Aufgabe: Beweisen Sie, daß die Leistungsfähigkeit eines neuronalen Netzwerks von der Zahl der Neuronen abhängt, und daß eine autonome Entscheidung bei entsprechendem Training mit einem geringeren Fehler behaftet ist als jede andere Art von Programmierung.

Lösung: Eine Entscheidung heißt autonom oder „frei“, wenn ein Ereignis A mehrere Entscheidungen $B_1, \dots, B_n$ auslösen kann, die entweder richtig (w) oder falsch (f) sein können, d.h. die n Subjunktionen

$$A \Rightarrow B_1, \quad \dots, \quad A \Rightarrow B_n,$$

mit $n \geq 2$, sind äquivalent zu $A \Rightarrow B_1 \vee \dots \vee B_n$. Dabei muß mindestens eine der Entscheidungen

$$B_1 \vee B_2 \vee \dots \vee B_n$$

richtig bzw. wahr sein, unabhängig davon, ob das Ereignis A eingetreten ist. Ist es nicht eingetreten bzw. falsch, können die Entscheidungen beliebig sein, denn es gilt folgende Wahrheitstabelle:

	A	$B_1 \vee \dots \vee B_n$	$A \Rightarrow B_1 \vee \dots \vee B_n$
(1)	w	w	w
(2)	w	f	f
(3)	f	w	w
(4)	f	f	w

Tabelle 1. Wahrheitstabelle der autonomen Entscheidungen

Somit ist gezeigt, daß von 2^n möglichen Entscheidungen nur eine einzige komplett falsch ist, eine optimal und alle anderen sub-optimal sind. Es müssen nämlich immer sämtliche Entscheidungen falsch sein, damit auch die Disjunktion falsch ist. Ähnlich verhält es sich in einem neuronalen Netzwerk. Je mehr Entscheidungen demnach möglich sind, desto geringer wird auch die Wahrscheinlichkeit, daß alle Entscheidungen falsch sind.

Wählen wir als einfaches Beispiel den Wurf einer Münze. Jede Münze hat zwei Seiten, Kopf oder Zahl. Wenn die Münze „richtig“ fällt, erscheint der Kopf, der von einer weisen Entscheidung zeugt. Fällt hingegen die Zahl, die symbolisch für die entstandenen Kosten steht, war die Entscheidung schlecht. Bei 2^n möglichen Entscheidungen, die getroffen werden können, entspricht dies genau n Münzen, die gleichzeitig (parallel) geworfen werden müssen, um am Ende abzählen zu können, wie viele Entscheidungen zu unseren Gunsten ausgefallen sind und wie viele nicht. Die Zufallswahrscheinlichkeit des Münzwurfes gehorcht einem Binomialgesetz:

$$P(n) = (w + f)^n = \sum_{k=0}^{n} \binom{n}{k} w^{n-k} f^k.$$

Unter der Annahme, daß Kopf und Zahl gleich wahrscheinlich sind, d.h. $w = f = 1/2$, gilt

$$P(n) = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k} = \frac{1}{2^n}\left[\binom{n}{0}+\binom{n}{1}+\ldots+\binom{n}{n}\right] = 1.$$

Für $n = 3$ bedeutet das zum Beispiel, daß gemäß

$$P(3) = \frac{1}{2^3}\sum_{k=0}^{3}\binom{3}{k} = \frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8} = 1$$

die Wahrscheinlichkeit, daß alle drei Münzen die Zahl zeigen, nur $1/8$ ist, und die Wahrscheinlichkeit, daß entweder ein oder zwei Neuronen auf die Zahl fallen, gleich $3/8$ beträgt. Die Gesamtwahrscheinlichkeit, daß von 8 zur Auswahl stehenden Möglichkeiten mindestens eine falsche Entscheidung enthalten ist, ist dann $7/8$. Die ODER-Verknüpfung folgt dem binomischen Satz, wonach von den 2^n Kombinationen aus n verknüpften Entscheidungen nur eine total falsch ist, d.h.

