Data Mining / Intelligente Datenanalyse WS 2017/2018
PD Dr.-Ing. habil. Christian Borgelt

Lösung des 4. Übungsblattes

Aufgabe 21 Naiver Bayes-Klassifikator

Um aus dem gegebenen Datensatz einen naiven Bayes-Klassifikator zu bestimmen, schätzt man die klassenspezifischen Mittelwerte und Varianzen, sowie die A-priori-Wahrscheinlichkeit der beiden Klassen. Da die beiden Klassen in dem gegebenen Datensatz gleich häufig auftreten, ist letztere für beide Klassen 0.5. Für die Mittelwerte ergibt sich:

$$\begin{aligned}
\mu_{a,x} &= \tfrac{1}{10}(3+3+4+4+5+6+7+7+8+9) &&= 5.6\\
\mu_{a,y} &= \tfrac{1}{10}(1+2+2+3+3+4+4+6+5+7) &&= 3.7\\
\mu_{b,x} &= \tfrac{1}{10}(1+2+2+3+4+5+5+6+7+7) &&= 4.2\\
\mu_{b,y} &= \tfrac{1}{10}(3+4+5+6+6+7+8+8+8+9) &&= 6.4
\end{aligned}$$

und für die Varianzen (erwartungstreuer Schätzer):

$$\begin{aligned}
\sigma^2_{a,x} &= \tfrac{1}{9}(3^2+3^2+4^2+4^2+5^2+6^2+7^2+7^2+8^2+9^2-10\cdot 5.6^2) &&\approx 4.49\\
\sigma^2_{a,y} &= \tfrac{1}{9}(1^2+2^2+2^2+3^2+3^2+4^2+4^2+6^2+5^2+7^2-10\cdot 3.7^2) &&\approx 3.57\\
\sigma^2_{b,x} &= \tfrac{1}{9}(1^2+2^2+2^2+3^2+4^2+5^2+5^2+6^2+7^2+7^2-10\cdot 4.2^2) &&\approx 4.62\\
\sigma^2_{b,y} &= \tfrac{1}{9}(3^2+4^2+5^2+6^2+6^2+7^2+8^2+8^2+8^2+9^2-10\cdot 6.4^2) &&\approx 3.82
\end{aligned}$$

Die Parameter des so erhaltenen Klassifikators stellen wir noch einmal übersichtlich dar:

Klasse	p	μ	Σ
a	0.5	$\begin{pmatrix} 5.6 \\ 3.7 \end{pmatrix}$	$\begin{pmatrix} 4.49 & 0 \\ 0 & 3.57 \end{pmatrix}$
b	0.5	$\begin{pmatrix} 4.2 \\ 6.4 \end{pmatrix}$	$\begin{pmatrix} 4.62 & 0 \\ 0 & 3.82 \end{pmatrix}$

Man beachte hier, daß die Varianzen in Kovarianzmatrizen eingetragen sind, bei denen nur die Diagonalelemente verschieden von 0 sind, da die (naive) Annahme der bedingten Unabhängigkeit der beschreibenden Attribute für numerische Attribute der Annahme verschwindender Kovarianzen entspricht.

Um die beiden zusätzlich gegebenen Fälle zu klassifizieren, setzen wir die Werte der Attribute x und y für diese beiden Fälle in die Klassifikationsformel ein:

$$\begin{aligned}
&P(C=a \mid X=8 \wedge Y=7)\\
&= \frac{1}{p_1}\cdot P(C=a)\cdot f(X=8\mid C=a)\cdot f(Y=7\mid C=a)\\
&= \frac{1}{p_1}\cdot P(C=a)\cdot \frac{1}{\sqrt{2\pi\cdot\sigma^2_{a,x}}}\exp\left(-\frac{(8-\mu_{a,x})^2}{2\cdot\sigma^2_{a,x}}\right)\cdot\frac{1}{\sqrt{2\pi\cdot\sigma^2_{a,y}}}\exp\left(-\frac{(7-\mu^2_{a,y})^2}{2\cdot\sigma^2_{a,y}}\right)\\
&= \frac{1}{p_1}\cdot 0.5\cdot\frac{1}{\sqrt{2\pi\cdot 4.49}}\exp\left(-\frac{(8-5.6)^2}{2\cdot 4.49}\right)\cdot\frac{1}{\sqrt{2\pi\cdot 3.57}}\exp\left(-\frac{(7-3.7)^2}{2\cdot 3.57}\right)\\
&\approx \frac{1}{p_1}\cdot 0.00228
\end{aligned}$$

$$
\begin{aligned}
&P(C = b \mid X = 8 \wedge Y = 7) \\
&= \frac{1}{p_1} \cdot P(C = b) \cdot f(X = 8 \mid C = b) \cdot f(Y = 7 \mid C = b) \\
&= \frac{1}{p_1} \cdot P(C = b) \cdot \frac{1}{\sqrt{2\pi \cdot \sigma_{b,x}^2}} \exp\left(-\frac{(8 - \mu_{b,x})^2}{2 \cdot \sigma_{b,x}^2}\right) \cdot \frac{1}{\sqrt{2\pi \cdot \sigma_{b,y}^2}} \exp\left(-\frac{(7 - \mu_{b,y}^2)^2}{2 \cdot \sigma_{b,y}^2}\right) \\
&= \frac{1}{p_1} \cdot 0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 4.62}} \exp\left(-\frac{(8 - 4.2)^2}{2 \cdot 4.62}\right) \cdot \frac{1}{\sqrt{2\pi \cdot 3.82}} \exp\left(-\frac{(7 - 6.4)^2}{2 \cdot 3.82}\right) \\
&\approx \frac{1}{p_1} \cdot 0.00379
\end{aligned}
$$

$$
\begin{aligned}
&P(C = a \mid X = 3 \wedge Y = 4) \\
&= \frac{1}{p_2} \cdot P(C = a) \cdot f(X = 3 \mid C = a) \cdot f(Y = 4 \mid C = a) \\
&= \frac{1}{p_2} \cdot 0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 4.49}} \exp\left(-\frac{(3 - 5.6)^2}{2 \cdot 4.49}\right) \cdot \frac{1}{\sqrt{2\pi \cdot 3.57}} \exp\left(-\frac{(4 - 3.7)^2}{2 \cdot 3.57}\right) \\
&\approx \frac{1}{p_2} \cdot 0.00925
\end{aligned}
$$

$$
\begin{aligned}
&P(C = b \mid X = 3 \wedge Y = 4) \\
&= \frac{1}{p_2} \cdot P(C = b) \cdot f(X = 3 \mid C = b) \cdot f(Y = 4 \mid C = b) \\
&= \frac{1}{p_2} \cdot 0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 4.62}} \exp\left(-\frac{(3 - 4.2)^2}{2 \cdot 4.62}\right) \cdot \frac{1}{\sqrt{2\pi \cdot 3.82}} \exp\left(-\frac{(4 - 6.4)^2}{2 \cdot 3.82}\right) \\
&\approx \frac{1}{p_2} \cdot 0.00763
\end{aligned}
$$

Die Normierungsfaktoren p_1 und p_2 werden aus

$$
\begin{aligned}
&P(C = a \mid X = 8 \wedge Y = 7) + P(C = b \mid X = 8 \wedge Y = 7) = \tfrac{1}{p_1}(0.00278 + 0.00379) = 1 \quad \text{und} \\
&P(C = a \mid X = 3 \wedge Y = 4) + P(C = b \mid X = 3 \wedge Y = 4) = \tfrac{1}{p_2}(0.00925 + 0.00763) = 1
\end{aligned}
$$

