Redes Neurais Artificiais

AULA 04 – Redes ADALINE

Prof. Ivan Nunes da Silva

1. Rede ADALINE

Aspectos históricos do ADALINE

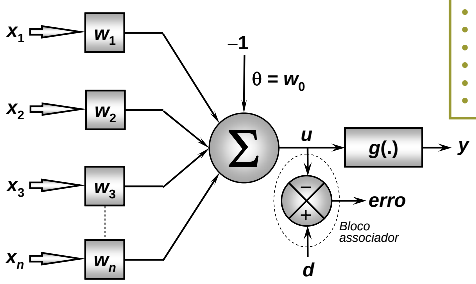
- O ***ADALINE*** (*Adaptive Linear Element*) foi idealizado por Widrow & Hoff em 1960.
- Sua principal aplicação estava em sistemas de chaveamento de circuitos telefônicos.
- Esta foi uma das primeiras aplicações industriais que envolveram efetivamente a utilização das redes neurais artificiais.
- Embora seja também um tipo de rede bem simples (um único neurônio), o ***ADALINE*** promoveu alguns outros avanços para a área de redes neurais, isto é:
  - Desenvolvimento do algoritmo de aprendizado **regra Delta**.
  - Aplicações em diversos problemas práticos envolvendo processamento de sinais analógicos.
  - Primeiras aplicações industriais de redes neurais artificiais.

2. Arquitetura do ADALINE

Aspectos Topológicos

- **Ilustração da rede *ADALINE*:**

- **Sinais de entrada** ➔ $\{x_1, x_2, \ldots, x_n\}$
- **Pesos sinápticos** ➔ $\{w_1, w_2, \ldots, w_n\}$
- **Combinador linear** ➔ $\{\Sigma\}$
- **Limiar de ativação** ➔ $\{\theta\}$
- **Potencial de ativação** ➔ $\{u\}$
- **Função de ativação** ➔ $\{g\}$
- **Sinal de saída** ➔ $\{y\}$
- **Valor desejado** ➔ $\{d\}$
- **Valor de erro** ➔ $\{erro\}$ entre $\{d\}$ e $\{u\}$

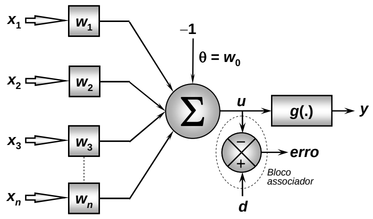


1) Rede ***ADALINE*** é constituída de *n* sinais de entrada e somente uma saída (composta de um único neurônio).
2) Assim como o ***Perceptron***, a rede ***ADALINE*** possui também arquitetura *feedforward* de camada única, pois o fluxo de informações é realizado sempre adiante.
3) Assim como o ***Perceptron***, a rede ***ADALINE***, devida ainda à sua simplicidade estrutural, tem sido também mais utilizada em problemas de **"Classificação de Padrões"**, envolvendo apenas duas classes distintas.

3. Princípio de Funcionamento

Resumo do funcionamento do ADALINE

- **Passos para obtenção da resposta (saída):**



$$\begin{cases} u = \sum_{i=1}^{n} w_i \cdot x_i - \theta = \sum_{i=0}^{n} w_i \cdot x_i \\ y = g(u) \end{cases}$$

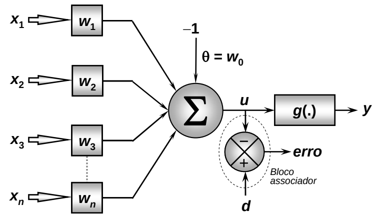
$$\boxed{erro = d - u}$$

1) Apresentação de um conjunto de valores que representam as variáveis de entrada do neurônio.
2) Multiplicação de cada entrada pelo seu respectivo peso sináptico.
3) Obtenção do potencial de ativação produzido pela soma ponderada dos sinais de entrada, subtraindo-se o limiar de ativação.
4) Aplicação de uma função de ativação apropriada, tendo-se como objetivo limitar a saída do neurônio.
5) Compilação da saída a partir da aplicação da função de ativação neural em relação ao seu potencial de ativação.
6) Presença de bloco associador, cuja função é produzir o sinal de "*erro*" entre "*d*" e "*u*" a fim de auxiliar no treinamento da rede.

