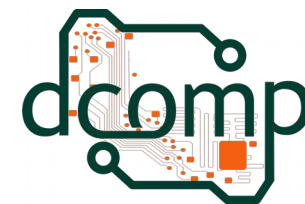




Universidade Federal do Espírito Santo
Centro de Ciências Agrárias – CCA UFES
Departamento de Computação



Paradigmas de Aprendizagem

Redes Neurais Artificiais

Site: <http://jeiks.net>

E-mail: jacsonrcsilva@gmail.com

Paradigmas de Aprendizagem

- Representação do Conhecimento.
- Paradigmas de Aprendizagem.
- Processos de Aprendizagem.

Representação do Conhecimento

- Definição genérica de conhecimento:
 - *Conhecimento se refere à informação armazenada ou a modelos utilizados por uma pessoa ou máquina para interpretar, prever e responder apropriadamente ao mundo exterior (citação de Fischler e Firschein, 1987, em Haykin, 1999).*
- Principais características da representação do conhecimento:
 - Quais informações realmente são expostas;
 - Como a informação realmente é codificada de forma física para ser utilizada posteriormente.
- Assim, podemos ver que a representação do conhecimento é direcionada a um bom objetivo.
- Entretanto, escolher os parâmetros corretos de uma RNA para o aprendizado dos dados reais costuma ser complicada devido a diversidade de escolhas existentes:
 - Isso transforma a construção (*design*) de uma RNA em um desafio.

Representação do Conhecimento

- A RNA deve:
 - Aprender um modelo do mundo (ambiente de trabalho);
 - Manter um modelo consistente com o mundo real, de forma a atingir os objetivos de interesse de onde está sendo aplicada.
- O conhecimento do mundo consiste em dois tipos de informação:
 - Pelo estado conhecido do mundo, representado pelos fatos que dizem o que é pelo que se conhece (*informação prévia*);
 - Pelas observações do mundo, geradas por medidas provenientes de sensores ou cálculos. São sobre essas medidas que a RNA trabalhará e produzirá resultados.

Essas observações formam *exemplos* que serão fornecidos à RNA.

Representação do Conhecimento

- Os *exemplos* podem ser:
 - Com rótulos: Para cada padrão de *entrada* existe uma *resposta desejada* já conhecida e agregada aos exemplos;
 - Sem rótulos: Onde os padrões de entrada possuem diferentes percepções (características) do ambiente por si só que devem ser agrupadas (*clustering*) pela RNA.
- De qualquer forma, quaisquer exemplos apresentam o conhecimento do ambiente para a RNA.
- Um conjunto de exemplos, consistindo de entrada e saída (com rótulos) ou somente de entradas (sem rótulos), são chamados de:
 - Conjunto de dados de treinamento; ou
 - Amostra de treinamento; ou
 - Padrões de treinamento.

Representação do Conhecimento

- Considere o problema de reconhecer um *dígito manuscrito* (digitalizado para uma imagem de 10 pixels):
 - Uma arquitetura da RNA deve ser criada:
 - Possuindo 10 neurônios de entrada, um para cada pixel;
 - Possuindo uma camada oculta com H neurônios;
 - Possuindo C saídas para a classificação desejada.
 - A RNA deve receber um subconjunto de treino e deve ser treinada por um algoritmo apropriado:
 - Fase chamada de **Aprendizagem**.
 - Deve-se então testar o reconhecimento da rede treinada:
 - Isso é feito comparando-se a saída da RNA com a saída desejada
 - Essa fase é chamada de **Generalização**, um termo da psicologia.
- Os exemplos podem ser tanto **positivos** quanto **negativos**.
- Os exemplos devem ser estandardizados (*standardization*).

Representação do Conhecimento

- Em uma Rede Neural Artificial,
 - A representação do conhecimento do meio ambiente é definida pelos valores dos **parâmetros livres** da rede.
 - Parâmetros Livres:
 - Pesos sinápticos;
 - Bias.
- O design (arquitetura, formação, distribuição dos neurônios e sinapses) da RNA constituem a forma de representação desses parâmetros livres.

