

MatrikelNr:



Klausur zur Vorlesung Adaptive Systeme Wintersemester 2013/2014

Datum: 10.02.2014

Vorname:
Name:
Matrikelnummer:
Geburtsdatum:
Studiengang:

Als BSc bearbeiten Sie bitte den Teil der Aufgaben, der mit „AS-1“ gekennzeichnet ist. (max 80 Minuten). Als MSc bearbeiten Sie den „AS-1“ (max. 80 Minuten) und/oder den „AS-2“-Teil (max. 100 Minuten, zusammen 180 Minuten). Im AS-1-Teil sind mit 80 Punkten 100% erreicht. Für den AS-2-Teil sind 100 Punkte für 100% nötig.

Die Punktzahl einer Aufgabe entspricht ungefähr der maximalen Bearbeitungsdauer der Aufgabe in Minuten. Die Gesamtpunktzahl der Aufgaben eines Teils übersteigt jeweils 100%, so dass Sie nicht alle Aufgaben lösen müssen.

Durch die Übungspunkte können maximal 10% der Klausurleistung erbracht werden. Als Hilfsmittel ist ein Taschenrechner erlaubt. Bitte benutzen Sie für Notizen die Rückseiten der Aufgabenblätter.

Viel Erfolg!

Wird vom Prüfer ausgefüllt:

AS1-1	AS1-2	AS1-3	AS1-4	AS1-5	AS1-6	AS1-7	AS1-8		Σ
/10	/15	/20	/8	/10	/16	/10	/6		/80
AS2-1	AS2-2	AS2-3	AS2-4	AS2-5	AS2-6	AS2-7	AS2-8	AS2-9	
/15	/20	/8	/15	/10	/10	/20	/15	/7	/100

Punkte Klausur:

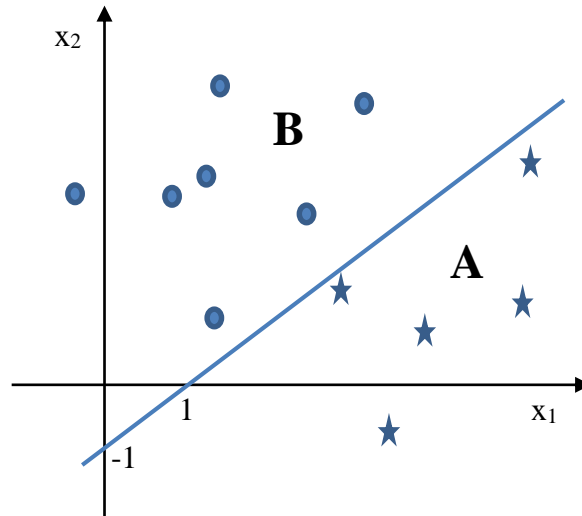
Punkte Übungen:

Punkte Gesamt:

Note:

AS-1.1 Klassifikation**10 Pkte**

Gegeben sei eine Klassentrennung von Mustern $\mathbf{x} = (x_1, x_2)$, visualisiert in dem folgenden Diagramm.



Wie lauten die Gewichte, die Aktivität und die Ausgabefunktion für ein binäres Neuron, das die Klassentrennung durchführt? Dazu soll eins für „ \mathbf{x} aus Klasse A“ und null für „ \mathbf{x} aus Klasse B“ ausgegeben werden.

AS-1.2 Perzeptron und Adaline**15 Pkte**

a) Vergleichen Sie die Perzeptron-Lernregel

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \gamma(L - y) \mathbf{x}$$

mit der Adaline-Lernregel

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \gamma(L - \mathbf{w}^T \mathbf{x}) \mathbf{x} / |\mathbf{x}|^2$$

Was sind die beiden wesentlichen Unterschiede? (5 Pkte)

b) Wie lässt sich eine Konvergenz der Lernregeln erreichen, obwohl die Muster nicht linear separierbar sind? (4 Pkte)

c) Was versteht man unter „overfitting“ und was kann man dagegen tun? (6 Pkte)

AS-1.3 PCA und ICA

20 Pkte

Ein schneller Fixpunkt-Algorithmus für PCA besteht aus folgenden Schritten:

1. Zentrieren der gesamten Eingabe $X_1 = \{\mathbf{x}\}$ der Dimension n . Sei $i=1$.
2. Initialisieren des Gewichtsvektors $\mathbf{w}_i(t=0)$
3. Bilden der Kovarianzmatrix $C_i = \langle \mathbf{x}\mathbf{x}^T \rangle$ mit \mathbf{x} aus X_i
4. Bilden von $\mathbf{w}_i(t+1) = C_i \mathbf{w}_i$ (*Fixpunktiteration*)
5. Orthogonalisieren des Vektors \mathbf{w}_i zu allen anderen vorher gefundenen Vektoren
6. Normieren von \mathbf{w}_i auf den Betrag 1.
7. Führe solange Schritte 4 - 6 aus, bis \mathbf{w}_i zum Eigenvektor \mathbf{e}_i konvergiert ist.
8. Wenn $i = n$, stop. Ansonsten bilde eine neue Eingabe $X_{i+1} = \{\mathbf{x} \mid \mathbf{x} = \mathbf{x} - (\mathbf{e}_i^T \mathbf{x}) \mathbf{e}_i\}$ aus der um die Komponenten in \mathbf{e}_i -Richtung reduzierten Eingabe.
9. Führe für jede weitere Dimension $i = 2..n$ die Schritte 2 bis 7 aus.

Stellen Sie die für den Fixpunkt-Algorithmus von Hyvärinen zur ICA notwendigen Schritte auf und vergleichen Sie diese mit dem PCA-Fixpunktalgorithmus. Zur Erinnerung: die Fixpunktiteration der ICA war $\mathbf{w}_i(t+1) = \langle (\mathbf{w}_i^T \mathbf{v})^3 \mathbf{v} \rangle - 3\mathbf{w}_i$

Welche Schritte sind unterschiedlich, welche gleich, und warum?

AS-1.4 Kohonenkarten**8 Pkte**

- a) Die Winner-take-all-Regel für Kohonenkarten lautet: Wähle Einheit k so, dass
- $$|\mathbf{w}_k - \mathbf{x}| = \min_i |\mathbf{w}_i - \mathbf{x}|$$

Wie lautet dagegen die klassische, korrelative winner-take-all-Regel? (3 Pkte)

- b) Wann stimmen beide überein? (2 Pkte)

- c) Geben Sie Beispiele für Nachbarschaftsfunktionen an (3 Pkte)

AS-1.5 RBF-Netze**10 Pkte**

- a) Warum konvergieren RBF-Netze mit schichten-separater Adaption schneller als solche mit Backpropagation-Learning? (5 Pkte)

- b) Warum lernen RBF-Netze schlechter als Multilayer-Perzeptrons mit sigmoidalen Ausgabefunktionen, wenn sie extrapolieren müssen? (5 Pkte)

AS-1.6 Fuzzy-Regelung**16 Punkte**

Angenommen, Sie haben drei Gemütszustände: schlechte Laune, normal, gute Laune sowie zwei Wahrnehmungen: im_Lotto_gewonnen und das Wetter mit den Zuständen (schlecht, gut).

- a) Entwerfen Sie dazu plausible Zugehörigkeitsfunktionen (8 Pkte)
- b) sowie eine plausible Fuzzy-Regelung für Ihren Gemütszustand. Sehen Sie dabei alle möglichen Prinzipien vor. (8 Pkte)

AS-1.7 Evolutionäre Algorithmen**10 Punkte**

Geben Sie die grundsätzlichen algorithmischen Iterationsschritte an, um mit Hilfe eines evolutionären Algorithmus das Optimum einer Zielfunktion zu finden.

AS-1.8 Online vs. Offline-Learning**6 Pkte**

a) Was ist der Unterschied zwischen offline-learning und online-learning? (3 Pkte)

b) Welche Methode konvergiert meist schneller und warum? (3Pkte)

Ende des Teils AS-1



Beginn Teil AS-2**AS-2.1 Formale Neuronen und Gatter****15 Pkte**

Geben Sie die Werte für die Gewichte für ein zweischichtiges neuronales Netz an, welches eine boolesche Funktion F implementiert. Diese sei für drei Eingaben x_1, x_2, x_3 folgendermaßen definiert: $F(x_1, x_2, x_3) = \overline{x_1}x_2x_3 + x_1\overline{x_2}x_3 + x_1x_2\overline{x_3}$

Im *hidden Layer* sollen drei binäre Neuronen mit $S(z) = \begin{cases} 1 & z > 0,5 \\ 0 & z \leq 0,5 \end{cases}$ zum Einsatz kommen.

