Informatik IV - Tutorium XVI & XVII

Tut Nr. 3 – Optimierung & Neuronale Netze

David Münch

Universität Karlsruhe (TH)
Fakultät für Informatik
IBDS Prautzsch

Dank an Yusuf für die Zusammenarbeit.

20. Mai 2009

Universität Karlsruhe (TH)
Forschungsuniversität • gegründet 1825

Inhaltsverzeichnis

Inhaltsverzeichnis

Inhaltsverzeichnis

Inhaltsverzeichnis

Organisatorisches

Email: muenchdavid@gmail.com

https://www.stud.uni-karlsruhe.de/~uhbro/

Tutorium 16: Freitags 8:00 Uhr - Raum -107
Tutorium 17: Freitags 9:45 Uhr - Raum -119

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009
- 02.07.2009
- 23.07.2009

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009
- 02.07.2009
- 23.07.2009

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Ab 140 Punkten gibt es einen Notenbonus von ca. 1/3 Note.

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009
- 02.07.2009
- 23.07.2009

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Ab 140 Punkten gibt es einen Notenbonus von ca. 1/3 Note.

Abgabe in Zweiergruppe erlaubt und ausdrücklich erwünscht!

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009
- 02.07.2009
- 23.07.2009

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Ab 140 Punkten gibt es einen Notenbonus von ca. 1/3 Note.

Abgabe in Zweiergruppe erlaubt und ausdrücklich erwünscht!

Bitte folgendes Deckblatt verwenden:
`http://www.stud.uni-karlsruhe.de/~unbdh/deckblatt/`

Literatur

Boehm, Prautzsch: Numerical Methods. AK Peters 1993. ISBN 3-528-06350-5
http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&fbt=7319953&nd=3204657

Ash: Information Theory. Dover 1990. ISBN 0-486-66521-6
http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&nd=9866904

Goos: Vorlesungen über Informatik. Bd. 4, Springer 1998. ISBN 3-540-60650-5
http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&fbt=9316367&nd=6568301

Was wollen wir heute erreichen?

Was wollen wir heute erreichen?

- 1. Scheinklausur besprechen

Was wollen wir heute erreichen?

- 1. Scheinklausur besprechen
- Stochastische Optimierungsalgorithmen kennen lernen

Was wollen wir heute erreichen?

- 1. Scheinklausur besprechen
- Stochastische Optimierungsalgorithmen kennen lernen
- Evolutionäre Algorithmen verstehen

Was wollen wir heute erreichen?

- 1. Scheinklausur besprechen
- Stochastische Optimierungsalgorithmen kennen lernen
- Evolutionäre Algorithmen verstehen
- Neuronale Netze

Aufgabe 1

Gegeben sei eine Untermenge $\mathcal{U}$ des Raums, dir durch folgende Ungleichungen definiert wird.

$$(*) \begin{cases} 2x + 3y + 4z + 3 & \geq 0 \\ 2x - 4y - 3z + 3 & \geq 0 \\ x - 2y + 2z - 2 & \leq 0 \\ -2x - 3y + 3z + 4 & \geq 0 \end{cases}$$

Sei $\mathcal{B} = [a, b] \times [c, d] \times [e, f]$ der minimale Quader, der $\mathcal{U}$ enthält. (Bounding Box von $\mathcal{U}$)

a) Geben Sie zum Simplex $(*)$ je eine Zielfunktion zur Berechnung von a und b an.

Aufgabe 1

Gegeben sei eine Untermenge $\mathcal{U}$ des Raums, dir durch folgende Ungleichungen definiert wird.

$$(*)\begin{cases} 2x + 3y + 4z + 3 & \geq 0 \\ 2x - 4y - 3z + 3 & \geq 0 \\ x - 2y + 2z - 2 & \leq 0 \\ -2x - 3y + 3z + 4 & \geq 0 \end{cases}$$

Sei $\mathcal{B} = [a, b] \times [c, d] \times [e, f]$ der minimale Quader, der $\mathcal{U}$ enthält. (Bounding Box von $\mathcal{U}$)

a) Geben Sie zum Simplex $(*)$ je eine Zielfunktion zur Berechnung von a und b an.

b) Berechnen Sie den Parameter a, ausgehend vom zulässigen Punkt $\vec{0}$, mithilfe des Simplexverfahrens.