Essa é a chave do desempenho das RNAs.

- Porém, a representação do conhecimento dentro da RNA é realmente complicada e existem somente quatro regras básicas para sua criação/desenvolvimento.

Representação do Conhecimento

- Regra 1:
 - Entradas similares de classes similares normalmente devem produzir representações similares no interior da RNA.
 - Portanto devem ser classificadas como pertencentes à mesma categoria.
 - Classificação de Entradas similares:
 - Distância euclidiana;
 - Tamanho da projeção de x_i em x_j ;
 - Distância de Mahalanobis: distância euclidiana entre x_i e sua média.

Representação do Conhecimento

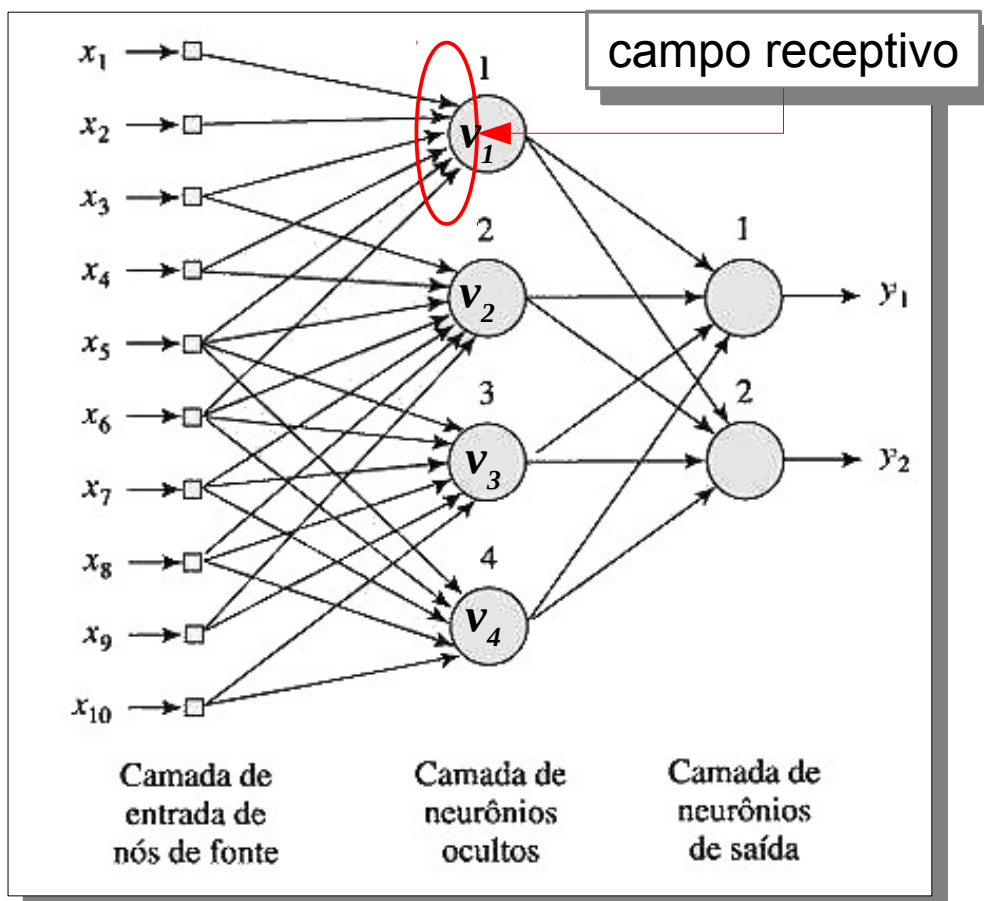
- Regra 2:
 - Devem ser atribuídas representações bem diferentes na rede a itens que devem ser categorizados como classes separadas;
Contrário da regra 1.
- Regra 3:
 - Se uma característica particular é importante, então deve haver um grande número de neurônios envolvidos na representação daquele item na RNA.
 - Deve haver um grande número de neurônios envolvidos na tomada de decisão para uma saída correta;
 - Também deve haver um grande número de neurônios envolvidos na tomada de decisão quando existem perturbações;
 - A probabilidade de resposta correta deve ser maior que a probabilidade de alarmes falsos (falso positivos).

