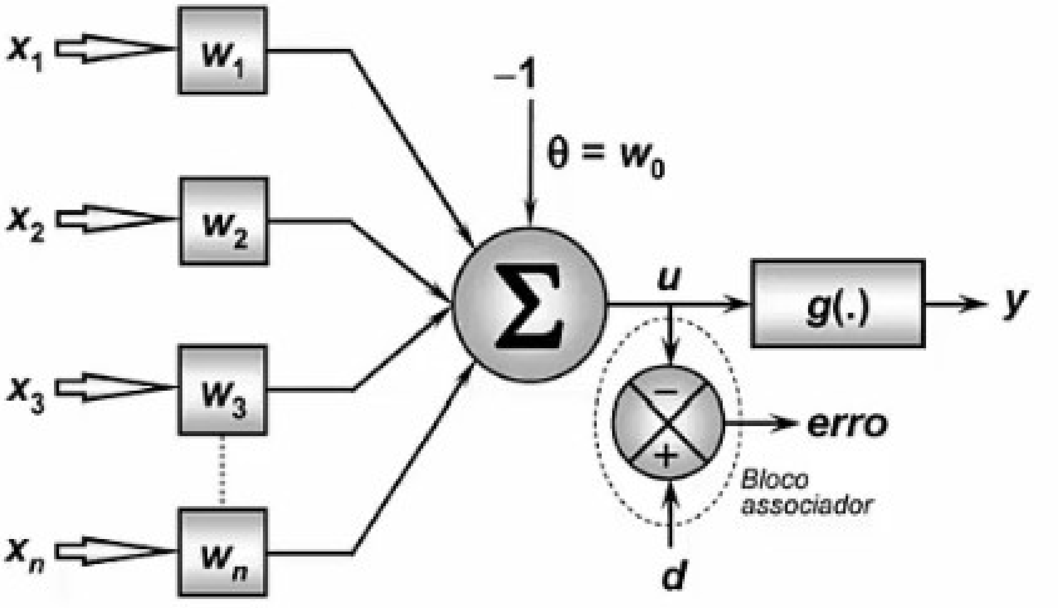
Rede Adaline

Adaline

- A rede Adaline (Adaptive Linear Element) proposta por Widrow e Hoff (1960) tem a mesma estrutura do Perceptron, diferenciando apenas no algoritmo de treinamento.

- Características da Adaline
  - Rede adaptativa
  - Inclusão de um algoritmo pioneiro para treinamento de redes de múltiplas camadas, o algoritmo de treinamento da Regra Delta
  - O algoritmo da Regra Delta é baseado no método dos mínimos quadrados.
  - Método de aprendizado mais suave

Rede Adaline



n sinais de entrada (x1...xn) e uma única saída (y).

- O processamento se dá pelas seguinte expressões:

$$u = \sum_{i=1}^{n} w_i \cdot x_i - \theta \Leftrightarrow u = \sum_{i=0}^{n} w_i \cdot x_i$$

$$y = g(u)$$

- Onde:
- x_i são os sinais de entrada,
- w_i são os pesos sinápticos associados,
- θ é o bias (limiar de ativação),
- g(.) é a função de ativação (degrau ou degrau bipolar) e
- u é o potencial de ativação.

- O bloco associador demonstrado no diagrama acima tem uma função auxiliar no processo de treinamento da rede.
-
- erro = d-u
-
- Erro é igual ao valor esperado menos o valor do potencial de ativação produzido pela rede.

Regra Delta

- O algoritmo de treinamento do Adaline é baseado no método dos mínimos quadrados, conhecido como Regra Delta.