Aufgabe 2

Gegeben sei die Bewertungsfunktion $f : \mathbb{R}^3 \to \mathbb{R}$ mit

$$[x\ y\ z]^\mathsf{T} \mapsto 5 - x^2 + xy - y/z.$$

Gesucht ist ein Punkt $\vec{x}$, der die Bewertung $f(\vec{x})$ maximiert.

a) Berechnen Sie den Gradienten von f für $z \neq 0$.

Aufgabe 2

Gegeben sei die Bewertungsfunktion $f : \mathbb{R}^3 \to \mathbb{R}$ mit

$$[x\ y\ z]^\mathsf{T} \mapsto 5 - x^2 + xy - y/z.$$

Gesucht ist ein Punkt $\vec{x}$, der die Bewertung $f(\vec{x})$ maximiert.

a) Berechnen Sie den Gradienten von f für $z \neq 0$.

b) Berechnen Sie zum Startwert $\vec{x}_0 = [0\ 2\ 1]^\mathsf{T}$ die nächste Näherungslösung $\vec{x}_1$ mithilfe des Gradientenverfahrens, wenn $h = 1/2$ die Schrittweite ist. Geben Sie die Bewertungen $f(\vec{x}_0)$ und $f(\vec{x}_1)$ aus.

Aufgabe 3

Lösen Sie das Ausgleichsproblem

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \vec{p} := \begin{bmatrix} -5 \\ -5 \\ -1 \\ 1 \end{bmatrix}$$

unter den Nebenbedingungen

$$y_1 - y_3 + 6 = 0 \text{ und } y_2 + y_4 + 2 = 0,$$

d.h. minimieren Sie $\|\vec{y} - \vec{p}\|_2$.

Aufgabe 4

Gegeben sei die Abbildung $f : \mathbb{R}^2 \to \mathbb{R}^3$ mit

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x + y \\ x - y \\ xy^2 \end{bmatrix}$$

und der Punkt $\vec{p} = [2\ 1\ 1]^\intercal$.
Gesucht ist ein $\vec{x}$, das den Term $\|f(\vec{y}) - \vec{p}\|^2$ minimiert.
Berechnen Sie zum Startwert $\vec{x}_0 = [1\ 1]^\intercal$ die nächste Näherungslösung $\vec{x}_1$ mit einem Schritt des Levenberg-Marquardt-Algorithmus und dem Dämpfungsfaktor $\mu = 1$.

Aufgabe 5

Gegeben seien zwei Punktmengen $\mathcal{P}_1 = \{ \begin{bmatrix} 0 \\ 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \end{bmatrix} \}$ und $\mathcal{P}_2 = \{ \begin{bmatrix} 4 \\ 0 \end{bmatrix} \begin{bmatrix} 8 \\ 0 \end{bmatrix} \begin{bmatrix} 8 \\ 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} \}$.

a) Berechnen Sie für jede Punktemenge die Trägheitsellipse.

Aufgabe 5

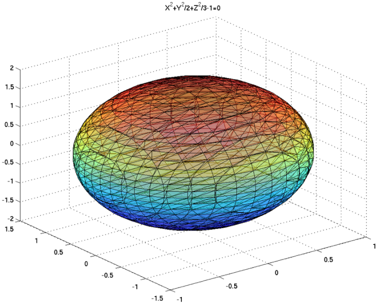
Gegeben seien zwei Punktmengen $\mathcal{P}_1 = \{ \begin{bmatrix} 0 \\ 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \end{bmatrix} \}$ und $\mathcal{P}_2 = \{ \begin{bmatrix} 4 \\ 0 \end{bmatrix} \begin{bmatrix} 8 \\ 0 \end{bmatrix} \begin{bmatrix} 8 \\ 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} \}$.

a) Berechnen Sie für jede Punktemenge die Trägheitsellipse.

b) Geben Sie die affine Transformation an, die die Ellipse von $\mathcal{P}_2$ auf die von $\mathcal{P}_1$ abbildet und insbesondere die Hauptachse der Ellipse von $\mathcal{P}_2$ auf die von $\mathcal{P}_1$ abbildet.

Trägheitsellipsoid

Trägheitsellipsoid TE ist definiert durch
$TE = \{\vec{x} | (\vec{x} - \vec{m})^t T (\vec{x} - \vec{m}) = 1\}$
mit Trägheitstensor $T = spur(K)E - K$. T ist (positiv definit) und symmetrisch und hat positive Eigenwerte.