Representação do Conhecimento

- Regra 4:
 - Informação prévia e variâncias devem ser incorporadas no projeto de uma rede neural,
 - Simplificando com isso o projeto da rede por não ter que aprendê-las.
 - Isso gera uma *rede especializada*. É desejável, pois:
 - As Redes Neurais Biológicas visuais e auditivas são muito especializadas;
 - Uma RNA especializada possui um número menor de parâmetros livres para serem ajustados. Assim, aprende mais rápido e generaliza melhor;
 - A taxa de transmissão é mais rápida;
 - O custo de construção da RNA é reduzido.

Representação do Conhecimento

- Incorporação de informação prévia no projeto da RNA:
 - Infelizmente não há regras para fazer isso;
 - Mas há alguns métodos que produzem resultados úteis (técnicas de LeCun et al., 1990a, citadas em Haykin, 1999).



- I. *Restringir a arquitetura da rede utilizando conexões locais conhecidas como campos receptivos;*
- II. *Restringir a escolha de pesos sinápticos com a utilização do compartilhamento de peso.*

$$v_j = \sum_{i=1}^6 w_i x_{i+j-1}, \quad j=1,2,3,4$$

Rede Convolutiva

Representação do Conhecimento

- Incorporação de invariâncias no projeto da RNA:
 - Invariância por Estrutura:
 - Incluir características nos pesos, como a replicação do mesmo valor de peso. Ex.: $w_{ji} = w_{jk}$ para todos os pixels com distância igual ao centro da imagem, fazendo com que a RNA seja invariante a rotações do plano.
 - Invariância por Treinamento:
 - Apresentar um número de exemplos diferentes para o mesmo objeto. Os exemplos devem ser escolhidos para corresponder a diferentes transformações do objeto.
 - Invariância do Espaço de Características:
 - Inicialmente, utilizar métodos para extrair características que caracterizem o conteúdo essencial e que sejam invariantes às entradas.
 - Após isso, fornecer essas características à RNA, aliviando assim seu processamento/transformações.

Paradigmas de Aprendizagem

- Aprendizagem Supervisionada.
- Aprendizagem Não-Supervisionada.

Aprendizagem supervisionada

- Também pode ser chamada de aprendizagem com um professor.
- Podemos considerar que o *professor* conhece o ambiente.
- Assim, o conhecimento é representado por um conjunto de exemplos com *entrada e saída desejada*.
- Como a RNA não conhece o ambiente, ela terá que aprender com os exemplos fornecidos.
- Se um novo padrão for apresentado:
 - O professor saberá responder corretamente;
 - Porém, a RNA responderá baseando-se somente no que aprendeu com os exemplos.
 - A resposta do professor é a resposta desejada e representa a ação ótima da RNA.

Aprendizagem supervisionada

- Os parâmetros da RNA são ajustados sob influência:
 - Do vetor de exemplos;
 - Do Sinal de Erro da rede.
- Sinal de Erro:
 - Diferença entre a resposta desejada e a resposta real da RNA.
- O ajuste dos parâmetros da RNA é realizado de forma iterativa ou determinística:
 - O objetivo é *emular* o professor;
 - Assim, o conhecimento do ambiente que está nos exemplos é transferido para a RNA;
 - Quando esta condição é alcançada, pode-se dispensar o professor, pois a RNA pode lidar com o ambiente por si mesma.

