

Adaptação local das taxas de aprendizado

Heurísticas para taxas de aprendizado individuais por peso de uma rede neural
Algoritmo delta-bar-delta e Silva-Almeida

Adriano Martins Moutinho

Redes Neurais I – Mestrado NCE 2003

Prof.: Antonio Carlos Thomé

Algoritmo delta-bar-delta

- 1) Cada peso tem sua própria taxa de aprendizado:

$$\Delta w_{ij}^k = -\eta_{ij}^k \cdot \partial E / \partial w_{ij}$$

Com a adaptação local da taxa de aprendizado, acredita-se possibilitar uma mais rápida convergência

Algoritmo delta-bar-delta (cont.)

2) A taxa de aprendizado é adaptada durante o treinamento:

- Quando o gradiente tem o mesmo sinal entre as épocas, a taxa de aprendizado é aumentada em uma constante.
- Quando o gradiente muda de sinal, a taxa de aprendizado sofre um decremento exponencial.

Algoritmo delta-bar-delta (cont.)

- A mudança na taxa de aprendizado é dada por, onde a é uma constante e b uma exponencial:

$$\eta_{ij}^k = \begin{cases} \eta_{ij}^{k-1} + a & \text{mesmo sinal no gradiente} \\ b\eta_{ij}^{k-1} & \text{mudança de sinal} \\ 0 & \text{caso contrário} \end{cases}$$

Algoritmo delta-bar-delta (cont.)

- Ao gradiente ainda se soma um termo de momento:

$$\delta_{ij} = \frac{\partial E}{\partial w_{ij}}$$

$$\delta_{ij}(k) \leftarrow (1 - \mathcal{G}) \cdot \delta_{ij}(k) + \mathcal{G} \cdot \delta_{ij}(k-1)$$

- O termo de momento, assim como a e b são ajustados pelo usuário. Valores recomendados:

$$10^{-4} \leq a \leq 0,1$$

$$0,5 \leq b \leq 0,9$$

$$0,1 \leq \mathcal{G} \leq 0,7$$

Back-propagation em lote com taxa de aprendizado local (Silva e Almeida)

- Neste algoritmo, também há uma taxa de aprendizado para cada peso.

$$\Delta w_{ij}^k(k) = -\eta_{ij}^k \cdot \left[\frac{\partial E}{\partial w_{ij}} + \beta \Delta w_{ij}(k-1) \right]$$

Back-prop. (Silva e Almeida) (cont.)

- A taxa de aprendizado é adaptativa conforme a equação abaixo:

$$\eta_{ij}^k = \begin{cases} a\eta_{ij}^{k-1} & \text{não há mudança} \\ b\eta_{ij}^{k-1} & \text{mudança de sinal} \end{cases}$$

- Neste caso, tanto a taxa de crescimento quanto a decrescimento é exponencial.

Back-prop. (Silva e Almeida) (cont.)

- Valores recomendados:

$$\left. \begin{array}{l} 1,1 \leq a \leq 1,3 \\ 0,75 \leq b \leq 0,9 \end{array} \right\} a \cong b^{-1}$$

$$\eta_{ij}^{(0)} = 10^{-3}, \beta = 0.1$$