Aufgabe 3

Betrachten Sie folgendes vereinfachtes Behandlungsproblem: Ein Patient klagt über Symptome, die entweder von einer vergleichsweise harmlosen Entzündung (Umweltzustand θ_1) oder von einer schweren Erkrankung stammen (Umweltzustand θ_2). Der Arzt habe zudem zwei Behandlungsmöglichkeiten: eine leichte symptombezogene Behandlung (Aktion a_1) oder eine spezifische Behandlung, bezogen auf die schwere Krankheit (Aktion a_2). Das Problem führe auf folgende Verlusttafel:

	θ_1	θ_2
a_1	0	30
a_2	10	0

Zur Unterstützung seiner Entscheidung führt der Arzt einen medizinischen Test durch. Dieser habe nur zwei verschiedene potentielle Ergebnisse: „positiv", d.h. Verdacht auf schwere Krankheit, und „negativ", also kein Verdacht. Ferner sei bekannt, dass der medizinische Test eine Sensitivität (Wahrscheinlichkeit eines positiven Testergebnisses bei Vorliegen der schweren Erkrankung) von 90% besitzt, sowie eine Spezifität (Wahrscheinlichkeit eines negativen Testergebnis bei Vorliegen von nur der leichten Erkrankung) von 80%.

(a) Formalisieren Sie dieses Problem als datengestütztes Entscheidungsproblem und skizzieren Sie den Entscheidungsbaum!

(b) Bestimmen Sie das zugehörige Auswertungsproblem!

Lösung Aufgabe 3

Verlusttafel des zugrundeliegenden datenfreien Entscheidungsproblems $(\mathbb{A}, \Theta, \ell)$:

	θ_1	θ_2
a_1	0	30
a_2	10	0

Statistisches Modell für medizinischen Test $(\mathcal{X}, \sigma(\mathcal{X}), (P_\theta)_{\theta\in\Theta})$ mit $\mathcal{X} = \{x_1, x_2\}$, wobei x_1 bedeutet, dass der Test ein negatives Ergebnis anzeigt und x_2 bedeutet, dass der Test auf die schwere Krankheit hindeutet, sowie $\sigma(\mathcal{X}) = 2^{\mathcal{X}}$ (Potenzmenge). Die Wahrscheinlichkeitsverteilungen $(P_\theta)_{\theta\in\Theta} = \{P_{\theta_1}, P_{\theta_2}\}$ sind bestimmt durch:

	θ_1	θ_2
x_1	$P_{\theta_1}(x_1) = 0.8$	$P_{\theta_2}(x_1) = 0.1$
x_2	$P_{\theta_1}(x_2) = 0.2$	$P_{\theta_2}(x_2) = 0.9$

(b) Auswertungsproblem:

- Menge aller möglichen Entscheidungsfunktionen: $\mathcal{D} = \{d_j : \mathcal{X} \to \mathbb{A} \mid j \in \{1, \ldots, 4\}\}$ mit

	x_1	x_2
d_1	$d_1(x_1) = a_1$	$d_1(x_2) = a_1$
d_2	$d_2(x_1) = a_1$	$d_2(x_2) = a_2$
d_3	$d_3(x_1) = a_2$	$d_3(x_2) = a_1$
d_4	$d_4(x_1) = a_2$	$d_4(x_2) = a_2$

- Bewertung der Entscheidungsfunktionen mit dem Risiko

$$R : \mathcal{D} \times \Theta \to \mathbb{R} \qquad \text{mit}$$

$$(d, \theta) \mapsto R(d, \theta) = \int_{\mathcal{X}} \ell(d(x), \theta) dP_\theta(x) = \ell(d(x_1), \theta)\, P_\theta(x_1) + \ell(d(x_2), \theta)\, P_\theta(x_2)$$

	θ_1	θ_2
d_1	$R(d_1, \theta_1) = 0 \cdot 0.8 + 0 \cdot 0.2 = 0$	$R(d_1, \theta_2) = 30 \cdot 0.1 + 30 \cdot 0.9 = 30$
d_2	$R(d_2, \theta_1) = 0 \cdot 0.8 + 10 \cdot 0.2 = 2$	$R(d_2, \theta_2) = 30 \cdot 0.1 + 0 \cdot 0.9 = 3$
d_3	$R(d_3, \theta_1) = 10 \cdot 0.8 + 0 \cdot 0.2 = 8$	$R(d_3, \theta_2) = 0 \cdot 0.1 + 30 \cdot 0.9 = 27$
d_4	$R(d_4, \theta_1) = 10 \cdot 0.8 + 10 \cdot 0.2 = 10$	$R(d_4, \theta_2) = 0 \cdot 0.1 + 0 \cdot 0.9 = 0$

- Das Auswertungsproblem $(\mathcal{D}, \Theta, R)$ stellt wieder ein datenfreies Entscheidungsproblem dar.