

Informationsverarbeitung in neuronalen Netzwerken

(Zusammenfassung von wichtigen Konzepten und Formeln, die man an der Prüfung wissen sollte)

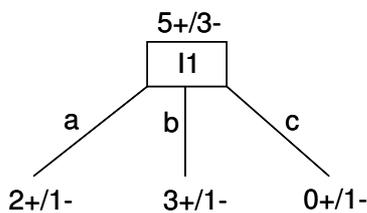
▪ ID3-Verfahren

- Entropie: $E = -p_+ \log p_+ - p_- \log p_-$ (p_+ = Anzahl +-Klassifizierungen, $p_- = 1 - p_+$)

- ID3: Maximale Entropie-Abnahme:

$$\Delta E = E_0 - \langle E_1 \rangle = \max$$

- Beispiel:



$$E_0 = -\frac{5}{8} \log \frac{5}{8} - \frac{3}{8} \log \frac{3}{8}$$

$$\langle E_1^I \rangle = \frac{3}{8} \left(-\frac{2}{3} \log \frac{2}{3} - \frac{1}{3} \log \frac{1}{3} \right) + \frac{4}{8} \left(-\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} \right) + 0$$

$$\rightarrow \Delta E^I = E_0 - \langle E_1^I \rangle$$

▪ Deterministische Neuronen

$$- S_i = f \left(\sum_j W_{ij} S_j \right)$$

- Aktivierungsfunktionen: $f(x) = \text{sign}(x) \rightarrow f'(x)$ nicht definiert

$$f(x) = x \rightarrow f'(x) = 1$$

$$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \rightarrow f'(x) = \frac{1}{2} [1 - f(x)^2]$$

$$f(x) = \frac{1}{1 + e^{-x}} \rightarrow f'(x) = f(x)[1 - f(x)]$$

▪ Perceptrons

$$- O = \text{sign} \left(\sum_{i=0}^d W_i I_i \right), \quad I_0 \equiv 1$$

- Kapazität: $N_c = 2(d+1) \rightarrow$ If $N < N_c$, dann fast sicher linear separabel
If $N > N_c$, dann fast sicher nicht linear separabel

- Perceptron-Lernverfahren: Lernbeispiele $\{I_i^y\} / D^y$

$$\rightarrow \Delta W_i^y = \eta (D^y - O^y) I_i^y, \quad O^y = \text{sign} \left(\sum_i W_i I_i^y \right)$$

▪ Gradientenverfahren für Feedforward-Netzwerke (Error-Backpropagation)

- Fehlerfunktion: $F = \sum_v F^v$, $F^v =$ Mass für Outputfehler bei Beispiel v

$$\rightarrow \Delta W_{ij} = -\eta \frac{\partial F}{\partial W_{ij}} \quad (\text{Batch})$$

$$\rightarrow \Delta W_{ij} = -\eta \frac{\partial F^v}{\partial W_{ij}} \quad (\text{Example-by-Example})$$

- Beispiel: $F = \frac{1}{2} (D_i - O_i)^2$

$$O_i = f\left(\sum_j W_{ij} S_j\right), \quad f(x) = \frac{1}{1 + e^{-x}}$$

$$\rightarrow \Delta W_{ij} = -\eta \frac{\partial F}{\partial W_{ij}} = -\eta \frac{\partial F}{\partial O_i} \frac{\partial O_i}{\partial W_{ij}} = \eta (D_i - O_i) f'\left(\sum_j W_{ij} S_j\right) \cdot S_j$$

$$\rightarrow \Delta W_{ij} = \eta (D_i - O_i) \cdot O_i \cdot (1 - O_i) \cdot S_j$$

▪ Hopfield-Netzwerke

- Dynamik: $S_i(t+1) = \text{sign}\left(\sum_j W_{ij} S_j(t)\right)$, $W_{ji} = W_{ij}$, $t =$ Iterationsschritt