3. Princípio de Funcionamento

Aplicabilidade em classificação de padrões

Aspectos de aplicabilidade:

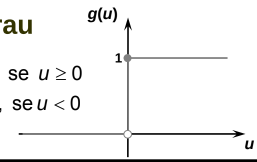


$$\begin{cases} u = \sum_{i=1}^{n} w_i \cdot x_i - \theta = \sum_{i=0}^{n} w_i \cdot x_i \\ y = g(u) \end{cases}$$

1) Assim como no ***Perceptron***, devido às suas características estruturais, as funções de ativação do ***ADALINE*** são a **"degrau"** ou **"degrau bipolar"**.
2) Assim, tem-se também apenas **"duas possibilidades"** de valores a serem produzidos pela sua saída, ou seja, valor 0 ou 1 (para a função de ativação Degrau), ou ainda, valor -1 ou 1 (para a função Degrau Bipolar).

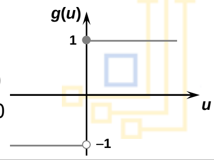
Degrau

$$g(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ 0, & \text{se } u < 0 \end{cases}$$



Degrau Bipolar

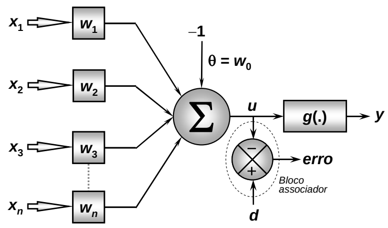
$$g(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ -1, & \text{se } u < 0 \end{cases}$$



3. Princípio de Funcionamento

Parâmetros característicos do ADALINE

Aspectos de aplicabilidade:



- **Sinais de entrada** → {x_1, x_2,..., x_n}
- **Pesos sinápticos** → {w_1, w_2,..., w_n}
- **Combinador linear** → {Σ}
- **Limiar de ativação** → {θ}
- **Potencial de ativação** → {*u*}
- **Função de ativação** → {*g*}
- **Sinal de saída** → {*y*}
- **Valor desejado** → {*d*}
- **Valor de erro** → {*erro*} entre {*d*} e {*u*}

Parâmetro	Variável Representativa	Tipo Característico
Entradas	x_i (*i*-ésima entrada)	Reais ou Binária (advindas externamente)
Pesos Sinápticos	w_i (associado a x_i)	Reais (iniciados aleatoriamente)
Limiar	θ	Real (iniciado aleatoriamente)
Saída	*y*	Binária
Função de Ativação	*g*(.)	Degrau ou Degrau Bipolar
Processo de Treinamento	-----	Supervisionado
Regra de Aprendizado	-----	Regra Delta

3. Princípio de Funcionamento

Processo de treinamento supervisionado

- **Aspectos do treinamento supervisionado:**

Parâmetro	Variável Representativa	Tipo Característico
Entradas	x_i (*i*-ésima entrada)	Reais ou Binária (advindas externamente)
Pesos Sinápticos	w_i (associado a x_i)	Reais (iniciados aleatoriamente)
Limiar	θ	Real (iniciado aleatoriamente)
Saída	***y***	**Binária**
Função de Ativação	***g*(.)**	**Degrau ou Degrau Bipolar**
Processo de Treinamento	**-----**	**Supervisionado**
Regra de Aprendizado	**-----**	**Regra Delta**

1) Assim como no ***Perceptron***, o ajuste dos pesos e limiar do ***ADALINE*** é efetuado utilizando processo de treinamento **"Supervisionado"**.
2) Então, para cada amostra dos sinais de entrada se tem a respectiva saída (resposta) desejada.
3) Como o ***ADALINE*** é tipicamente usado em problemas de classificação de padrões, a sua saída pode assumir somente dois valores possíveis, os quais se associam às **"duas classes"** que o mesmo estará identificando.
4) A diferença principal entre o ***ADALINE*** e o ***Perceptron*** está principalmente na regra de aprendizado utilizada para os ajustes dos pesos e limiar.

7

3. Princípio de Funcionamento

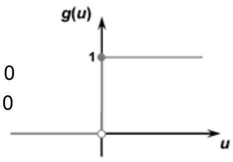
Mapeamento de problemas de classificação de padrões

- **Aspectos de aplicabilidade:**

1) Assim como no ***Perceptron***, para problemas de classificação dos sinais de entrada, tem-se então duas classes possíveis para a saída do ***ADALINE***, denominadas de ***Classe A*** e ***Classe B***.
2) Então, como se tem somente **"duas possibilidades"** de valores a serem produzidos na saída do ***ADALINE***, podem-se ter as seguintes associações para a sua saída { $y = g(u)$ }:

Degrau

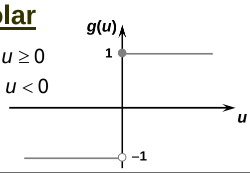
$$g(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ 0, & \text{se } u < 0 \end{cases}$$



$$g(u) = \begin{cases} 1, & \text{se } u \geq 0 \Rightarrow \boldsymbol{x} \in \textit{Classe A} \\ 0, & \text{se } u < 0 \Rightarrow \boldsymbol{x} \in \textit{Classe B} \end{cases}$$