$$f^n = \frac{1}{2^n}.$$

In einem gut trainierten neuronalen Netzwerk sollte die Zahl weniger oft fallen als der Kopf. Unter Berücksichtigung der Beziehung $w = 1 - f$ läßt sich die binomische Reihe zerlegen in eine Summe zweier Wahrscheinlichkeiten für wahre und falsche Entscheidungen:[1]

$$P(n) = P_n^w(w) + P_n^f(f) = w^n + \sum_{k=1}^{n}\binom{n}{k}(1-f)^{n-k} f^k.$$

Unter der Annahme, daß in dem besagten, gut trainierten neuronalen Netzwerk weniger als die Hälfte aller Entscheidungen (Münzwürfe) falsch sind, daß es also nicht nur zwei diskrete Möglichkeiten gibt, macht die Wahrscheinlichkeit für einen Fehler nur den Bruchteil $1/m$ der Wahrscheinlichkeit für eine korrekte Entscheidung aus, d.h.

$$f = \frac{w}{m} \quad \text{bzw.} \quad \left(1+\frac{1}{m}\right)w = 1,$$

wobei m eine natürliche Zahl größer oder gleich 2 ist. Damit folgt aus

$$f = \frac{1}{1+m} \quad \text{und} \quad w = \frac{m}{1+m}$$

die Relation

[1] Genauer gesagt teilweise falsch

$$P_n^f(m) = \frac{m^n}{(1+m)^n} \sum_{k=1}^{n} \frac{1}{m^k} \binom{n}{k} = \frac{1}{(1+m)^n} \left[\binom{n}{1} m^{n-1} + \binom{n}{2} m^{n-2} + \ldots\ldots + \binom{n}{n} \right].$$

Für die 7 ersten Neuronen haben wir demnach folgende Fehlerwahrscheinlichkeiten:

$$P_2^f(m) = \frac{1}{(1+m)^2}(2m+1),$$

$$P_3^f(m) = \frac{1}{(1+m)^3}(3m^2+3m+1),$$

$$P_4^f(m) = \frac{1}{(1+m)^4}(4m^3+6m^2+4m+1),$$

$$P_5^f(m) = \frac{1}{(1+m)^5}(5m^4+10m^3+10m^2+5m+1),$$

$$P_6^f(m) = \frac{1}{(1+m)^6}(6m^5+15m^4+20m^3+15m^2+6m+1),$$

$$P_7^f(m) = \frac{1}{(1+m)^7}(7m^6+21m^5+35m^4+35m^3+21m^2+7m+1).$$

Wir definieren nun

$$m \equiv \sum_{k=0}^{n} \binom{n}{k} = 2^n$$

als die maximal mögliche Zahl von Ausgabemustern des neuronalen Netzwerks[2] und berechnen damit die Fehlerwahrscheinlichkeit in Abhängigkeit von der Zahl der Trainingsmuster. Das Ergebnis ist in Tab. 2 dargestellt.

m	2	4	8	16	32	64	128
$P_2^f(m)$	0,5556	0,3600					
$P_3^f(m)$	0,7037	0,4880	0,2977				
$P_4^f(m)$	0,8025	0,5904	0,3757	0,2153			
$P_5^f(m)$	0,8683	0,6723	0,4451	0,2615	0,1426		
$P_6^f(m)$	0,9122	0,7379	0,5067	0,3049	0,1686	0,0888	
$P_7^f(m)$	0,9415	0,7903	0,5615	0,3458	0,1938	0,1028	0,0530

Tabelle 2. Fehlerwahrscheinlichkeit eines neuronalen Netzwerks mit *n* Neuronen in Abhängigkeit von der Zahl der Trainingsmuster

Die minimale Zahl der Trainingsmuster kann 2 nicht unterschreiten, das entspricht dem Wurf einer einzigen Münze. Der äußert linke und äußerst rechte Zahlenwert in jeder Zeile gibt zu-

[2] z.B. in einem Perzeptron

gleich die Intervallgrenze an, innerhalb derer sich der Fehler bewegen kann. Obige Fehlerfunktionen sind in Abb. 1 nochmals graphisch aufgetragen.

