30. Mai 2007	**Prüfung aus Einführung in die Mustererkennung — SS 2006**	11.00 – 12.30
Matr.Nummer: Kennzahl:	Nachname: Vorname:	

Bei der vorliegenden Prüfung können Sie eine maximale Anzahl von 30 Punkten erreichen. Bitte verwenden Sie den für die Beantwortung der Frage vorgesehenen Platz und beantworten Sie die folgenden Fragen kurz aber aussagekräftig. Sie können die Fragen auf Englisch oder Deutsch beantworten. **Keine Unterlagen sind erlaubt.**

1 Kurze Fragen

1. Ist das Multilayer Perceptron (MLP) neuronale Netz ein überwachtes oder unüberwachtes Lernverfahren? Die nicht-passende Antwort durchstreichen. (1 Punkt)

   Überwachtes Verfahren / ~~Unüberwachtes Verfahren~~

   Begründung: weil das Trainingsset aus Eingabe- und den zugehörigen Ausgabevektoren besteht

2. Ist Klassifikation anhand der Mahalanobis Distanz ein parametrisches oder nicht-parametrisches Verfahren? Die nicht-passende Antwort durchstreichen. (1 Punkt)

   Parametrisches Verfahren / Nicht-parametrisches Verfahren

   Begründung: weil eine pdf-Funktion bekannt ist

3. Die Aktivierungsfunktion in einem Backpropagation neuronalen Netz muss

   eine tanh Funktion / differenzierbar / ~~eine Signum Funktion~~ / ~~integrierbar~~

   sein. Die nicht-passenden Antworten durchstreichen. (1 Punkt)

2 Rechenbeispiele

1. Sie bekommen den Auftrag, eine Machine zu entwickeln, die automatisch zwischen zwei wiederverwertbaren Müllarten, Glasflaschen (Klasse 1, $\omega = 1$) und Plastikflaschen (Klasse 2, $\omega = 2$), unterscheidet. Ein diskretes Helligkeitsmerkmal x, das 3 Werte annehmen kann, wird für jede Flasche gemessen ($x = 1, 2$ oder 3). Die *a priori* Klassenwahrscheinlichkeiten sind: $P(\omega = 1) = 0.4$ und $P(\omega = 2) = 0.6$. Vom Trainingsset wird die folgende Merkmalsverteilung für jede Klasse $P(x = i|\omega = j)$ experimentell gemessen:

i	1	2	3
$P(x = i\|\omega = 1)$	0.8	0.2	0.0
$P(x = i\|\omega = 2)$	0.4	0.3	0.3

(a) Berechnen Sie $P(x = i)$ für $i = 1, 2, 3$. Füllen Sie die folgende Tabelle aus (1 Punkt):

$P(x=1) = 0.8 \times 0.4 + 0.4 \times 0.6 = 0.56$

[illegible]

$P(x=2) = 0.2 \times 0.4 + 0.3 \times 0.6 = 0.26$

$P(x=3) = 0.0 \times 0.4 + 0.3 \times 0.6 =$

i	1	2	3
$P(x = i)$	0.56	0.26	0.18

(b) Wie berechnen sich gemäß Bayes-Theorem die *a posteriori* Wahrscheinlichkeiten (posteriors)? (1 Punkt)

$P(\omega=j|x=i) = \frac{P(x=i|\omega=j)\,P(\omega=j)}{[illegible]}$

(c) Berechnen Sie die *a posteriori* Wahrscheinlichkeiten und füllen Sie die folgende Tabelle aus (1 Punkt):

$P(\omega=1|x=1) = \frac{0.8 \cdot 0.4}{0.56}$ 0.57

$P(\omega=2|x=1) = \frac{0.4 \times 0.6}{0.56}$

j	1	2
$P(\omega = j\|x = 1)$	0.57	0.43
$P(\omega = j\|x = 2)$	0.31	0.69
$P(\omega = j\|x = 3)$	0	1

(d) Welche Klasse soll laut Bayes-Entscheidungsregel (*Bayes Decision Rule*) und der oben ausgerechneten *a posteriori* Klassenwahrscheinlichkeiten für die drei möglichen x-Werte gewählt werden? Was ist der korrespondierende *conditional error* $P(error|x)$? Füllen Sie die folgende Tabelle aus (1 Punkt):