- Stabilität: $S_i \cdot \sum_{j=1}^N W_{ij} S_j > 0$, $i = 1, \dots, N$, $N =$ Anzahl Neuronen

- Energiefunktion: $E = -\sum_{i < j} W_{ij} S_i S_j$

- Hebb-Regel: $W_{ij} = \sum_{v=1}^M S_i^v S_j^v$, $M =$ Anzahl Muster

- Kapazität: $\alpha_c = \frac{M_{\max}}{N} \approx 0.15$

▪ Winner-Take-All Netzwerke, Kompetitives Lernen

- Winner-Neuron i^* : $d_{i^*} = \min_i d_i$, $d_i = \sum_j (X_j - W_{ij})^2$, $X_j =$ Inputs

- Einfaches kompetitives Lernen:

$$W_{i^*j} = W_{i^*j} + \eta(t) \cdot (X_j - W_{i^*j})$$

$$W_{ij} = W_{ij} \quad \text{für } i \neq i^*$$

$\eta(t)$ abnehmend

▪ Stochastische Neuronen

- Stochastische Neuronen: $S_i = \begin{cases} 1 & \text{mit } p(V_i) \\ 0 & \text{mit } 1-p(V_i) \end{cases}$, $p(V_i) = \frac{1}{1+e^{-V_i}}$, $V_i = \sum_j W_{ij} S_j$

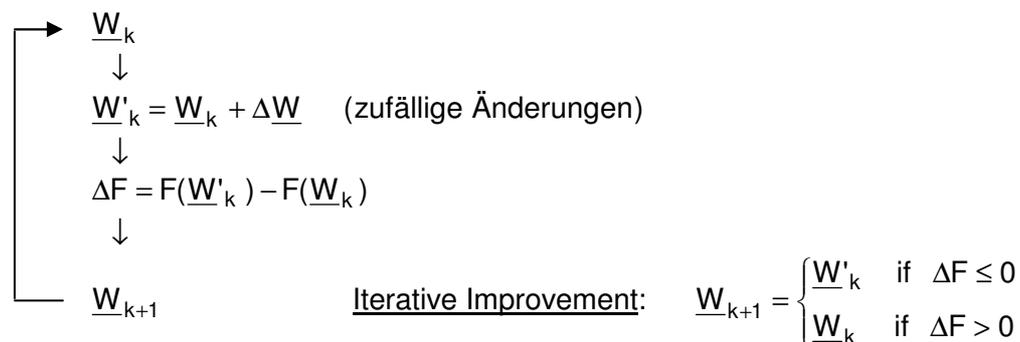
- A_{R-P}-Lernverfahren (ohne Bestrafung):

$$\Delta W_{ij} = \eta \cdot r \cdot [S_i - p(V_i)] \cdot S_j, \quad 0 \leq r \leq 1 \text{ (Bewertung)}$$

▪ Stochastische Lernstrategien

- $F(\underline{W}) = \min$

Iterationsschritt $k \rightarrow k+1$:



▪ Optimales Lernen und Verallgemeinern

- Skalierung der Inputs: (Mittelwert ≈ 0 , Wertebereich $\approx [-1, +1]$)

- Wahl der Anfangsgewichte: Faustregel: $W_{ij}^0 \in [-W_0, +W_0]$, $W_0 \approx \frac{2.4}{M_i}$
 (M_i = Anzahl Inputs des Neurons i)

- Anzahl freie Parameter (Gewichte): # Gewichte \leq # Lernbeispiele

- Weight Sharing

- Weight Decay

- Stopped Training

- Crossvalidation

- Kombination von neuronalen Netzwerken (arithmetisches Mittel, Majority)

- Active Learning Strategien (Space-Filling, Maximale Modellierungsunsicherheit)

Hinweis: Diese Zusammenfassung ist für die Prüfungsvorbereitung gedacht und darf an der Prüfung nicht verwendet werden!

(An der Prüfung sind ausser Taschenrechner keine Hilfsmittel zugelassen.)