Aprendizagem supervisionada

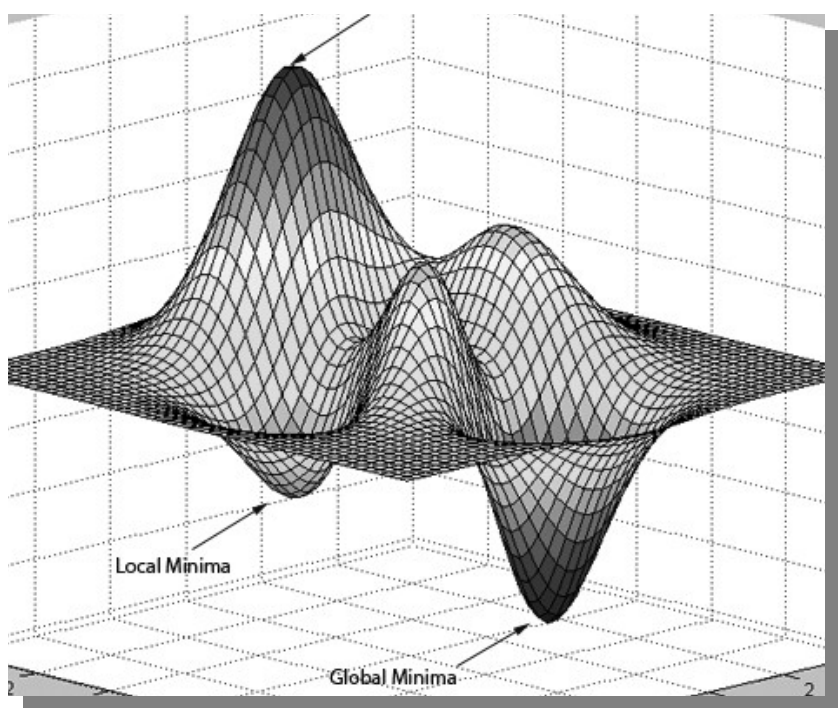
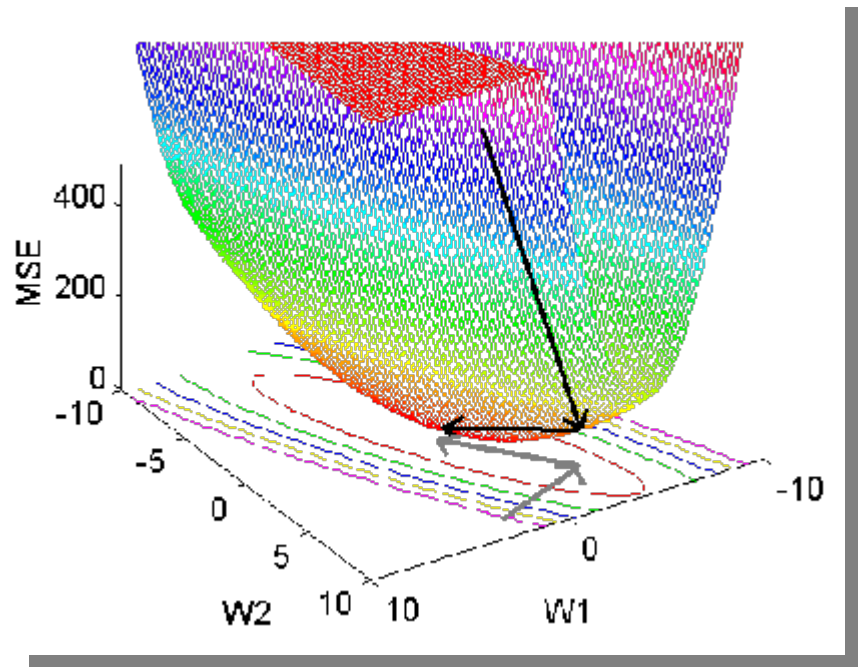
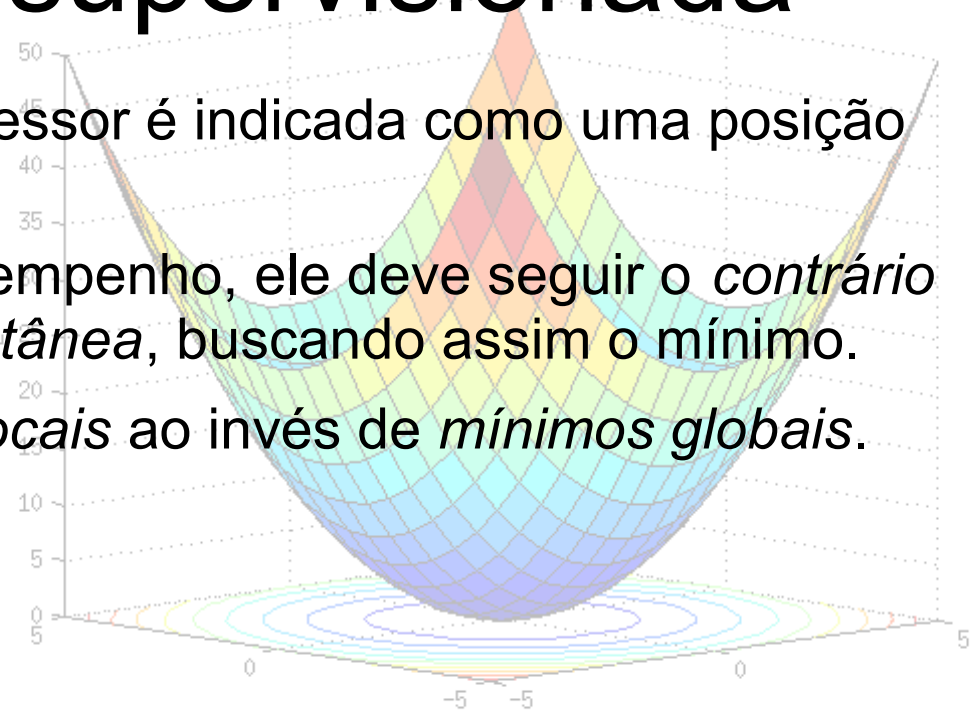
- Métodos determinísticos:
 - Para sistemas lineares;
 - Exemplos:
 - Mínimos Quadrados; ou
 - Utilização da Pseudo Inversa.
- Métodos iterativos:
 - Atuam sobre a correção de erro;
 - Exemplos:
 - Regra de Delta;
 - Descida do gradiente.