bestimmt. Man erhält dann die Wahrscheinlichkeiten

Klasse	Tupel 1	Tupel 2
a	0.375	0.548
b	0.625	0.452

Tupel 1 wird also als zur Klasse b, Tupel 2 als zur Klasse a gehörend klassifiziert (Wahl der Klasse mit der größeren Wahrscheinlichkeit). Dies ist nicht sehr plausibel, da man wegen der Lage der Tupel des Datensatzes eher eine Zuordnung des Tupels 1 zur Klasse a und des Tupels 2 zur Klasse b erwartet (siehe Diagramme zu Aufgabe 27). Das Problem besteht offenbar darin, daß die Annahme der bedingten Unabhängigkeit der beiden beschreibenden Attribute x und y gegeben die Klasse für den gegebenen Datensatz nicht erfüllt ist, denn durch achsenparallele Normalverteilungen lassen sich die Klassenverteilungen nicht angemessen beschreiben.

Aufgabe 22 Voller Bayes-Klassifikator

Um aus dem gegebenen Datensatz einen vollen Bayes-Klassifikator zu bestimmen, müssen wir (zusätzlich zu den Parametern des bereits in Aufgabe 25 berechneten naiven Bayes-Klassifikators) lediglich die beiden klassenspezifischen Kovarianzen schätzen:

$$
\begin{aligned}
\sigma^2_{a,xy} &= \tfrac{1}{9}(3\cdot 1 + 3\cdot 2 + 4\cdot 2 + 4\cdot 3 + 5\cdot 3 + 6\cdot 4 + 7\cdot 4 + 7\cdot 6 + 8\cdot 5 + 9\cdot 7 - 10\cdot 5.6\cdot 3.7)\\
&\approx 3.76\\
\sigma^2_{b,xy} &= \tfrac{1}{9}(1\cdot 3 + 2\cdot 4 + 2\cdot 5 + 3\cdot 6 + 4\cdot 6 + 5\cdot 7 + 5\cdot 8 + 6\cdot 8 + 7\cdot 8 + 7\cdot 9 - 10\cdot 4.2\cdot 6.4)\\
&\approx 4.02
\end{aligned}
$$

Die Parameter des Klassifikators stellen wir noch einmal übersichtlich dar, wobei wir auch gleich die für die Auswertung des Klassifikators benötigten Determinanten und Inversen der Kovarianzmatrizen angeben.

Klasse	p	μ	Σ	$\lvert\Sigma\rvert$	Σ^{-1}
a	0.5	$\begin{pmatrix} 5.6 \\ 3.7 \end{pmatrix}$	$\begin{pmatrix} 4.49 & 3.76 \\ 3.76 & 3.57 \end{pmatrix}$	1.91	$\begin{pmatrix} 1.87 & -1.97 \\ -1.97 & 2.35 \end{pmatrix}$
b	0.5	$\begin{pmatrix} 4.2 \\ 6.4 \end{pmatrix}$	$\begin{pmatrix} 4.62 & 4.02 \\ 4.02 & 3.82 \end{pmatrix}$	1.49	$\begin{pmatrix} 2.57 & -2.70 \\ -2.70 & 3.10 \end{pmatrix}$

Um die beiden zusätzlich gegebenen Fälle zu klassifizieren, setzen wir wieder die Werte der Attribute x und y für diese beiden Fälle in die Klassifikationsformel ein:

$$
\begin{aligned}
&P(C = a \mid X = 8 \wedge Y = 7)\\
&\quad = \frac{1}{p_1} \cdot P(C = a) \cdot f(X = 8 \wedge Y = 7 \mid C = a)\\
&\quad = \frac{1}{p_1} \cdot P(C = a) \cdot \frac{1}{2\pi \cdot \sqrt{|\Sigma_a|}} \exp\left(-\frac{1}{2} \cdot (8 - \mu_{a,x}, 7 - \mu_{a,y})\ \Sigma_a^{-1} \begin{pmatrix} 8 - \mu_{a,x} \\ 7 - \mu_{a,y} \end{pmatrix}\right)\\
&\quad = \frac{1}{p_1} \cdot 0.5 \cdot \frac{1}{2\pi \cdot \sqrt{1.91}} \exp\left(-\frac{1}{2} \cdot (8 - 5.6, 7 - 3.7) \begin{pmatrix} 1.87 & -1.97 \\ -1.97 & 2.35 \end{pmatrix} \begin{pmatrix} 8 - 5.6 \\ 7 - 3.7 \end{pmatrix}\right)\\
&\quad \approx \frac{1}{p_1} \cdot 0.00425
\end{aligned}
$$

$$
\begin{aligned}
&P(C = b \mid X = 8 \wedge Y = 7)\\
&\quad = \frac{1}{p_1} \cdot P(C = b) \cdot f(X = 8 \wedge Y = 7 \mid C = b)\\
&\quad = \frac{1}{p_1} \cdot P(C = b) \cdot \frac{1}{2\pi \cdot \sqrt{|\Sigma_b|}} \exp\left(-\frac{1}{2} \cdot (8 - \mu_{b,x}, 7 - \mu_{b,y})\ \Sigma_b^{-1} \begin{pmatrix} 8 - \mu_{b,x} \\ 7 - \mu_{b,y} \end{pmatrix}\right)\\
&\quad = \frac{1}{p_1} \cdot 0.5 \cdot \frac{1}{2\pi \cdot \sqrt{1.49}} \exp\left(-\frac{1}{2} \cdot (8 - 4.3, 7 - 6.4) \begin{pmatrix} 2.57 & -2.70 \\ -2.70 & 3.10 \end{pmatrix} \begin{pmatrix} 8 - 4.3 \\ 7 - 6.4 \end{pmatrix}\right)\\
&\quad \approx \frac{1}{p_1} \cdot 1.18 \cdot 10^{-7}
\end{aligned}
$$

$$
\begin{aligned}
&P(C = a \mid X = 3 \wedge Y = 4)\\
&\quad = \frac{1}{p_2} \cdot P(C = a) \cdot f(X = 3 \wedge Y = 4 \mid C = a)\\
&\quad = \frac{1}{p_2} \cdot 0.5 \cdot \frac{1}{2\pi \cdot \sqrt{1.91}} \exp\left(-\frac{1}{2} \cdot (3 - 5.6, 4 - 3.7) \begin{pmatrix} 1.87 & -1.97 \\ -1.97 & 2.35 \end{pmatrix} \begin{pmatrix} 3 - 5.6 \\ 4 - 3.7 \end{pmatrix}\right)\\
&\quad \approx \frac{1}{p_2} \cdot 2.00 \cdot 10^{-5}
\end{aligned}
$$

$$
\begin{aligned}
&P(C = b \mid X = 3 \wedge Y = 4) \\
&= \frac{1}{p_2} \cdot P(C = b) \cdot f(X = 3 \wedge Y = 4 \mid C = b) \\
&= \frac{1}{p_2} \cdot 0.5 \cdot \frac{1}{2\pi \cdot \sqrt{1.49}} \exp\left(-\frac{1}{2} \cdot (3 - 4.3, 4 - 6.4) \begin{pmatrix} 2.57 & -2.70 \\ -2.70 & 3.10 \end{pmatrix} \begin{pmatrix} 3 - 4.3 \\ 4 - 6.4 \end{pmatrix} \right) \\
&\approx \frac{1}{p_2} \cdot 0.00322
\end{aligned}
$$

