Multivariate Statistik, Übung 7

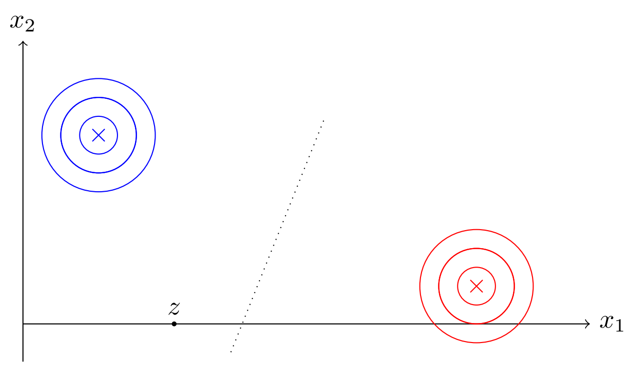
Henry Haustein

Aufgabe 1

(a) Um a zu bestimmen, müssen wir zuerst $\hat{\Sigma}$ invertieren:

$$\hat{\Sigma}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$
$$a = \hat{\Sigma}^{-1}(\bar{x}^{C_1} - \bar{x}^{C_2}) = (-10, 4)'$$
$$t = a' \frac{\bar{x}^{C_1} + \bar{x}^{C_2}}{2} = -58$$

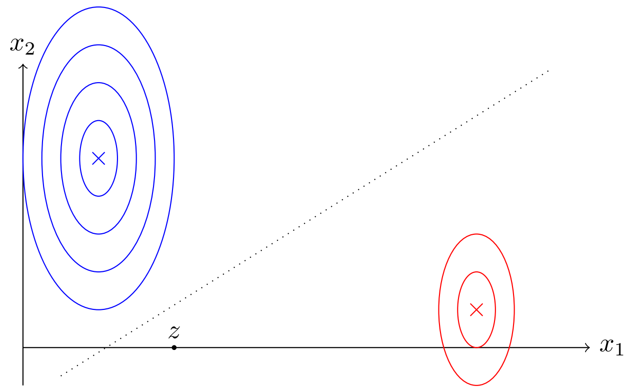
Um ein Objekt in Klasse 1 einzuordnen, muss gelten $-10x_1 + 4x_2 > -58$. Die diskriminierende Kurve lautet also: $x_2 = -\frac{58}{4} + \frac{10}{4}x_1$. Tatsächlich wird der Punkt z Klasse 1 zugeordnet.



(b) Analog zu (a) ergibt sich

$$\hat{\Sigma}^{-1} = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 1 \end{pmatrix}$$
$$a = \hat{\Sigma}^{-1}(\bar{x}^{C_1} - \bar{x}^{C_2}) = (-2.5, 4)'$$
$$t = a' \frac{\bar{x}^{C_1} + \bar{x}^{C_2}}{2} = -5.5$$

Um ein Objekt in Klasse 1 einzuordnen, muss gelten $-2.5x_1 + 4x_2 > -5.5$. Die diskriminierende Kurve lautet also: $x_2 = -\frac{11}{8} + \frac{5}{8}x_1$. Tatsächlich wird der Punkt z Klasse 2 zugeordnet.



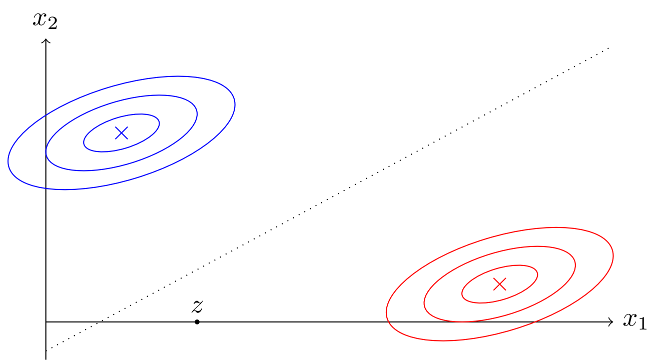
(c) Analog zu (a) ergibt sich

$$\hat{\Sigma}^{-1} = \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{4}{3} \end{pmatrix}$$

$$a = \hat{\Sigma}^{-1}(\bar{x}^{C_1} - \bar{x}^{C_2}) = \left(-\frac{14}{3}, \frac{26}{3}\right)'$$