Aprendizagem supervisionada

- Para a medida de desempenho do sistema, pode-se utilizar:
 - Erro médio quadrático;
 - Soma dos erros quadrados.
- Cada uma pode ser definida como *função dos parâmetros livres do sistema*.
- Essa função pode ser visualizada como:
 - *Superfície multidimensional de desempenho de erro*; ou como
 - *Superfície de erro*.
- A verdadeira superfície de erro é obtida pela média sobre todos os exemplos possíveis de entrada e saída.

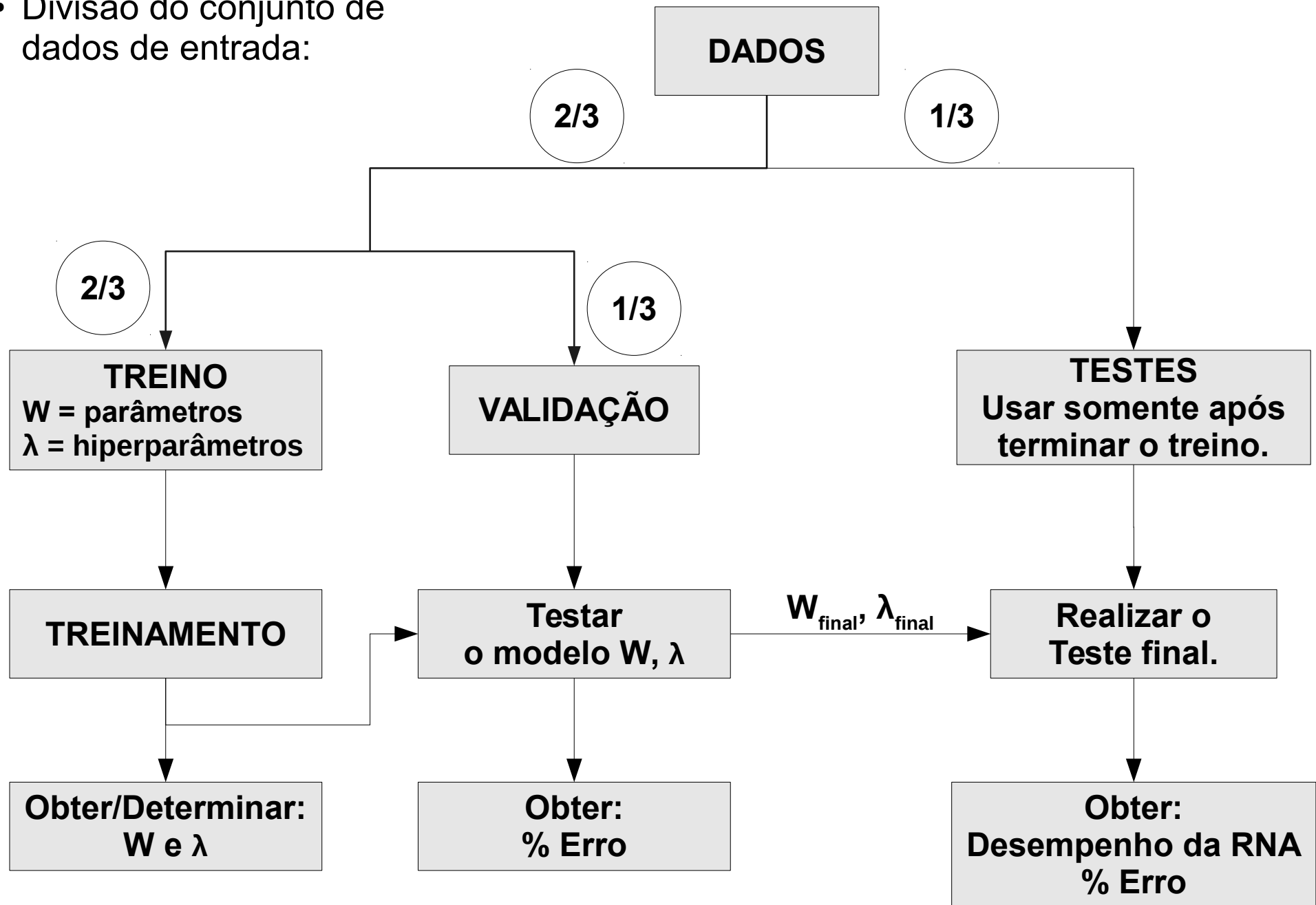
Aprendizagem supervisionada

- Cada operação realizada pelo professor é indicada como uma posição na superfície de erro.
- Para que o sistema melhore o desempenho, ele deve seguir o *contrário do gradiente* ou a *estimativa instantânea*, buscando assim o mínimo.
- O problema é encontrar *mínimos locais* ao invés de *mínimos globais*.



Aprendizagem supervisionada

- Divisão do conjunto de dados de entrada:



Aprendizagem supervisionada

- Estandarização dos dados (*standardization*):

- Se não fizer, a RNA pode não aprender;

Isso ficará mais claro ao trabalharmos com a descida do gradiente.

- Com os padrões de entrada, obtenha:

- μ = média (função *mean* do Octave);
- σ = desvio padrão (função *std* do Octave).