	Klasse (1 oder 2)	$P(error\|x)$
$x = 1$	1	0,45
$x = 2$	2	0,24
$x = 3$	2	0

(e) Berechnen Sie die Fehlerrate $P(error)$ für dieses Beispiel. Schreiben Sie die berechnete Fehlerrate ins Rechteck unten. (1 Punkt)

[illegible]

[illegible]

Fehlerrate: [illegible]

(f) Beweisen Sie, dass die Bayes-Entscheidungsregel im Zwei-Klassen Fall die Fehlerrate $P(error)$ immer minimiert. (2 Punkte)

[illegible]

2. Sie haben zwei Klassen und wollen die Entscheidungsgrenze zwischen diesen zwei Klassen berechnen. Das Trainingsset für jede Klasse besteht aus 4 zwei-dimensionalen Vektoren. Die Trainingsvektoren für Klasse 1 sind:

$$\mathbf{x}_{11} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \quad \mathbf{x}_{12} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \quad \mathbf{x}_{13} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \mathbf{x}_{14} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} \tag{1}$$

und die Trainingsvektoren für Klasse 2 sind:

$$\mathbf{x}_{21} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}, \quad \mathbf{x}_{22} = \begin{pmatrix} 7 \\ 3 \end{pmatrix}, \quad \mathbf{x}_{23} = \begin{pmatrix} 6 \\ 2 \end{pmatrix}, \quad \mathbf{x}_{24} = \begin{pmatrix} 6 \\ 4 \end{pmatrix} \tag{2}$$

(a) Berechnen Sie die geschätzten Mittelwertvektoren $\hat{\mu}_j$ für diese zwei Klassen. Schreiben Sie diese Vektoren in die unteren Rechtecke. (1 Punkt)

Mittelwertvektor $\hat{\mu}_1 = \boxed{}$ Mittelwertvektor $\hat{\mu}_2 = \boxed{}$

(b) Berechnen Sie die geschätzte Kovarianzmatrix $\hat{\Sigma}_1$ für Klasse 1. Schreiben Sie diese Matrix ins Rechteck auf der nächsten Seite. (1 Punkt)

Kovarianzmatrix $\hat{\Sigma}_1 = \boxed{}$

(c) Sie modellieren die Klassen durch zwei-dimensionale Gauß'sche Verteilungen:

$$p(\mathrm{x}|\omega_j) = \frac{1}{(2\pi)^{\frac{a}{2}} \left|\hat{\Sigma}_j\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathrm{x}-\hat{\mu}_\mathrm{j})^T \hat{\Sigma}_j^{-1}(\mathrm{x}-\hat{\mu}_\mathrm{j})\right] \tag{3}$$

wobei $a = 2$. Schreiben Sie die Diskriminanten-Funktion $g_j(\mathrm{x}) = \ln p(\omega_j|\mathrm{x})$ hier aus (die Gauß'sche Verteilung für $p(\mathrm{x}|\omega_j)$ soll auch in diesen Ausdruck hineingenommen werden). N.B. Der Ausdruck soll eine Funktion von $\hat{\mu}_\mathrm{j}$, $\hat{\Sigma}_j$, ω_j, $P(\omega_j)$, $p(\mathrm{x})$ und x sein. Klassen-spezifische Werte sollen nicht verwendet werden. (1 Punkt)

(d) Warum kann man die Terme: $-\ln 2\pi - \ln p(\mathrm{x})$ auslassen, ohne dass die Entscheidungsregionen und Entscheidungsgrenzen sich ändern? (1 Punkt)

(e) Sie nehmen an, dass $P(\omega_1) = P(\omega_2) = \frac{1}{2}$. Die Diskriminanten-Funktionen für die zwei Klassen schauen so aus:

$$g_1(\mathrm{x}) = -\frac{3}{4}x_1^2 + 3x_1 - \frac{3}{4}x_2^2 + \frac{9}{2}x_2 - \frac{39}{4} - \frac{1}{2}\ln\left(\frac{4}{9}\right) + \ln\left(\frac{1}{2}\right) \qquad (4)$$

$$g_2(\mathrm{x}) = -\frac{3}{4}x_1^2 + 9x_1 - \frac{3}{4}x_2^2 + \frac{9}{2}x_2 - \frac{135}{4} - \frac{1}{2}\ln\left(\frac{4}{9}\right) + \ln\left(\frac{1}{2}\right) \qquad (5)$$

wobei x_1 und x_2 die zwei Komponenten vom einem belibiegen Vektor x sind. Berechnen Sie den Ausdruck für die Entscheidungsgrenze. (1 Punkt)

3 Detaillierte Fragen

1. Dieser Datensatz soll für das Training eines Entscheidungsbaumes verwendet werden. Der Baum soll prognostizieren ob eine Person glücklich (G) oder unglücklich (U) ist, basiert auf der Farbe der Schuhe, ob eine Perücke getragen wird und der Anzahl von Ohren.