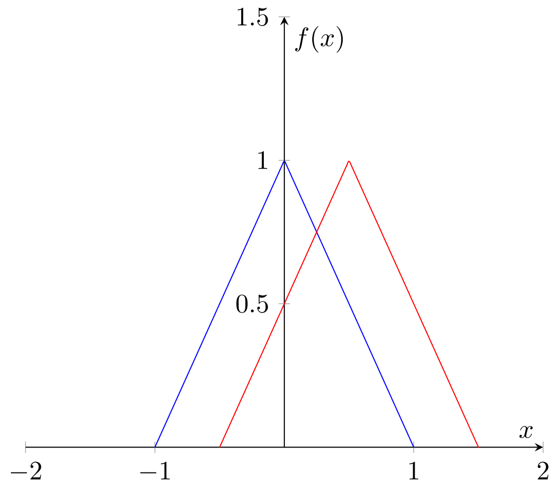
$$t = a'\frac{\bar{x}^{C_1} + \bar{x}^{C_2}}{2} = -\frac{20}{3}$$

Um ein Objekt in Klasse 1 einzuordnen, muss gelten $-\frac{14}{3}x_1 + \frac{26}{3}x_2 > -\frac{20}{3}$. Die diskriminierende Kurve lautet also: $x_2 = -\frac{10}{13} + \frac{7}{13}x_1$. Tatsächlich wird der Punkt z Klasse 2 zugeordnet.



Aufgabe 2

(a) Diagramm



(b) Für den Fall gleicher Fehlklassifikationskosten und gleicher a-priori-Wahrscheinlichkeiten gilt

$$R_1 = \{x \mid f_1(x) \geq f_2(x)\}$$
$$R_2 = \{x \mid f_1(x) < f_2(x)\}$$

Wir müssen also den Schnittpunkt der beiden Dichtefunktionen bestimmen.

$$1 - |x| = 1 - |x - 0.5|$$
$$|x| = |x - 0.5|$$

Fallunterscheidung: Entweder ist $x \geq 0$ oder nicht:

- $x \geq 0$:

$$x = |x - 0.5|$$

  Fallunterscheidung: Entweder ist $x \geq 0.5$ oder nicht:

  - $x \geq 0.5$:

$$x = x - 0.5$$
$$0 = -0.5 \Rightarrow \text{Widerspruch}$$

  - $x < 0.5$:

$$x = -(x - 0.5)$$
$$x = -x + 0.5$$
$$2x = 0.5$$
$$x = 0.25$$

- $x < 0$:

$$-x = |x - 0.5|$$

  Fallunterscheidung: Entweder ist $x \geq 0.5$ oder nicht:

  - $x \geq 0.5$: Widerspruch, da $x < 0$ schon angenommen ist

- $x < 0.5$:

$$\begin{aligned} -x &= -(x - 0.5) \\ x &= x - 0.5 \\ 0 &= -0.5 \Rightarrow \text{Widerspruch} \end{aligned}$$

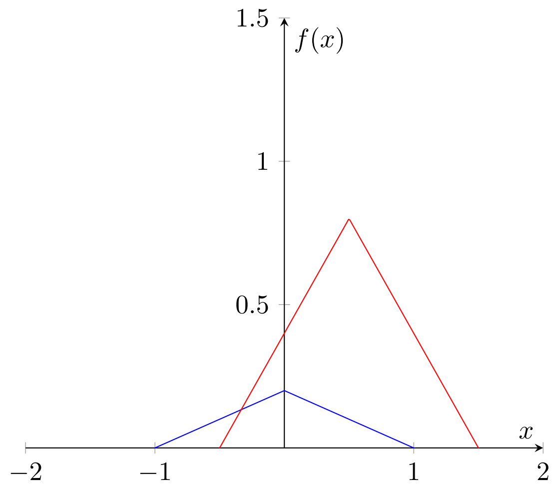
Der Schnittpunkt der beiden Dichten ist also bei $x = \frac{1}{4}$. Damit ist

$$\begin{aligned} R_1 &= \left[-1, \frac{1}{4}\right] \\ R_2 &= \left(\frac{1}{4}, \frac{3}{2}\right] \end{aligned}$$

(c) Jetzt lauten die Klassifikationsregionen

$$\begin{aligned} R_1 &= \{x \mid f_1(x) \cdot 0.2 \geq f_2(x) \cdot 0.8\} \\ R_2 &= \{x \mid f_1(x) \cdot 0.2 < f_2(x) \cdot 0.8\} \end{aligned}$$

Wir interessieren uns wieder für den Schnittpunkt von $0.2 \cdot f_1(x)$ und $0.8 \cdot f_2(x)$. Die Rechnung ist ziemlich analog wie oben, der Schnittpunkt ist bei $x = -\frac{1}{3}$.