- Faça sobre todo o conjunto de treinamento \mathbf{X} :

$$X = \frac{X - \mu}{\sigma}$$

- Isso uniformizará os dados de entrada para que todos tenham uma variância única.

Revisão: *variância*?

- Variância:
 - Média da diferença quadrática entre \mathbf{X} e sua média.
 - Símbolo: σ^2
 - Variância de uma **população** com N elementos:

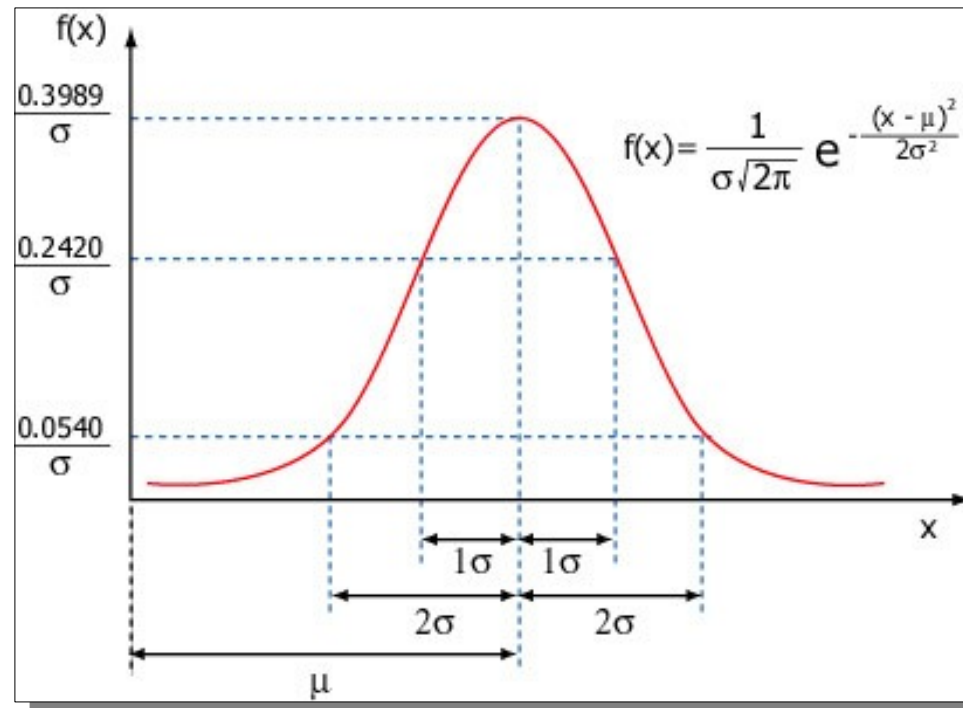
$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \frac{\sum_{i=1}^N x_i^2}{N} - \mu^2$$

- Onde:
 - μ : média;
 - $(x_i - \mu)$: distância de x_i até a média;
- A variância de uma **amostra** é obtida por:
- Octave: `var(x_i - mu, 1)`; e `var(x_i - mu)`;

$$s^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N - 1}$$

Revisão: *desvio padrão*?

- Desvio padrão:
 - Mede quão espalhados os dados de um conjunto estão.
 - É simplesmente a raiz da variância:
$$\sigma = \sqrt{\sigma^2}$$
 - Na função gaussiana:



Aprendizagem Não Supervisionada

- Também pode ser chamada de aprendizagem sem um professor.
- Assim, o conhecimento é representado por um conjunto de exemplos somente com as *entradas*.
- O objetivo da aprendizagem não supervisionada é descobrir estruturas entre os dados, fazendo uma aglomeração (*clusters*).
- Ela deve trabalhar sobre uma *medida fornecida* para se ajustar as regularidades estatísticas dos dados de entrada.
- Formas de aprender sem supervisão:
 - Utilização da regra de Hebb; e
 - Aprendizagem competitiva.