Die Normierungsfaktoren p_1 und p_2 werden (wie oben) aus

$$
\begin{aligned}
P(C = a \mid X = 8 \wedge Y = 7) + P(C = b \mid X = 8 \wedge Y = 7) &= \tfrac{1}{p_1}(0.00425 + 1.18 \cdot 10^{-7}) = 1, \\
P(C = a \mid X = 3 \wedge Y = 4) + P(C = b \mid X = 3 \wedge Y = 4) &= \tfrac{1}{p_2}(2.00 \cdot 10^{-5} + 0.00322) = 1
\end{aligned}
$$

bestimmt. Dies führt zu den Wahrscheinlichkeiten

Klasse	Tupel 1	Tupel 2
a	1	0.006
b	0	0.994

Man erhält also eine (fast) eindeutige Zuordnung des Tupels 1 zur Klasse a und des Tupels 2 zur Klasse b, wie man sie auch bei Betrachtung der Lage der Tupel relativ zu den Trainingstupeln erwartet (siehe Diagramme in Aufgabe 27).

Aufgabe 23 Visualisierung von Bayes-Klassifikatoren

Der naive Bayes-Klassifikator aus Aufgabe 25 läßt sich sehr leicht visualisieren. Durch ihn werden zweidimensionale Normalverteilungen beschrieben, die achsenparallel liegen (d.h. die Hauptachsen der $k\sigma$-Ellipsen sind parallel zu den Koordinatenachsen). Um den Klassifikator zu visualisieren, zeichnet man das Zentrum (Erwartungswerte) der beiden Verteilungen und z.B. die $\sigma-$ und die 2σ-Ellipse. Dazu trägt man vom Zentrum aus die Standardabweichung (Quadratwurzel der Varianz) für X in horizontaler und für Y in vertikaler Richtung ab. Die so erhaltenen vier Punkte zeigen bereits hinreichend deutlich die Lage der Ellipse (siehe Abbildung 1 rechts, in der die Punkte markiert sind).

Um den vollen Bayes-Klassifikator durch Angabe von σ- und 2σ-Ellipsen zu visualisieren, muß man zunächst aus der Kovarianzmatrix Σ „die Wurzel ziehen", d.h., eine untere bzw. linke Dreiecksmatrix L bestimmen, so daß $\Sigma = LL^T$. Eine solche Matrix L erhält man z.B. mit Hilfe der in der Vorlesung angegebenen Cholesky-Zerlegung, die für 2×2-Matrizen

$$
\Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \qquad \text{definiert ist als} \qquad L = \begin{pmatrix} \sigma_x & 0 \\ \dfrac{\sigma_{xy}}{\sigma_x} & \dfrac{1}{\sigma_x}\sqrt{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2} \end{pmatrix}.
$$

In der vorliegenden Aufgabe ergibt sich damit

$$
L_a = \begin{pmatrix} 2.12 & 0 \\ 1.77 & 0.65 \end{pmatrix} \qquad \text{und} \qquad L_b = \begin{pmatrix} 2.15 & 0 \\ 1.87 & 0.57 \end{pmatrix}.
$$

Wie in der Vorlesung erläutert, beschreiben diese Matrizen (zusammen mit einer Verschiebung um den Vektor zum Zentrum der Verteilung) Abbildungen, durch die ein Einheitskreis um den Ursprung auf die σ-Ellipse der Verteilung abgebildet wird. Entsprechend wird ein Kreis mit dem Radius k auf die $k\sigma$-Ellipse abgebildet. Erinnert man sich nun noch an einen wichtigen