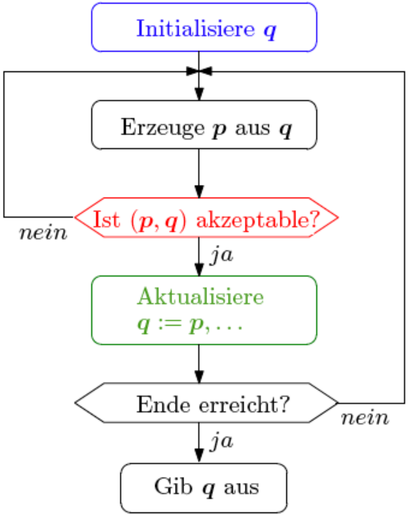
Wir wollen eine Transformation $f(\vec{x}) = A\vec{x} + \vec{b}$ angeben, die ein TE B nach A transformiert.

Transformation

- Mittelpunkt $\vec{m}_B$ von B muss auf Mittelpunkt $\vec{m}_A$ von A abgebildet werden.
  $\Rightarrow$ 1. Bedingung: $f(\vec{m}_B) = A\vec{m}_B + \vec{b} = \vec{m}_A$
- Die Eigenvektoren geben die Richtung und die Eigenwerte die dazugehörigen halben Längen der Hauptachsen an.
  $\Rightarrow$ 2. Bedingung: $A(\vec{m}_B + \frac{\vec{v}_{Bi}}{\sqrt{\lambda_{Bi}}}) + \vec{b} \in \{\vec{m}_A \pm \frac{\vec{v}_{Ai}}{\sqrt{\lambda_{Ai}}}\}$

Stochastische Verfahren

Zufallsgesteuerte Optimierung

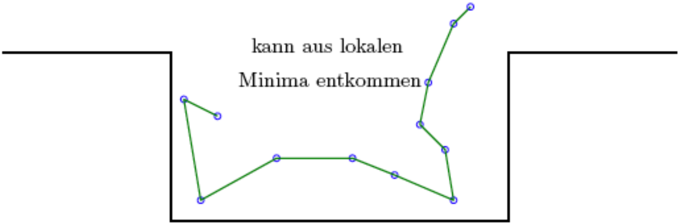


Algorithm 1 Simuliertes Tempern

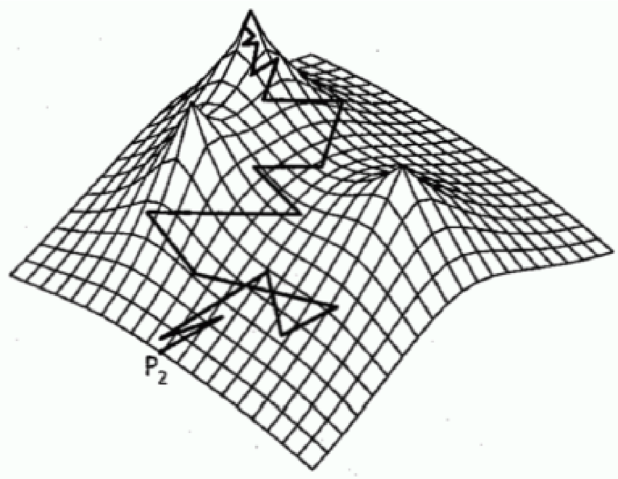
input : initialer Zustand q_0, initiale Temperatur T_0, Kostenfkt. c
output: beste Lösung

$q \leftarrow q_0$, $T \leftarrow T_0$
while *Stopp-Kriterium nicht erfüllt* **do**
    while *noch nicht im Gleichgewicht* **do**
        $q' \leftarrow$ irgendein zufälliger, benachbarter Zustand zu q;
        $\Delta c = c(q') - c(q)$;
        $Probability \leftarrow min(1, e^{\frac{-\Delta c}{T}})$;
        if $random(0,1) \leq Probability$ **then**
            $q \leftarrow q'$;
        end
    end
    Aktualisiere Temperatur T;
end
Gib die beste Lösung aus.