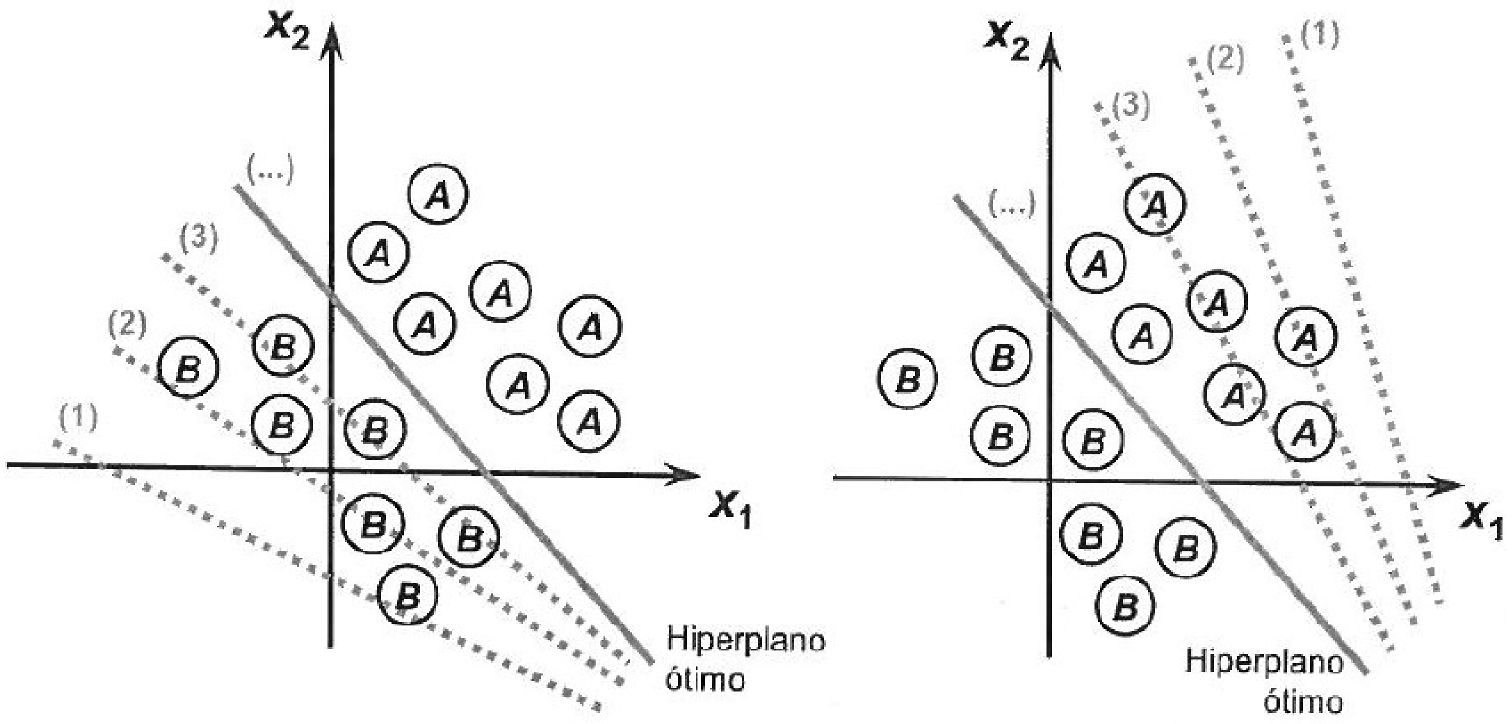
- O algoritmo utiliza o método do gradiente descendente com a intenção de diminuir o valor da função de erro possibilitando assim a convergência para um mínimo da função de erro
- Objetivo: Realizar iterações locais para obter o ponto mínimo da função de erro e, assim, fixar os valores dos pesos quando o mínimo foi encontrado
- Funcionamento: Através de um ponto arbitrário (vetor de pesos iniciado aleatoriamente), o algoritmo percorre, a cada iteração a superfície da função de erro em direção ao ponto de mínimo.

- Então, temos:

$$\mathbf{w} \leftarrow \mathbf{w} + \eta \cdot (d^{(k)} - u) \cdot \mathbf{x}^{(k)}$$

- onde: $\mathbf{w} = [\theta \quad w_1 \quad w_2 \ \dots \ w_n]^T$ é o vetor contendo o limiar e os pesos;
- $\mathbf{x}^{(k)} = [-1 \quad x_1^{(k)} \quad x_2^{(k)} \dots \ x_n^{(k)}]^T$ é a k-ésima amostra de treinamento;
- $d^{(k)}$ é o valor desejado para a k-ésima amostra de treinamento;
- u é o valor de saída do combinador linear; e
- η é uma constante que define a taxa de aprendizagem da rede.
- Dessa forma o processo de treinamento da rede Adaline procura mover continuamente o vetor de pesos buscando minimizar o erro quadrático em relação as amostras de dados disponíveis para aprendizado.

Situação de convergência



- No Perceptron o hiperplano de separação das classes pode ter “infinitas disposições” pois a configuração final vetor de pesos (w) é dependente dos valores iniciais aleatoriamente atribuídos.

- Na Adaline, a inclinação do hiperplano é ajustada por intermédio do método dos mínimos erros quadrados (LMS — least mean square), ou seja, independentemente dos valores inicialmente atribuídos ao vetor de pesos ocorrerá a busca pelo ótimo.