Die Ausgabe wird von einem linearen Neuron erzeugt.

AS-2.2 Klassifizierung und stochastische Mustererkennung**20 Pkte**

- a) Gegeben seien die bedingten Wahrscheinlichkeiten von Mustern und Klassen. Wie lautet die optimale Klassifikationsregel? (4P)
- b) Welche Maße kennen Sie, um die Güte eines Klassifikationssystems mit Hilfe einer einzigen reellen Zahl zu beschreiben? (6P)
- c) Angenommen, ein Diagnosesystem wird durch einen Parameter p gesteuert. Welche Schritte müssen Sie durchführen, um die ROC des Diagnosesystems zu bestimmen? (10P)

AS-2.3 Approximationseigenschaften Neuronaler Netze**8 Pkte**

Angenommen, Sie haben ein neuronales Netz.

- a) Wie gut kann das Netz eine vorgegebene Funktion approximieren? (3 Pkte)
- b) Was sind die Vorbedingungen dafür? (5 Pkte)

AS-2.4 Training und Testen**15 Pkte**

Angenommen, Sie haben von 3 Klassen je 10 Datenquellen (Individuen) und von jeder Quelle 3 Muster für Training, Testen und Validieren.

- a) Wie müssen Sie Trainings- und Testmengen bilden und was müssen Sie dabei beachten? (5 Pkte)

- b) Wie funktioniert die Kreuzvalidierung? (5 Pkte)

- c) Wie funktioniert „stopped training“? (5 Pkte)

AS-2.5 TLMSE**10 Pkte**

- a) Was ist der Total Least Mean Squared Error? (4 Pkte)

- b) Warum ist er ein besseres Maß für die Güte einer linearen Approximation als der „normale“ quadratische Fehler MSE? (3 Pkte)

- c) Wie wird die approximierende Hyperebene mit kleinstem TLMSE berechnet? (3 Pkte)

AS-2.6 Ljapunov-Funktion**10 Pkte**

a) Was ist eine Ljapunov-Funktion und wozu benötigt man sie? (4 Pkte)

b) Warum erfüllt der Gradientenabstieg die Ljapunov-Bedingung? (6 Pkte)

AS-2.7 Lagrange-Optimierung**20 Pkte**

Angenommen, Sie haben Signale, die nur positiv sein können wie etwa die Spikefrequenz von Neuronen, und modellieren sie mit einer stochastischen Variablen x aus $[0, +\infty]$. Welche Wahrscheinlichkeitsverteilung $p^*(x)$ trägt die meiste Information unter allen Funktionen $p(x)$, die endlichen Mittelwert m haben? Leiten Sie sich die Form von p her mit Hilfe der Lagrang'schen Methode und beachten Sie dabei die beiden Nebenbedingungen, die Wahrscheinlichkeitsdichte $p(x)$ ist normiert auf eins $\langle 1 \rangle = 1$, und der Mittelwert ist endlich $\langle x \rangle = m$.

AS-2.8 Fixpunktgleichungen**15 Pkte**

Angenommen, Sie haben eine Iteration der Form

$$w(t+1) = w(t) - \gamma(w(t) - a)w(t) - 9 \quad \text{mit } \gamma = 4, a = 5$$

Zu welchen Werten w^* konvergiert diese Iteration? Bestimmen Sie dazu die stabilen und labilen Fixpunkte.

Zur Erinnerung: Die Lösung von $x^2 + px + q = 0$ ist $x_{1,2} = -p/2 \pm \sqrt{(p/2)^2 - q}$

AS-2.9 RBF**7 Pkte**

Was sind die wesentlichen Vorteile und Nachteile von RBF-Netzen gegenüber Multi-Layer-Perzeptrons?