Degrau Bipolar

$$g(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ -1, & \text{se } u < 0 \end{cases}$$



$$g(u) = \begin{cases} 1, & \text{se } u \geq 0 \Rightarrow \boldsymbol{x} \in \textit{Classe A} \\ -1, & \text{se } u < 0 \Rightarrow \boldsymbol{x} \in \textit{Classe B} \end{cases}$$

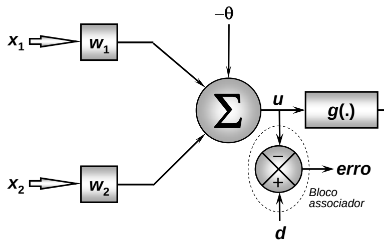
8

3. Princípio de Funcionamento

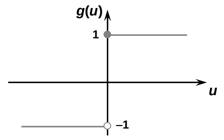
O quê acontece internamente no ADALINE?

● Análise matemática do *ADALINE*:

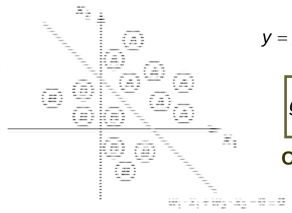
- Assumindo-se aqui um ***ADALINE*** com apenas duas entradas, tem-se:



$$\begin{cases} u = \sum_{i=1}^{n=2} w_i \cdot x_i - \theta \Rightarrow w_1 \cdot x_1 + w_2 \cdot x_2 - \theta \\ y = g(u) \end{cases}$$



- Usando a função de ativação sinal, a saída do ***ADALINE*** seria dada por:



$$y = g(u) = \begin{cases} 1, & \text{se } \sum w_i \cdot x_i - \theta \geq 0 \Leftrightarrow w_1 \cdot x_1 + w_2 \cdot x_2 - \theta \geq 0 \\ -1, & \text{se } \sum w_i \cdot x_i - \theta < 0 \Leftrightarrow w_1 \cdot x_1 + w_2 \cdot x_2 - \theta < 0 \end{cases}$$

$$\boxed{g(u) = \begin{cases} 1, & \text{se } u \geq 0 \Rightarrow \boldsymbol{x} \in \textbf{Classe A} \\ -1, & \text{se } u < 0 \Rightarrow \boldsymbol{x} \in \textbf{Classe B} \end{cases}}$$

Conclusão: Assim como o ***Perceptron***, o ***ADALINE*** é também um **classificador linear** que pode ser usado para classificar duas classes.

9

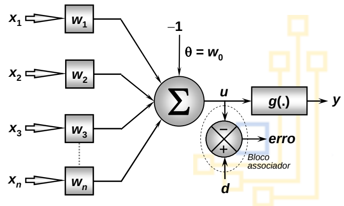
4. Processo de Treinamento

Princípio da regra Delta

● Aspectos do processo de treinamento:

- O processo de ajuste dos pesos e limiar do ***ADALINE*** é baseado no algoritmo de aprendizado denominado de regra **Delta**, também conhecida pelos seguintes nomes:
  - Regra de aprendizado de Widrow-Hoff.
  - Algoritmo *LMS (Least Mean Square)*.
  - Método do Gradiente Descendente.
- Assumindo-se a disponibilidade de ***p*** amostras de treinamento, o princípio envolvido com a aplicação da regra **Delta**, objetivando ajustar os pesos e limiar do neurônio, é a seguinte:
  - **Minimizar a diferença (*erro*)** entre a saída desejada ***d*** e a resposta do combinador ***u***, considerando-se para tanto todas as ***p*** amostras.
  - Mais especificamente, utiliza-se a minimização do **Erro Quadrático** entre ***u*** e ***d*** com o intuito de ajustar o vetor de pesos {***w***} da rede.

$$\boldsymbol{w} = [\theta \quad w_1 \quad w_2 \; \ldots \; w_n]^T$$



10

4. Processo de Treinamento

Derivação do algoritmo regra Delta (I)

- **Objetivo do algoritmo:**
  - Consiste de obter um $\boldsymbol{w}^*$ ótimo tal que o **Erro Quadrático** $\{E(\boldsymbol{w}^*)\}$ sobre todo o conjunto de amostras seja o mínimo possível, isto é:

$$E(\boldsymbol{w}^*) \le E(\boldsymbol{w}), \quad \text{para } \forall \boldsymbol{w} \in \Re^{n+1}$$

  - A função **Erro Quadrático** em relação às p amostras de treinamento é definida por:

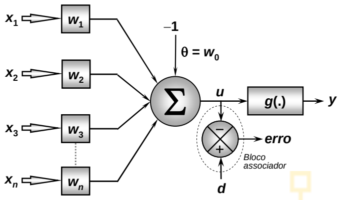
$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{k=1}^{p}(d^{(k)} - u)^2$$

  - Usando o valor de u na expressão acima, obtém-se:

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{k=1}^{p}\left(d^{(k)} - \left(\sum_{i=1}^{n} w_i \cdot x_i^{(k)} - \theta\right)\right)^2$$

  - Em forma vetorial, tem-se:

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{k=1}^{p}(d^{(k)} - (\boldsymbol{w}^T \cdot \boldsymbol{x}^{(k)} - \theta))^2$$



$$\begin{cases} u = \sum_{i=1}^{n} w_i \cdot x_i^{(k)} - \theta \\ y = g(u) \end{cases}$$

Conclusão: Esta expressão acima totaliza o erro quadrático contabilizando-se as p amostras de treinamento.

4. Processo de Treinamento

Derivação do algoritmo regra Delta (II)

- **Aplicação do operador Gradiente:**
  - Consiste de aplicar o operador gradiente em relação ao vetor $\boldsymbol{w}$, tendo-se como objetivo a busca do valor ótimo para o erro quadrático, isto é:

$$\nabla E(\boldsymbol{w}) = \frac{\partial E(\boldsymbol{w})}{\partial \boldsymbol{w}}$$

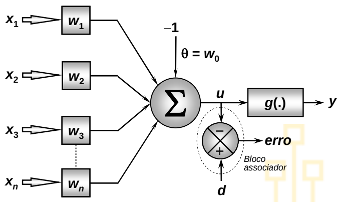
$$\boxed{E(\boldsymbol{w}) = \frac{1}{2}\sum_{k=1}^{p}(d^{(k)} - (\boldsymbol{w}^T \cdot \boldsymbol{x}^{(k)} - \theta))^2}$$

  - Considerando o valor de $E(\boldsymbol{w})$ na expressão acima, obtém-se:

$$\nabla E(\boldsymbol{w}) = \sum_{k=1}^{p}(d^{(k)} - \underbrace{(\boldsymbol{w}^T \cdot \boldsymbol{x}^{(k)} - \theta)}_{u}) \cdot (-\boldsymbol{x}^{(k)})$$

  - Retornando o valor de u na expressão acima, tem-se:

$$\nabla E(\boldsymbol{w}) = -\sum_{k=1}^{p}(d^{(k)} - u) \cdot \boldsymbol{x}^{(k)}$$



$$\begin{cases} u = \sum_{i=1}^{n} w_i \cdot x_i^{(k)} - \theta \\ y = g(u) \end{cases}$$

Conclusão: Esta expressão acima fornece o valor do gradiente do erro em relação a todas as amostras de treinamento.

4. Processo de Treinamento

Derivação do algoritmo regra Delta (III)

Ajuste do vetor de pesos:

- Agora, deve-se associar o valor do gradiente ao procedimentos de ajustes do vetor de pesos $\{\boldsymbol{w}\}$.
- Assim, a adaptação do vetor de pesos devem ser efetuado na direção oposta àquela do gradiente, pois o objetivo da otimização é minimizar o erro quadrático médio entre ***d*** e ***u***.
- Nesta condição, a variação $\Delta\boldsymbol{w}$ a ser efetivada no vetor de pesos do ***ADALINE*** é dada por:

$$\Delta\boldsymbol{w} = -\eta \cdot \nabla E(\boldsymbol{w})$$

$$\boxed{\nabla E(\boldsymbol{w}) = -\sum_{k=1}^{p}(d^{(k)} - u)\cdot \boldsymbol{x}^{(k)}}$$

- Inserindo o resultado de $\nabla E(\boldsymbol{w})$ na expressão acima, tem-se:

$$\Delta\boldsymbol{w} = \eta \sum_{k=1}^{p}(d^{(k)} - u)\cdot \boldsymbol{x}^{(k)}$$

- De forma complementar, pode-se exprimir $\Delta\boldsymbol{w}$ por:

$$\boldsymbol{w}^{atual} = \boldsymbol{w}^{anterior} + \eta \sum_{k=1}^{p}(d^{(k)} - u)\cdot \boldsymbol{x}^{(k)}$$

4. Processo de Treinamento

Derivação do algoritmo regra Delta (IV)

Simplificação algorítmica:

- Por razões de simplificação computacional, a atualização de ***w*** pode ser também realizada de forma discreta.
- Neste caso, um passo de ajuste em ***w*** é realizado após a apresentação de cada *k*-ésima amostra de treinamento, ou seja:

$$\boldsymbol{w}^{atual} = \boldsymbol{w}^{anterior} + \eta \sum_{k=1}^{p}(d^{(k)} - u)\cdot \boldsymbol{x}^{(k)}$$

$$\boldsymbol{w}^{atual} = \boldsymbol{w}^{anterior} + \eta \cdot (d^{(k)} - u)\cdot \boldsymbol{x}^{(k)}, \quad k = 1 \ldots p$$

- Em notação algorítmica, tem-se:

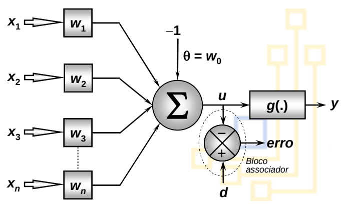
$$\boxed{\boldsymbol{w} \leftarrow \boldsymbol{w} + \eta \cdot (d^{(k)} - u)\cdot \boldsymbol{x}^{(k)}, \quad k = 1 \ldots p}$$

onde: $\boldsymbol{w} = [\theta \;\; w_1 \;\; w_2 \; \ldots \; w_n]^T$

$\boldsymbol{x}^{(k)} = [-1 \;\; x_1^{(k)} \;\; x_2^{(k)} \ldots \; x_n^{(k)}]^T$

$d^{(k)}$ é o valor desejado para *k*-ésima amostra de treinamento.

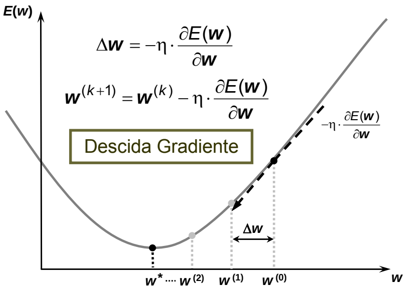
η é a taxa de aprendizagem $\{0 < \eta < 1\}$.



4. Processo de Treinamento

Derivação do algoritmo regra Delta (V)

- **Interpretação geométrica (processo de convergência do *ADALINE*):**
  - A interpretação geométrica frente aos passos de atualização der $\boldsymbol{w}$, rumo ao ponto de minimização $\boldsymbol{w}^*$ da função erro quadrático $\{E(\boldsymbol{w})\}$, é dada pela seguinte ilustração:
    1) Parte-se de um valor inicial $\boldsymbol{w}^{(0)}$;
    2) O próximo valor de $\boldsymbol{w}$ (dado por $\boldsymbol{w}^{(1)}$) será obtido considerando-se a direção oposta ao vetor gradiente em relação a $\boldsymbol{w}^{(0)}$;
    3) Para o próximo passo de atualização, o ajuste de $\boldsymbol{w}$ (representado agora por $\boldsymbol{w}^{(2)}$) será realizado considerando-se o valor do gradiente em relação à $\boldsymbol{w}^{(1)}$;



4) Aplicando-se tais passos sucessivamente, haverá a convergência em direção a $\boldsymbol{w}^*$;
5) Este valor de $\boldsymbol{w}^*$ será a configuração ótima para os parâmetros livres do ***ADALINE***;
6) O valor de $E(\boldsymbol{w}^*)$ será então sempre menor que quaisquer $E(\boldsymbol{w})$ calculados nos passos anteriores.

15

4. Processo de Treinamento

Derivação do algoritmo regra Delta (VI)

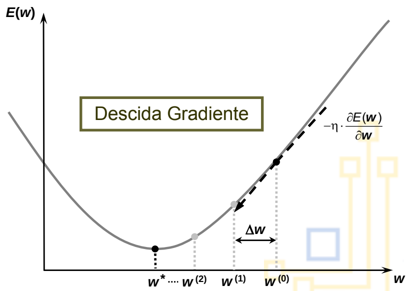
- **Critério de parada do algoritmo:**
  - Com a aplicação dos passos iterativos anteriores, o vetor $\boldsymbol{w}$ caminhará em direção ao seu valor ótimo $\{\boldsymbol{w}^*\}$.
  - Para se detectar o alcance de $\boldsymbol{w}^*$, a fim de cessar o processo de convergência, há a necessidade de se estabelecer um critério de parada.
  - Para tanto, utiliza-se o valor do erro quadrático médio $\{E_{qm}(\boldsymbol{w})\}$ em relação a todas as amostras de treinamento, sendo este definido por:

$$E_{qm}(\boldsymbol{w}) = \frac{1}{p}\sum_{k=1}^{p}(d^{(k)} - u)^2$$

  - O algoritmo converge quando o erro quadrático médio entre duas épocas sucessivas for bem pequeno, isto é:

$$\left| E_{qm}(\boldsymbol{w}^{atual}) - E_{qm}(\boldsymbol{w}^{anterior}) \right| \leq \varepsilon$$

onde ε é a precisão requerida para o processo de convergência



16

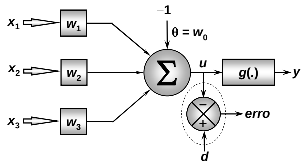
5. Implementação Computacional

Aspectos de preparação de dados

Montagem de conjuntos de treinamento:

- Supõe-se que um problema a ser mapeado pelo ***ADALINE*** tenha três entradas $\{x_1, x_2, x_3\}$, conforme a figura ao lado (abaixo).
- Assume-se que se tem quatro amostras, constituída dos seguintes valores de entrada:

Amostra 1 ➔ Entrada: [0,1 0,4 0,7] → Saída desejada: [1]
Amostra 2 ➔ Entrada: [0,3 0,7 0,2] → Saída desejada: [-1]
Amostra 3 ➔ Entrada: [0,6 0,9 0,8] → Saída desejada: [-1]
Amostra 4 ➔ Entrada: [0,5 0,7 0,1] → Saída desejada: [1]



- Então, de forma similar ao ***Perceptron***, pode-se converter tais sinais para que estes possam ser usados no treinamento do ***ADALINE***:

Conjunto de treinamento

$x^{(1)} = [-1 \;\; 0{,}1 \;\; 0{,}4 \;\; 0{,}7]^T \rightarrow$ com $d^{(1)} = 1$
$x^{(2)} = [-1 \;\; 0{,}3 \;\; 0{,}7 \;\; 0{,}2]^T \rightarrow$ com $d^{(2)} = -1$
$x^{(3)} = [-1 \;\; 0{,}6 \;\; 0{,}9 \;\; 0{,}8]^T \rightarrow$ com $d^{(3)} = -1$
$x^{(4)} = [-1 \;\; 0{,}5 \;\; 0{,}7 \;\; 0{,}1]^T \rightarrow$ com $d^{(4)} = 1$

forma matricial

$$\Omega^{(x)} = \begin{array}{c} \\ x_0 \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{cccc} x^{(1)} & x^{(2)} & x^{(3)} & x^{(4)} \\ \hline -1 & -1 & -1 & -1 \\ 0{,}1 & 0{,}3 & 0{,}6 & 0{,}5 \\ 0{,}4 & 0{,}7 & 0{,}9 & 0{,}7 \\ 0{,}7 & 0{,}2 & 0{,}8 & 0{,}1 \end{array} \;;$$

$$\Omega^{(d)} = \begin{array}{cccc} d^{(1)} & d^{(2)} & d^{(3)} & d^{(4)} \\ \hline 1 & -1 & -1 & 1 \end{array}$$

- Geralmente, as amostras de treinamento são disponibilizadas em sua forma matricial (por meio de arquivo texto ou planilha).

5. Implementação Computacional

Algoritmo de aprendizagem (fase de treinamento)

Pseudocódigo para fase de treinamento:

Início {Algoritmo *Adaline* – Fase de Treinamento}

<1> Obter o conjunto de amostras de treinamento $\{\boldsymbol{x}^{(k)}\}$;
<2> Associar a saída desejada $\{d^{(k)}\}$ para cada amostra obtida;
<3> Iniciar o vetor $\boldsymbol{w}$ com valores aleatórios pequenos;
<4> Especificar taxa de aprendizagem $\{\eta\}$ e precisão requerida $\{\varepsilon\}$;
<5> Iniciar o contador de número de épocas $\{$ *época* $\leftarrow 0\}$;
<6> Repetir as instruções:
<6.1> $E_{qm}^{anterior} \leftarrow E_{qm}(\boldsymbol{w})$;
<6.2> Para todas as amostras de treinamento $\{\boldsymbol{x}^{(k)}, d^{(k)}\}$, fazer:
<6.2.1> $u \leftarrow \boldsymbol{w}^T \cdot \boldsymbol{x}^{(k)}$;
<6.2.2> $\boldsymbol{w} \leftarrow \boldsymbol{w} + \eta \cdot (d^{(k)} - u) \cdot \boldsymbol{x}^{(k)}$;
<6.3> *época* $\leftarrow$ *época* + 1;
<6.4> $E_{qm}^{atual} \leftarrow E_{qm}(\boldsymbol{w})$;
Até que: $\left| E_{qm}^{atual} - E_{qm}^{anterior} \right| \leq \varepsilon$

Fim {Algoritmo *Adaline* – Fase de Treinamento}

Critério de parada ➔ O algoritmo converge quando o erro quadrático médio $\{E_{qm}\}$ entre duas épocas sucessivas for menor ou igual à precisão $\{\varepsilon\}$ requerida ao problema mapeado pelo ***ADALINE***.