Ein neuronales Netzwerk kann nicht besser trainiert werden, als es die maximale Zahl von Trainingsmustern zuläßt. Es hat z.B. keinen Sinn, ein Netzwerk aus 2 Ausgangsneuronen mit 8 Trainingsmustern trainieren zu wollen, weil die Kapazität nur für maximal 4 Trainingsmuster ausgelegt ist.

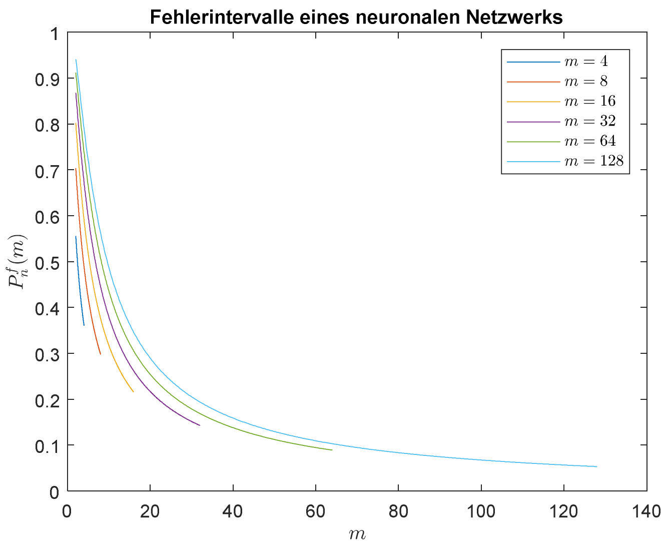

Abbildung 1. Die Abhängigkeit der Fehlerfunktionen von den Trainingseinheiten

Die Zahl der möglichen Entscheidungen in einem Netzwerk von *n* Neuronen beträgt 2^n, was der Quersumme der Binomialkoeffizienten entspricht. Damit hängt der Fehler nur noch von der Neuronenzahl ab, und in der Näherung für große *n* gilt für den Grenzwert

$$P_n^f(\infty) = \lim_{n\to\infty} P_n^f(n) = \lim_{n\to\infty}\left(\frac{2^n}{1+2^n}\right)^n \sum_{k=1}^{n}\left(\frac{1}{2^n}\right)^k \binom{n}{k} \approx \lim_{n\to\infty}\left(\frac{2^n}{1+2^n}\right)^n \frac{n}{2^n} = 0,$$

d.h. ein neuronales Netzwerk kann theoretisch und für beliebig viele Neuronen einen verschwindenden Fehler haben, sofern es nur ausreichend trainiert ist. Überschlägig kann für hinreichend großes *n* eine theoretisch erreichbare Fehleruntergrenze von

$$P_n(n >> 1) \simeq \frac{n}{2^n}$$

erzielt werden. Ein Netzwerk mit 30 Neuronen hat demnach einen Fehler, der nicht größer ist als $2{,}8\cdot 10^{-8}$. Mit 100 Neuronen läßt sich der maximale Fehler theoretisch auf $7{,}9\cdot 10^{-29}$ begrenzen. In der Regel kann man aber aus rein praktischen Erwägungen dem Netzwerk längst nicht alle Trainingsmuster einspeisen. Daher kann der Fehler auch um Größenordnungen über dem theoretischen Grenzwert liegen.

Dieses Ergebnis wirft auch Fragen bezüglich einer modifizierten Interpretation von Intelligenz auf. Ein Elektronengehirn[3] ist um so intelligenter, je mehr Neuronen miteinander verknüpft sind und je kleiner die daraus resultierende Wahrscheinlichkeit für eine Fehlleistung ist.[4] Intelligenz wäre demnach gleichbedeutend damit, möglichst wenige Fehler zu begehen.

[3] Und wahrscheinlich auch ein menschliches
[4] Die ausschließlich vom Trainingszustand des Gehirns abhängt