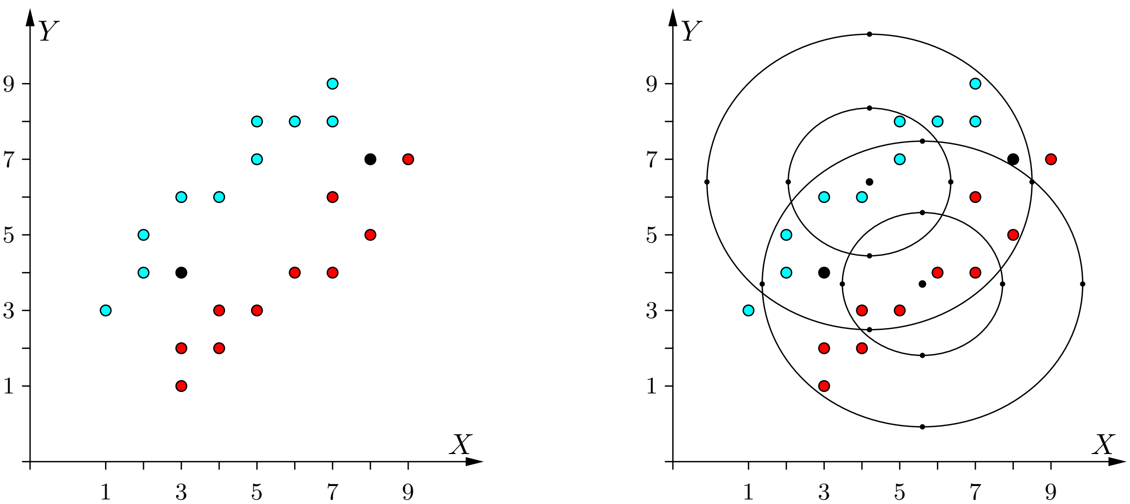


Abbildung 1: Daten und naiver Bayes-Klassifikator

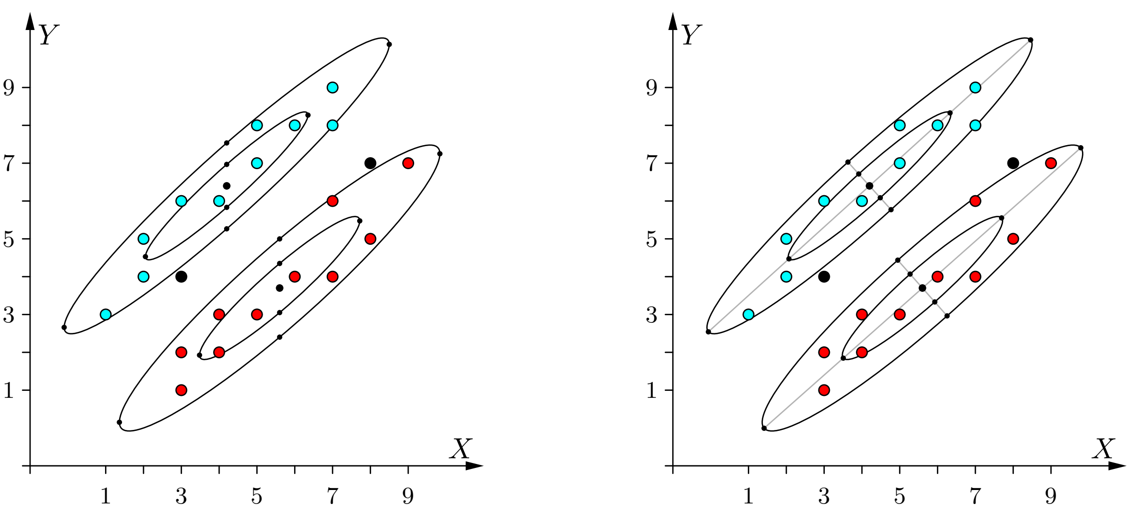


Abbildung 2: Voller Bayes-Klassifikator (links: Cholesky-, rechts: Eigenwertzerlegung).

