19. Juni 2008	**Prüfung aus Einführung in die Mustererkennung — SS 2008**	16.00 – 17.30
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 in Englisch oder Deutsch beantworten. **Es sind keine Unterlagen erlaubt.**

1 Kurze Fragen

1. Ist das CART Lernverfahren ein überwachtes oder unüberwachtes Lernverfahren? Die nicht-passende Antwort durchstreichen. (1 Punkt)

   Überwachtes Verfahren / Unüberwachtes Verfahren

   Begründung:

2. Ist die Klassifikation mittels ein Perceptron ein parametrisches oder nicht-parametrisches Verfahren? Die nicht-passende Antwort durchstreichen. (1 Punkt)

   Parametrisches Verfahren / Nicht-parametrisches Verfahren

   Begründung:

3. Die zwei Zufallsvariablen A und B sind statistisch voneinander *unabhängig*. Welcher der folgenden Ausdrücke ist **FALSCH**? Schreiben Sie die richtige Antwort ins untere Rechteck. (1 Punkt)

   (a) $P(A, B) = P(A|B)P(B)$

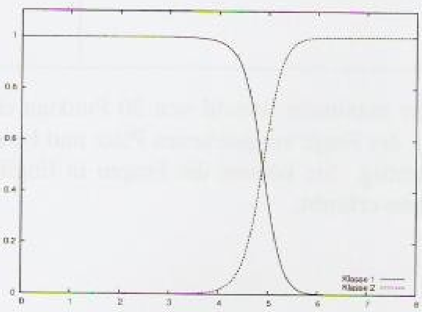
   (b) $P(A, B) = P(B|A)P(A)$

   (c) $P(A|B) = P(A)$

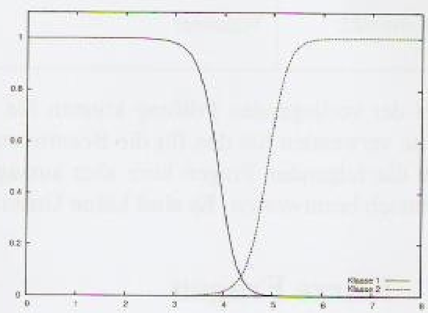
   (d) $P(A|B) = P(B)$

   Dieser Ausdruck ist FALSCH: $\boxed{}$

4. Welches der folgenden Kurvenpaare stellt *a posteriori* Wahrscheinlichkeiten für ein Klassifikationsproblem mit 2 Klassen und einem Merkmal dar? (1 Punkt)



(a)

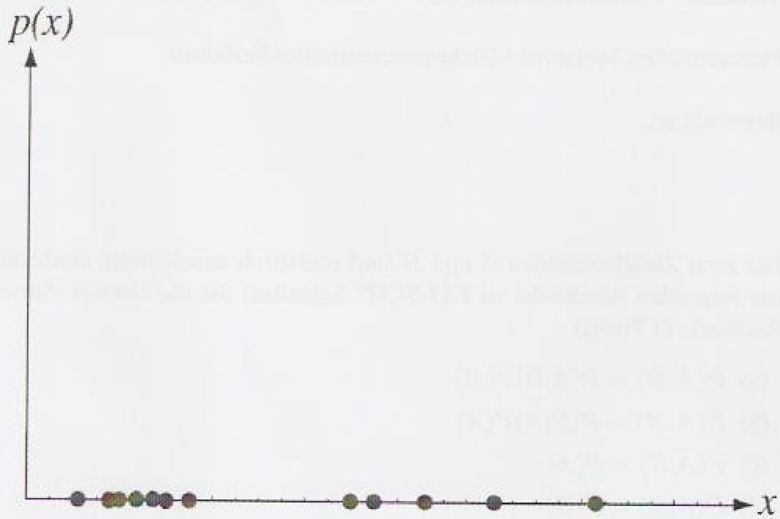


(b)

(a) oder (b): $\boxed{}$ (stellt *a posteriori* Wahrschlichkeiten dar.)

Begründung:

5. Zeichnen Sie eine mögliche *probability density function* (pdf) für die Merkmalsausprägungen x unten (sie gehören alle zu einer Klasse). (1 Punkt)



6. Welche der folgenden Funktionen können in einem Backpropagation neuronalen Netz als Aktivierungsfunktion verwendet werden? Die nicht-passende Antwort(en) durchstreichen. (1 Punkt)

   (a) tanh Funktion: $y = \tanh(x)$

   (b) Signum-Funktion: $y = \begin{cases} 1 & \text{wenn} \quad x > 0 \\ 0 & \text{wenn} \quad x = 0 \\ -1 & \text{wenn} \quad x < 0 \end{cases}$

   (c) sigmoid Funktion: $y = \frac{1}{1+\exp(-x)}$

7. Wenn PCA für Gesichtserkennung eingesetzt wird, heißen die Eigenvektoren oft

   ______________________________ (1 Punkt)

2 Rechenbeispiele

1. Der *Formfaktor* von einem 2-dimensionalen Object S ist so definiert: $\frac{4\pi A(S)}{[P(S)]^2}$, wobei $A(S)$ die Fläche und $P(S)$ der Umfang vom Object S sind. Schreiben sie den Wert vom Formfactor für: (1 Punkt)

   (a) einen Kreis mit Radius r.