Simuliertes Tempern



Simuliertes Tempern



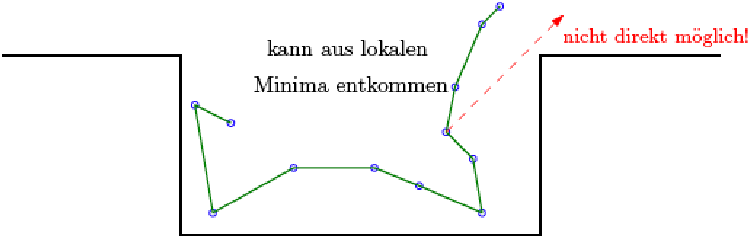
Schwellwertalgorithmus

Nachfolger p von q wird akzeptiert, wenn

$$c(p) \leq c(q) + \sigma$$

wobei σ der Schwellwert ist, welcher regelmässig abgesenkt wird.

Schwellwertalgorithmus



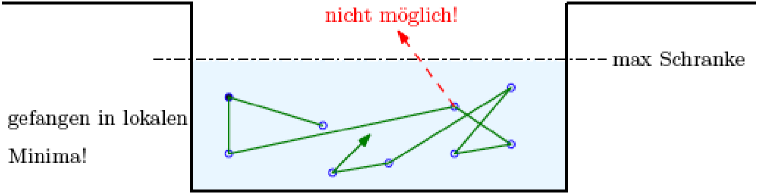
Sintflut-Algorithmus

Nachfolger p von q wird akzeptiert, wenn

$$c(p) \leq H$$

wobei H die Sintflutgrenze ist, die regelmässig abgesenkt wird.

Sintflut-Algorithmus



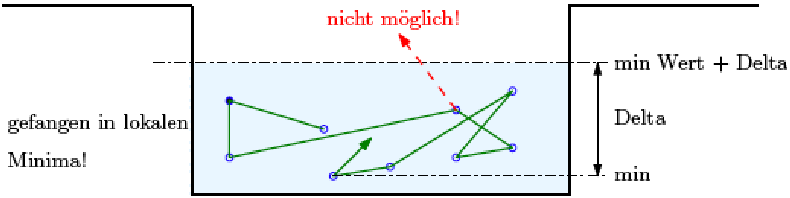
Rekordjagt-Algorithmus

Es wird die beste Bewertung gespeichert ($record$). Diese darf nur einen bestimmten Wert δ überschritten werden.
Nachfolger p von q wird akzeptiert, wenn

$$c(p) \leq record + \delta$$

wobei der Maximalwert δ regelmässig abgesenkt wird.

Rekordjagt-Algorithmus



Zusammenfassung

Zusammenfassung

- Schwellwert-Verfahren und Simuliertes Tempern können aus kleinen, aber tiefen lokalen Minima entkommen, die anderen Verfahren nicht. Aber bei grösser werdenden Problemen werden die lokalen Minima weniger markant und tief.

Zusammenfassung

- Schwellwert-Verfahren und Simuliertes Tempern können aus kleinen, aber tiefen lokalen Minima entkommen, die anderen Verfahren nicht. Aber bei grösser werdenden Problemen werden die lokalen Minima weniger markant und tief.
- Sintflut- und Schwellwert-Verfahren sind dem Simulierten Tempern überlegen.

Zusammenfassung

- Schwellwert-Verfahren und Simuliertes Tempern können aus kleinen, aber tiefen lokalen Minima entkommen, die anderen Verfahren nicht. Aber bei grösser werdenden Problemen werden die lokalen Minima weniger markant und tief.
- Sintflut- und Schwellwert-Verfahren sind dem Simulierten Tempern überlegen.
- Der Sintflut-Algorithmus ist fast genauso gut wie das Schwellwert-Verfahren, nicht ganz so stabil in der Qualität der Lösung, dafür jedoch etwas schneller.

Zusammenfassung

- Schwellwert-Verfahren und Simuliertes Tempern können aus kleinen, aber tiefen lokalen Minima entkommen, die anderen Verfahren nicht. Aber bei grösser werdenden Problemen werden die lokalen Minima weniger markant und tief.
- Sintflut- und Schwellwert-Verfahren sind dem Simulierten Tempern überlegen.
- Der Sintflut-Algorithmus ist fast genauso gut wie das Schwellwert-Verfahren, nicht ganz so stabil in der Qualität der Lösung, dafür jedoch etwas schneller.
- Das Schwellwert-Verfahren ist die Methode der Wahl, also vorzuziehen.