Merksatz der Matrizenrechnung, nämlich: „Die Spalten sind die Bilder der Einheitsvektoren.“, so kann man leicht die Koordinaten von vier Punkten je Ellipse bestimmen:

$$
\begin{aligned}
p_{1a} &= \begin{pmatrix} 5.6 \\ 3.7 \end{pmatrix} + \begin{pmatrix} 2.12 \\ 1.77 \end{pmatrix} = \begin{pmatrix} 7.72 \\ 5.47 \end{pmatrix}, \qquad & p_{2a} &= \begin{pmatrix} 5.6 \\ 3.7 \end{pmatrix} + \begin{pmatrix} 0 \\ 0.65 \end{pmatrix} = \begin{pmatrix} 5.6 \\ 4.35 \end{pmatrix}, \\
p_{3a} &= \begin{pmatrix} 5.6 \\ 3.7 \end{pmatrix} - \begin{pmatrix} 2.12 \\ 1.77 \end{pmatrix} = \begin{pmatrix} 3.48 \\ 1.93 \end{pmatrix}, \qquad & p_{4a} &= \begin{pmatrix} 5.6 \\ 3.7 \end{pmatrix} - \begin{pmatrix} 0 \\ 0.65 \end{pmatrix} = \begin{pmatrix} 5.6 \\ 3.05 \end{pmatrix}, \\
p_{1b} &= \begin{pmatrix} 4.2 \\ 6.4 \end{pmatrix} + \begin{pmatrix} 2.15 \\ 1.87 \end{pmatrix} = \begin{pmatrix} 6.35 \\ 8.27 \end{pmatrix}, \qquad & p_{2b} &= \begin{pmatrix} 4.2 \\ 6.4 \end{pmatrix} + \begin{pmatrix} 0 \\ 0.57 \end{pmatrix} = \begin{pmatrix} 4.2 \\ 6.97 \end{pmatrix}, \\
p_{3b} &= \begin{pmatrix} 4.2 \\ 6.4 \end{pmatrix} - \begin{pmatrix} 2.15 \\ 1.87 \end{pmatrix} = \begin{pmatrix} 2.05 \\ 4.53 \end{pmatrix}, \qquad & p_{4b} &= \begin{pmatrix} 4.2 \\ 6.4 \end{pmatrix} - \begin{pmatrix} 0 \\ 0.57 \end{pmatrix} = \begin{pmatrix} 4.2 \\ 5.83 \end{pmatrix}.
\end{aligned}
$$

Entsprechend für die 2σ-Ellipsen. Aus diesen Punkten läßt sich die Lage der Ellipsen bereits recht gut ablesen (siehe Abbildung 2 links, in der die Punkte markiert sind).

Eine andere Möglichkeit, eine „Wurzel“ aus der Kovarianzmatrix zu ziehen, besteht darin, eine Eigenwertzerlegung dieser Matrix vorzunehmen. Mit dem auf den Folien verwendeten Verfahren der Jacobi-Transformation:

$$\begin{aligned}
\phi &= \tfrac{1}{2}\arctan\tfrac{2\sigma_{xy}}{\sigma_x^2-\sigma_y^2},\\
\sigma_1 &= \sqrt{\cos^2\phi\,\sigma_x^2+\sin^2\phi\,\sigma_y^2+2\sin\phi\cos\phi\,\sigma_{xy}},\\
\sigma_2 &= \sqrt{\sin^2\phi\,\sigma_x^2+\cos^2\phi\,\sigma_y^2-2\sin\phi\cos\phi\,\sigma_{xy}},\\
\mathbf{T} &= \begin{pmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{pmatrix}\begin{pmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{pmatrix},
\end{aligned}$$

erhält man so:

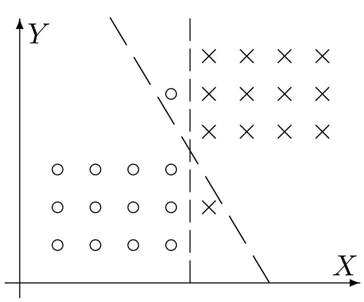
$$\begin{aligned}
\phi_a &= 41.5^\circ, & \phi_b &= 42.2^\circ,\\
\sigma_{1a} &= 2.796, & \sigma_{1a} &= 2.874,\\
\sigma_{2a} &= 0.492, & \sigma_{2b} &= 0.424,\\
\mathbf{T}_a &= \begin{pmatrix}2.09 & -0.33\\ 1.85 & 0.37\end{pmatrix}, & \mathbf{T}_b &= \begin{pmatrix}2.13 & -0.28\\ 1.93 & 0.31\end{pmatrix}.
\end{aligned}$$