5. Implementação Computacional

Algoritmo de aprendizagem (fase de operação)

Pseudocódigo para fase de operação:

Início {Algoritmo *Adaline* – Fase de Operação}

<1> Obter uma amostra a ser classificada { $\boldsymbol{x}$ };
<2> Utilizar o vetor $\boldsymbol{w}$ ajustado durante o treinamento;
<3> Executar as seguintes instruções:
<3.1> $u \leftarrow \boldsymbol{w}^T \cdot \boldsymbol{x}$;
<3.2> $y \leftarrow \text{sinal}(u)$;
<3.3> Se $y = -1$
<3.3.1> Então: amostra $\boldsymbol{x} \in$ *{Classe A}*
<3.4> Se $y = 1$
<3.4.1> Então: amostra $\boldsymbol{x} \in$ *{Classe B}*

Fim {Algoritmo *Adaline* – Fase de Operação}

Obs. 1 ➔ A "Fase de Operação" é usada somente após a fase de treinamento, pois aqui a rede já está apta para ser usada no processo.

Obs. 2 ➔ A "Fase de Operação" é então utilizada para realizar a tarefa de classificação de padrões frente às novas amostras que serão apresentadas à rede.

Obs. 3 ➔ Lembrar de incluir o valor -1 dentro do vetor $\boldsymbol{x}$, pois será multiplicado por $\boldsymbol{w}$.

$$\boldsymbol{x} = [-1 \quad x_1 \quad x_2 \dots x_n]^T$$
$$\boldsymbol{w} = [\theta \quad w_1 \quad w_2 \dots w_n]^T$$

19

5. Implementação Computacional

Algoritmo de aprendizagem (erro quadrático médio)

Pseudocódigo para o erro quadrático médio:

Início {Algoritmo EQM}

<1> Obter a quantidade de padrões de treinamento { p };
<2> Iniciar a variável E_{qm} com valor zero { $E_{qm} \leftarrow 0$ };
<3> Para todas as amostras de treinamento { $\boldsymbol{x}^{(k)}$, $d^{(k)}$ }, fazer:
<3.1> $u \leftarrow \boldsymbol{w}^T \cdot \boldsymbol{x}^{(k)}$;
<3.2> $E_{qm} \leftarrow E_{qm} + (d^{(k)} - u)^2$;
<4> $E_{qm} \leftarrow \dfrac{E_{qm}}{p}$;

Fim {Algoritmo EQM}

Obs. 1 ➔ O algoritmo do erro quadrático médio deve ser implementado por meio de uma sub-rotina (função).

Obs. 2 ➔ Todas as amostras de treinamento devem ser contabilizadas no cálculo do valor do erro quadrático médio.

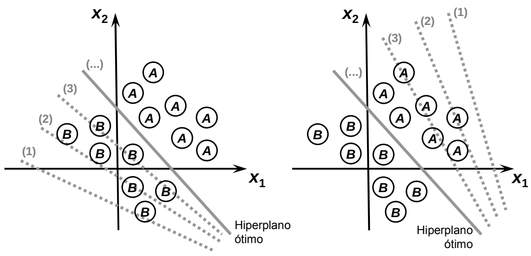
20

5. Implementação Computacional

Aspectos de convergência (I)

Ilustração do processo de convergência:

- Processo de treinamento do ***ADALINE*** tende a mover continuamente o seu vetor de pesos, tendo-se o intuito de minimizar o erro quadrático em relação a todas as amostras disponíveis para o aprendizado.
- As figuras mostram duas situações que ilustram a convergência do ***ADALINE*** rumo à estabilização, considerando-se para fins didáticos apenas duas entradas {x_1 e x_2}.



- Observa-se que nestas duas situações, cujos vetores de pesos iniciais $\boldsymbol{w}^{(0)}$ foram iniciados em posições distintas, o processo de convergência encaminha o hiperplano sempre para o mesmo local.
- Nesses casos, o valor correspondente a $\boldsymbol{w}^*$ é sempre aquele que minimiza a função erro quadrático.

