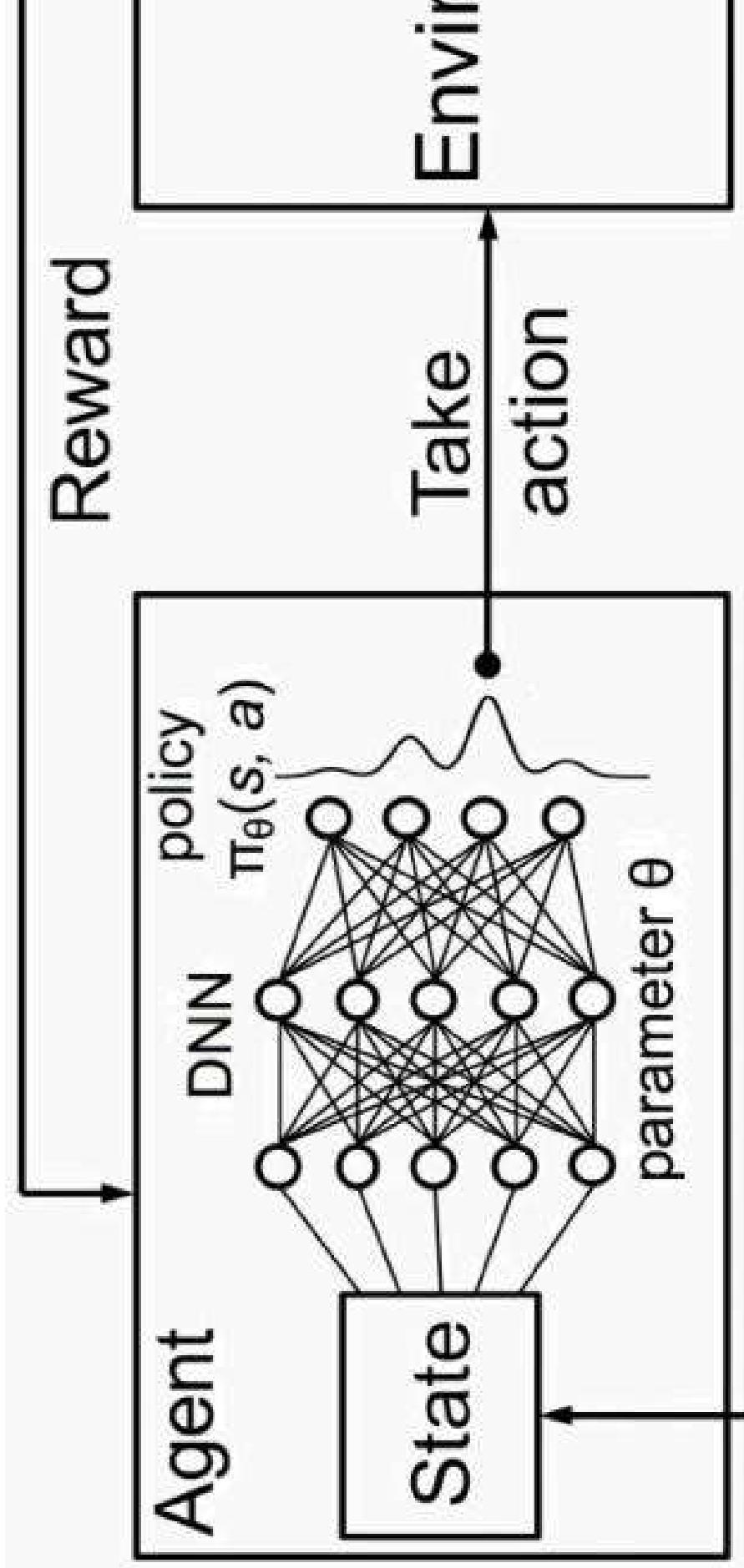
Environment

WHAT ABOUT REINFORCEMENT LEARNING?



1. Objetivo do algoritmo de aprendizagem: Aprender regras a partir da resposta do ambiente
1. Q-Table (Estados vs Ações)

WHAT ABOUT REINFORCEMENT LEARNING?



MORE LEARNING

- Self-Supervised Learning
- Contrastive Learning
- Transfer Learning
- Active Learning
- Etc.



REFERENCES

- Goodfellow, Bengio & Courville (2016) “Deep Learning”
The MIT Press (<https://www.deeplearningbook.org>)
- Simon Haykin (2001) “Redes Neurais: Princípio
2 ed. Bookman
- Antônio Braga, Teresa Ludemir e André Carvalho
“Redes Neurais Artificiais: Teoria e Aplicações”