Die Bilder der Einheitsvektoren sind damit

$$\begin{aligned}
p_{1a} &= \begin{pmatrix}5.6\\3.7\end{pmatrix}+\begin{pmatrix}2.09\\1.85\end{pmatrix}=\begin{pmatrix}7.69\\5.55\end{pmatrix}, & p_{2a} &= \begin{pmatrix}5.6\\3.7\end{pmatrix}+\begin{pmatrix}-0.33\\0.37\end{pmatrix}=\begin{pmatrix}5.27\\4.07\end{pmatrix},\\
p_{3a} &= \begin{pmatrix}5.6\\3.7\end{pmatrix}-\begin{pmatrix}2.09\\1.85\end{pmatrix}=\begin{pmatrix}3.51\\1.85\end{pmatrix}, & p_{4a} &= \begin{pmatrix}5.6\\3.7\end{pmatrix}-\begin{pmatrix}-0.33\\0.37\end{pmatrix}=\begin{pmatrix}5.93\\3.33\end{pmatrix},\\
p_{1b} &= \begin{pmatrix}4.2\\6.4\end{pmatrix}+\begin{pmatrix}2.13\\1.93\end{pmatrix}=\begin{pmatrix}6.33\\8.33\end{pmatrix}, & p_{2b} &= \begin{pmatrix}4.2\\6.4\end{pmatrix}+\begin{pmatrix}-0.28\\0.31\end{pmatrix}=\begin{pmatrix}3.92\\6.71\end{pmatrix},\\
p_{3b} &= \begin{pmatrix}4.2\\6.4\end{pmatrix}-\begin{pmatrix}2.13\\1.93\end{pmatrix}=\begin{pmatrix}2.07\\4.47\end{pmatrix}, & p_{4b} &= \begin{pmatrix}4.2\\6.4\end{pmatrix}-\begin{pmatrix}-0.28\\0.31\end{pmatrix}=\begin{pmatrix}4.48\\6.09\end{pmatrix}.
\end{aligned}$$

Entsprechend für die 2σ-Ellipsen. Aus diesen Punkten läßt sich die Lage der Ellipsen sehr leicht ablesen (siehe Abbildung 2 rechts, in der die Punkte markiert sind).

Aufgabe 24 Bayes-Klassifikatoren: Nutzen von Attributen



Weder für naive noch für volle Bayes-Klassifikatoren gilt, daß die Zahl der Fehlklassifikationen auf den Trainingsdaten bei Hinzufügen weiterer Attribute nur sinken kann. Bei ungünstiger Lage der Datenpunkte kann sie auch zunehmen. Man betrachte dazu das in dem nebenstehenden Diagramm gezeigte Beispiel. Wird nur das Attribut X verwendet, so können die beiden Klassen (markiert durch $\circ$ und $\times$) perfekt getrennt werden: Die vertikale, gestrichelte Linie ist die sowohl von einem naiven als auch von einem vollen Bayes-Klassifikator (die sich ja im eindimensionalen Fall nicht unterscheiden) berechnete Klassengrenze. Werden dagegen beide Attribute X und Y verwendet, so berechnet sowohl ein naiver als auch ein voller Bayes-Klassifikator auf der linken Seite der schrägen, gestrichelten Linie eine Zugehörigkeit zur Klasse $\circ$ und auf der rechten Seite eine Zugehörigkeit zur Klasse $\times$ (da die Unabhängigkeitsannahme hier in sehr guter Näherung erfüllt ist, unterscheiden sich naiver und voller Bayes-Klassifikator kaum). Dadurch kommt es jedoch zu zwei Fehlklassifikationen.