Zusammenfassung

- Schwellwert-Verfahren und Simuliertes Tempern können aus kleinen, aber tiefen lokalen Minima entkommen, die anderen Verfahren nicht. Aber bei grösser werdenden Problemen werden die lokalen Minima weniger markant und tief.
- Sintflut- und Schwellwert-Verfahren sind dem Simulierten Tempern überlegen.
- Der Sintflut-Algorithmus ist fast genauso gut wie das Schwellwert-Verfahren, nicht ganz so stabil in der Qualität der Lösung, dafür jedoch etwas schneller.
- Das Schwellwert-Verfahren ist die Methode der Wahl, also vorzuziehen.
- Wenn man ein reichhaltiges Sortiment von Veränderungsschritten zulässt, wird auch besser optimiert.

Evolutionäre Algorithmen

Es wird versucht das Evolutionsprinzip der Biologie nachzubilden. Die Elemente $q \in Q$ sind die Induviduen einer Population. Dabei besitzt jedes Individuum:

- einen Genotyp (Merkmalsvektor) $k \in (R)^n$ und
- einen Phänotyp (Bewertung). (Fitnessfunktion $fit(q)$.

Individuen können sich fortpflanzen und dabei ihren Genotyp verändern (Mutation). Bei einem Elter: Klon; bei mehreren Eltern: Kreuzung.

Ein Individuum wird durch das Paar (k, σ) repräsentiert.

Ein Individuum wird durch das Paar (k, σ) repräsentiert.
σ ist der Streuungsvektor, der dir maximale Veränderung bei der Mutation angibt. Kann auch geändert werden.

Ein Individuum wird durch das Paar (k, σ) repräsentiert.
σ ist der Streuungsvektor, der dir maximale Veränderung bei der Mutation angibt. Kann auch geändert werden.
Nachkommen mutieren nach der Regel: $k^{t+1} = k^t + d$, wobei die Koordinaten von d normalverteilt sind mit dem Streuungsvektor σ.

Plus-Strategie $(\mu + \lambda)$

Aus den μ Eltern werden λ Nachkommen geklont (Eltern werden gleichverteilt unter den μ Individuen gewählt, jedes Individuum kann auch mehrmals Elter werden)
Die μ Individuen mit der besten Fitness-Funktion bilden die nächste Generation (Eltern können also überleben)
Unter μ Eltern und λ Nachkommen werden die μ stärksten (u.a. mit $fit(q)$) ausgewählt, die anderen λ sterben.

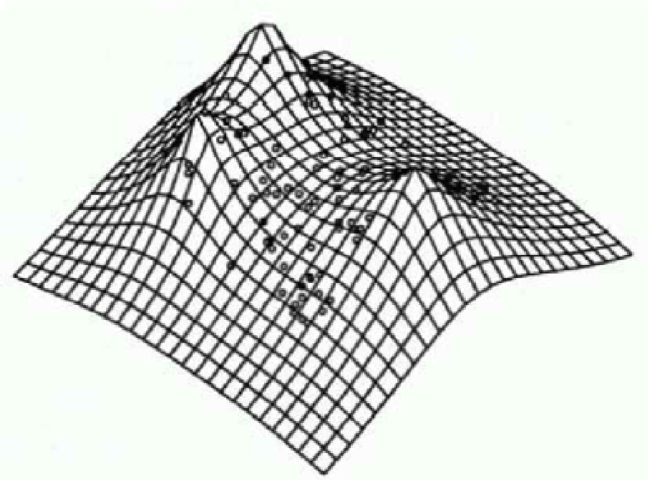
Plus-Strategie $(\mu + \lambda)$

Aus den μ Eltern werden λ Nachkommen geklont (Eltern werden gleichverteilt unter den μ Individuen gewählt, jedes Individuum kann auch mehrmals Elter werden)
Die μ Individuen mit der besten Fitness-Funktion bilden die nächste Generation (Eltern können also überleben)
Unter μ Eltern und λ Nachkommen werden die μ stärksten (u.a. mit $fit(q)$) ausgewählt, die anderen λ sterben.