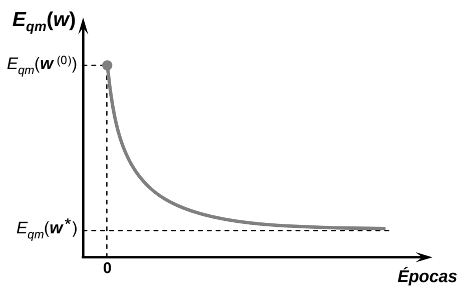
21

5. Implementação Computacional

Aspectos de convergência (II)

Comportamento da função erro quadrático:

- Para fins de interpretação, a figura mostra o comportamento do erro quadrático médio (E_{qm}) em função do número de épocas de treinamento.



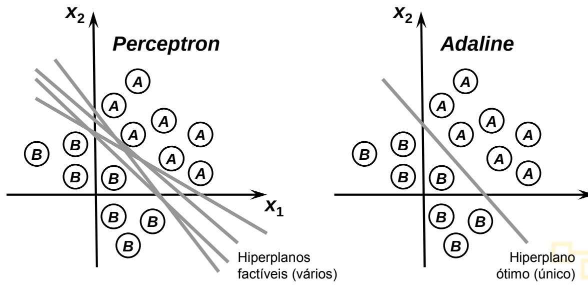
- Constata-se então que a curva do erro quadrático médio para o ***ADALINE*** é sempre descendente.
- Esta decresce na medida em que as épocas de treinamento vão sendo executadas.
- Finalmente, estabiliza-se num valor constante quando o ponto de mínimo da função erro quadrático médio for alcançado

22

6. Aspectos Comparativos

Comparação entre ADALINE e Perceptron (I)

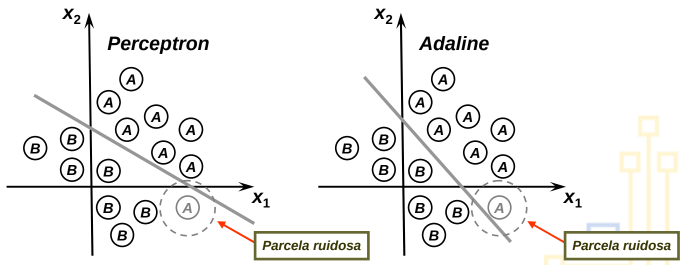
- **Diferenças entre treinamento do *ADALINE* e *Perceptron*:**
  - ***ADALINE*** ➔ Treinamento usando Regra Delta (Minimização do erro em relação a todas as amostras de treinamento).
    - Independentemente dos valores iniciais atribuídos a ***w***, o hiperplano de separabilidade (após convergência) sempre será o mesmo.
  - ***Perceptron*** ➔ Treinamento usando Regra de Hebb (Avaliação das respostas produzidas após apresentação de cada amostra de treinamento).
    - Quaisquer hiperplanos posicionados dentro da faixa de separabilidade entre as classes são considerados soluções factíveis.



6. Aspectos Comparativos

Comparação entre ADALINE e Perceptron (II)

- **Aspectos de robustez (fase de operação):**
  - ***ADALINE*** ➔ A inclinação do hiperplano é ajustada por intermédio do método dos mínimos quadrados dos erros (processo otimizado).
    - **Maior tendência** de imunidade a eventuais ruídos que podem afetar o processo em que a mesma está mapeando.
  - ***Perceptron*** ➔ O hiperplano que separa as classes pode ter infinitas disposições, pois sua configuração final é fortemente do valores inicial de ***w***.
    - **Menor tendência** de imunidade a eventuais ruídos que podem afetar o processo em que a mesma está mapeando.



  - ***ADALINE*** ➔ probabilidade menor de classificar a amostra incorretamente.
  - ***Perceptron*** ➔ probabilidade maior de classificar a amostra incorretamente.

6. Questões do ADALINE

Reflexões, observações e aspectos práticos

1) Considerando-se que um problema a ser mapeado pelo ***ADALINE*** não seja linearmente separável, explique então se para esta situação o processo de treinamento (por meio do algoritmo regra Delta) também convergirá. ***{Exercício 1}***

2) Explique por que o treinamento do *ADALINE* se processa normalmente de forma mais rápido que aquele do ***Perceptron***. ***{Exercício 2}***

3) Discorra se a afirmação seguinte é verdadeira ou falsa. Independentemente dos valores iniciais assumidos para o vetor de pesos do ***ADALINE***, uma mesma configuração final para ***w**** será sempre obtida após a sua convergência. ***{Exercício 8}***

4) Explique, considerando a questão anterior, se o número de épocas de treinamento será também igual, independentemente do seu vetor de pesos iniciais. ***{Exercício 9}***

5) Reflexões sobre os exercícios 7 e 10 do livro.

Fim da Apresentação

Núcleo de Pesquisa em
AUTOMAÇÃO INTELIGENTE
de Processos e Sistemas