   (b) ein Quadrat mit Kantenlänge ℓ.

2. Sie bekommen den Auftrag, eine Maschine zu entwickeln, die automatisch zwischen zwei Fischarten, Forelle (Klasse 1, $\omega = 1$) und Karpfen (Klasse 2, $\omega = 2$), unterscheidet. Ein diskretes Helligkeitsmerkmal x, das 3 Werte annehmen kann, wird für jeden Fisch 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	0.1	0.1
$P(x = i\|\omega = 2)$	0.1	0.4	0.5

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

i	1	2	3
$P(x = i)$			

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

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

j	1	2
$P(\omega = j \vert x = 1)$		
$P(\omega = j \vert x = 2)$		
$P(\omega = j \vert x = 3)$		

(d) Welche Klasse soll laut Bayes-Entscheidungsregel 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)	$P(error \vert x)$
$x = 1$		
$x = 2$		
$x = 3$		

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

Fehlerrate:

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

3. 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:

$$\mathbf{x}_{11} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \quad \mathbf{x}_{12} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \quad \mathbf{x}_{13} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \quad \mathbf{x}_{14} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \tag{1}$$

und die Trainingsvektoren für Klasse 2 sind:

$$\mathbf{x}_{21} = \begin{pmatrix} 0 \\ 4 \end{pmatrix}, \quad \mathbf{x}_{22} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \quad \mathbf{x}_{23} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \quad \mathbf{x}_{24} = \begin{pmatrix} 1 \\ 5 \end{pmatrix} \tag{2}$$

(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:

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

wobei $a = 2$. Sie nehmen an, dass $P(\omega_1) = P(\omega_2) = \frac{1}{2}$. Sie verwenden als Diskriminanten-Funktion $g_j(\mathbf{x}) = \ln p(\omega_j|\mathbf{x})$. Die Klassen-spezifische Diskriminanten-Funktionen schauen so aus:

$$g_1(\mathbf{x}) = -\frac{3}{4}x_1^2 + 6x_1 - \frac{3}{4}x_2^2 + \frac{3}{2}x_2 - \frac{51}{4} - \frac{1}{2}\ln\left(\frac{4}{9}\right) \quad (4)$$

$$g_2(\mathbf{x}) = -\frac{3}{4}x_1^2 + \frac{3}{2}x_1 - \frac{3}{4}x_2^2 + 6x_2 - \frac{51}{4} - \frac{1}{2}\ln\left(\frac{4}{9}\right) \quad (5)$$

wobei x_1 und x_2 die zwei Komponenten vom einem beliebiegen Vektor $\mathbf{x}$ sind. Berechnen Sie den Ausdruck für die Entscheidungsgrenze. (1 Punkt)

(d) Begründen Sie mit Hilfe einer graphischen Darstellung, warum die von Ihnen berechnete Entscheidungsgrenze richtig ist. (1 Punkt)

3 Detaillierte Fragen

1. Was bedeutet die *impurity* (Unreinheit) eines Knotens in einem Entscheidungsbaum? Warum ist die *entropy impurity*

$$i(N) = -\sum_j P(\omega_j) \log_2 P(\omega_j) \tag{6}$$

ein geeignetes Maß für die *impurity* eines Knotens? Besprechen Sie nur den Fall mit einer Entscheidung zwischen zwei Klassen. (Zur Erinnerung, $P(\omega_j)$ gibt den Anteil der Muster in Knoten N, die zur Klasse ω_j gehören, an). (2 Punkte)

2. Wie wird ein Klassifikator als eine Gruppe von Diskriminantenfunktionen dargestellt? Was sind *Entscheidungsregionen* und *Entscheidungsgrenzen*? Wie werden diese anhand der Diskriminantenfunktionen definiert? Zeichnen Sie ein 2-dimensionnelles Beispiel mit 3 Klassen worin diese Definitionen klar erklärt werden. (3 Punkte)

4 Aufsatz

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

1. Gesichtserkennung mit PCA.
2. Bias, Varianz und Generalisierungsfähigkeit.
3. Parametrische und nicht-parametrische Lernverfahren: Algorithmen, Vorteile und Nachteile.