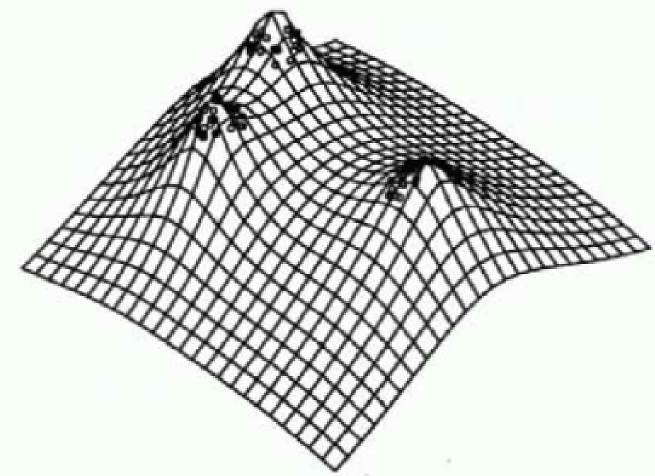
Komma-Strategie (μ, λ)

Aus den μ Eltern werden λ Nachkommen geklont (Eltern werden gleichverteilt gewählt, jedes Individuum kann auch mehrmals Elter werden)
Die Eltern überleben nicht, es werden nur unter den Nachkommen die μ stärksten ausgesucht.