Farbe	Perücke	Anzahl von Ohren	Emotion
G	J	2	U
G	N	2	U
G	N	2	U
B	N	2	U
B	N	2	G
R	N	2	G
R	N	2	G
R	N	2	G
R	J	3	G

Zeichnen Sie einen vollständigen binären Entscheidungsbaum für diese Daten (Sie sollen annehmen, dass keine *Pruning* gemacht wird). (3 Punkte)

2. Wie wird Klassifikation anhand der Mahalanobis Distanz durchgeführt? Schreiben Sie den Ausdruck für die Mahalanobis Distanz. Wie werden die Parameter während der Trainingsphase geschätzt? Auf dem Diagramm sind Trainingsvektoren, die zu einer Klasse gehören, geplottet. Zeichnen Sie ein paar Konturlinien, die einem konstanten Wert für die Mahalanobis Distanz ab dieser Klasse entsprechen (die Konturlinien sollen ungefähr die richtige Form zeigen, aber müssen nicht genau sein). (3 Punkte)

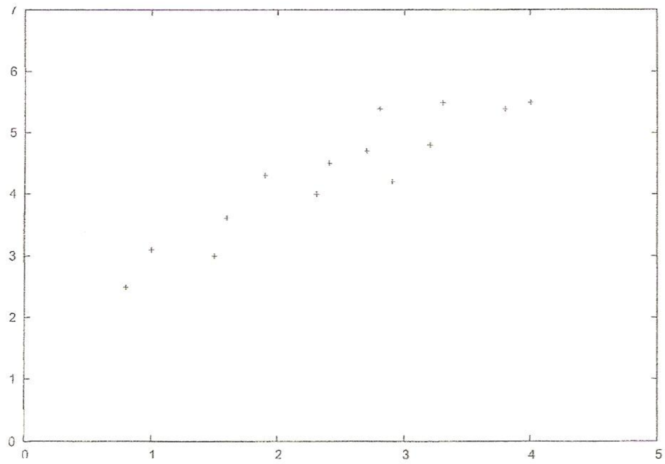
Trainingsphase: Schätzen Mittelwert u. Kovarianz

Dichtefunktion (Gauß pdf) eines normalverteilten Zufallsvektor $x \sim N(\mu, \Sigma)$ mit Mittelwert μ und Kovarianzmatrix Σ:

$$p(x) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

Mahalanobis Distanz

[illegible]


7
6
5
4
3
2
1
0
0
1
2
3
4
5

4 Aufsatz

Wählen Sie **eines** der folgenden Themen und schreiben Sie 150–200 Wörter darüber (6 Punkte).
Die Themen sind:

1. Parametrische und nicht-parametrische Verfahren für die Schätzung von Wahrscheinlichkeits-Dichtefunktionen.
2. Bias, Varianz und Generalisierungsfähigkeit.
3. Merkmalsextraktion in der Biometrie.

(1) Nicht parametrische Verfahren: keine Annahme [illegible] die pdf Funktion

Parametrische Verfahren: Annahme über die Form der pdf Funktion
z.B. Gaussian Kurve
d.h. Schätzen die Parameter der Kurve

Vorteil: Weil wir die Form der pdf kennen → weniger Parameter zu schätzen und oft kann [illegible] Trainingsset ausreichend

Nachteil: nicht so genau

(3) PCA Gesichtserkennung

Elastic graph matching: Ein Graph mit Merkmalen in jedem Knoten gespeichert ist als Modell eines Gesichts verwendet
Dieser Graph wird an Bilder von Gesichtern angepasst um einen Bildgraph zu erzeugen. Ähnlichkeit zwischen dem Bildgraph und Modellen [illegible]