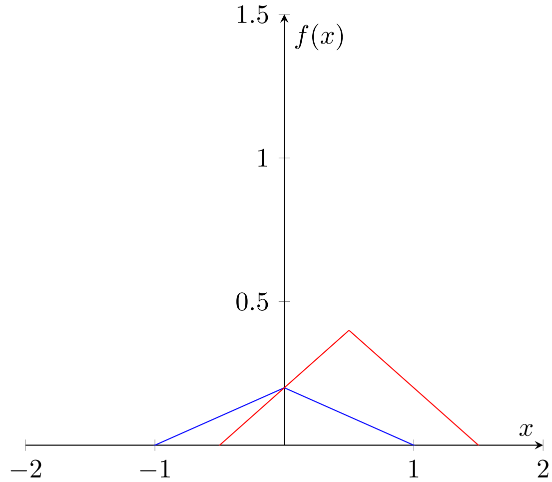
Die Regionen sind also

$$\begin{aligned} R_1 &= \left[-1, -\frac{1}{3}\right] \\ R_2 &= \left(-\frac{1}{3}, \frac{3}{2}\right] \end{aligned}$$

(d) Jetzt lauten die Klassifikationsregionen

$$\begin{aligned} R_1 &= \left\{x \mid f_1(x) \cdot 0.2 \geq f_2(x) \cdot 0.8 \cdot \frac{0.5}{1}\right\} \\ R_2 &= \left\{x \mid f_1(x) \cdot 0.2 < f_2(x) \cdot 0.8 \cdot \frac{0.5}{1}\right\} \end{aligned}$$

Wir interessieren uns wieder für den Schnittpunkt von $0.2 \cdot f_1(x)$ und $0.4 \cdot f_2(x)$. Die Rechnung ist ziemlich analog wie oben, der Schnittpunkt ist bei $x = 0$.



Die Regionen sind also

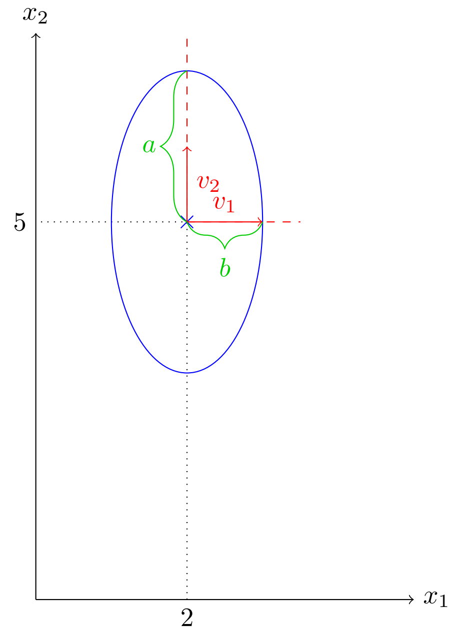
$$R_1 = [-1, 0]$$
$$R_2 = \left(0, \frac{3}{2}\right]$$

Berechnung der Höhenlinien

Weil es nicht ganz einfach ist, die Höhenlinien auszurechnen, hier noch mal eine detaillierte Erklärung anhand $\hat{\Sigma}$ von (b): Die Eigenwerte von $\hat{\Sigma}^{-1}$ sind $\lambda_1 = 1$ und $\lambda_2 = \frac{1}{4}$ und die Eigenvektoren $v_1 = (1, 0)'$ und $v_2 = (0, 1)'$. Wir wählen als Abstand $d = 1$ und damit sind die Achsenlängen der Ellipsen der Höhenlinien

$$a = \sqrt{\frac{1^2}{1}} = 1$$
$$b = \sqrt{\frac{1^2}{0.25}} = 2$$

Zeichnen wir das ganze ein.


x_2
a
v_2
v_1
5
b
2
x_1