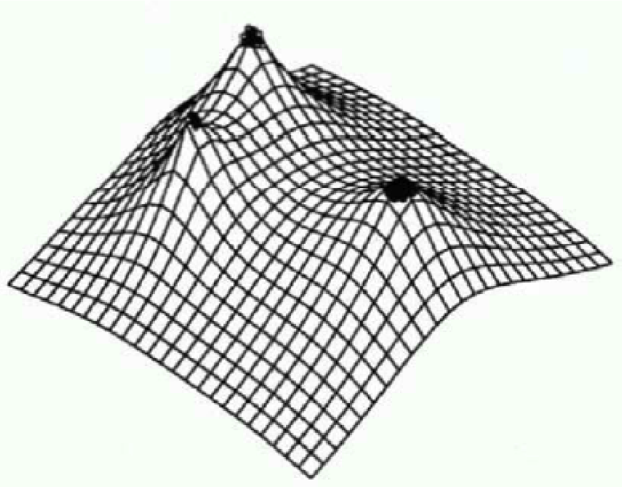
ES(150+1000), Generation 1



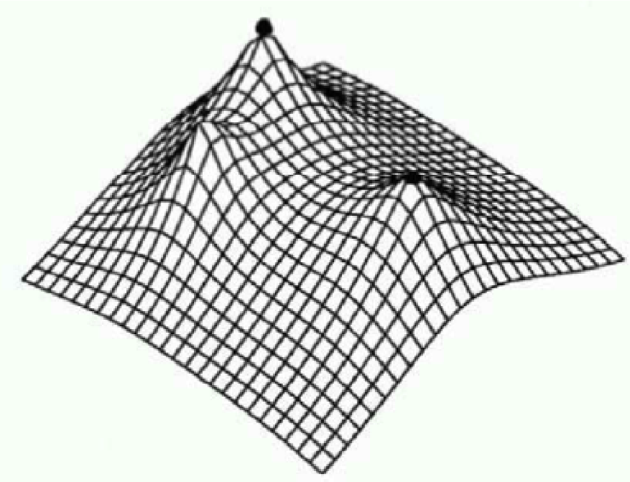
ES(150+1000), Generation 2



ES(150+1000), Generation 3



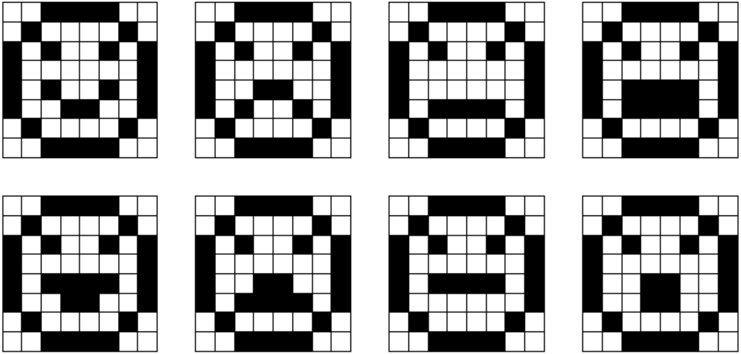
ES(150+1000), Generation 6



Aufgabe

Betrachten Sie die hier aufgeführten Emotionen als 8×8 Pixel Bilder die jeweils die Werte 1 (Schwarz) und 0 (Weiss) annehmen können.
Entwerfen Sie ein neuronales Netz, das in der Lage ist zwischen den vier verschiedenen Emotionen zu unterscheiden.
Beschreiben Sie die Eingabe, Ausgabe des Netzes und geben Sie eine Topologie des Netzes mit möglichst wenig Neuronen an.
Geben Sie ausserdem die Gewichte sowie Schwellwerte der Neuronen an. Gehen Sie dabei von der binären Sprungfunktion als Aktivierungsfunktion für die Neuronen aus.

Neuronale Netze



Lachend Traurig Neutral Erstaunt

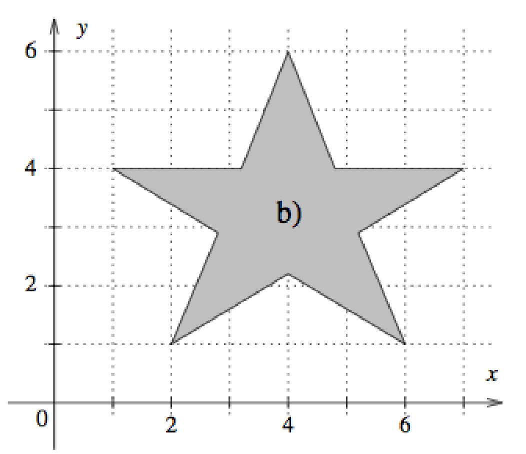
Aufgabe

Ein Perzeptron soll Punkte im Innern geometrischer Figuren, hier: Polygone, als solche korrekt klassifizieren.

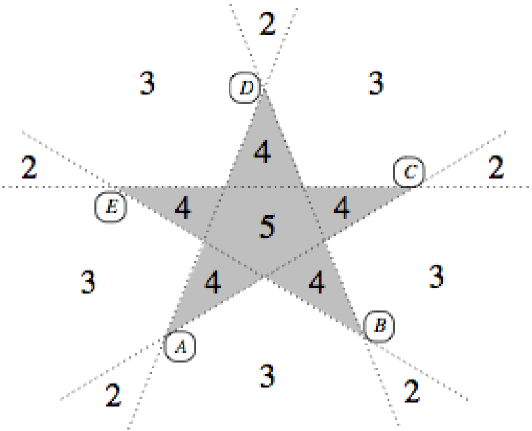
Dazu stehen binäre $\{0,1\}$-Neuronen zur Verfügung mit der Eigenschaft, daß ihre Ausgabe q *undefiniert*[1] ist, wenn der Erregungszustand $\sum w_i q_i$ genau mit dem Schwellwert b übereinstimmt. Wie üblich sei für $\sum w_i q_i > b$ die Ausgabe $q = 1$, und $q = 0$ für $\sum w_i q_i < b$.

Geben Sie ein neuronales Netz – mit sämtlichen Gewichten und Schwellwerten – an, das genau die Figur b) erkennt. Benötigen Sie mehr als 6 Neuronen?

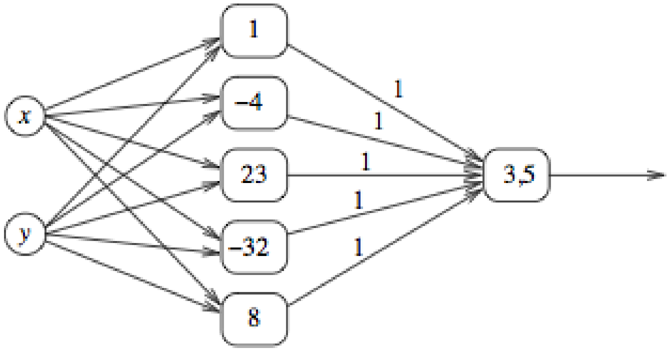
Aufgabe



Anzahl erregter Neuronen



Lösung



Noch Fragen?

Vorschau

Vorschau

- Lernen von Neuronalen Netzen

Vorschau

- Lernen von Neuronalen Netzen
- Hopfield-Netze

Vorschau

- Lernen von Neuronalen Netzen
- Hopfield-Netze
- Informationstheorie

Bis zum nächsten Mal