Formelsammlung

July 9, 2018

Contents

1 Bayes'scher Ansatz

1.1 Entscheidungsregionen

Bei gleicher Klassifizierungswahrscheinlichkeit gilt

$$p(x|C_1) = p(x|C_2)$$

$$p(x|C_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot e^{\frac{-\frac{1}{2}(x-\mu_i)^2}{\sigma_i^2}}$$

1.1.1 Ungleiche Klassifizierungswahrscheinlichkeit

Wenn eine Klasse wahrscheinlicher ist so gilt

$$p(C_1) \cdot p(x|C_1) = p(x|C_2) \cdot p(C_2)$$

$$\frac{p(C_1)}{p(C_2)} \cdot \frac{\sqrt{2\pi\sigma_2^2}}{\sqrt{2\pi\sigma_1^2}} = e^{\frac{-\frac{1}{2}(x-\mu_2)^2}{\sigma_2^2} - \frac{-\frac{1}{2}(x-\mu_1)^2}{\sigma_1^2}}$$

$\frac{p(C_1)}{p(C_2)}$ definieren wir als $\mathcal{P}$. Als Reziproge von $\mathcal{P}$ bezeichnen wir$\mathcal{P}'$. Bei gleichen Wahrscheinlichkeiten $p(C_{1/2}) = 0,5$ ist $\mathcal{P} = 1$ und kann ignoriert werden.

1.1.2 Fehlklassifizierung

Wenn eine Verlustmatrix V gegeben ist so wird die Gewichtung der Fehlklassifikation pro Klasse berücksichtigt über

$$V_{1,2} \cdot p(x|C_1) = p(x|C_2) \cdot V_{2,1}$$

$$\frac{V_{1,2}}{V_{2,1}} \cdot \frac{\sqrt{2\pi\sigma_2^2}}{\sqrt{2\pi\sigma_1^2}} = e^{\frac{-\frac{1}{2}(x-\mu_2)^2}{\sigma_2^2} - \frac{-\frac{1}{2}(x-\mu_1)^2}{\sigma_1^2}}$$

$\frac{V_{1,2}}{V_{2,1}}$ definieren wir als $\mathcal{V}$. Bei gleicher Verlustgewichtung $V_{(1/2),(2/1)} = x$ ist $\mathcal{V} = 1$ und kann ignoriert werden.

1.1.3 Umformung

Mit diesen Parametern kann die Formel wie folgt umgeformt werden

$$ln(\mathcal{P} \cdot \mathcal{V} \cdot \frac{\sqrt{\sigma_2^2}}{\sqrt{\sigma_1^2}}) = \frac{-\frac{1}{2}(x-\mu_2)^2}{\sigma_2^2} - \frac{-\frac{1}{2}(x-\mu_1)^2}{\sigma_1^2} \tag{1}$$

$$\sigma_2^2 \cdot \sigma_1^2 \cdot ln(\mathcal{P} \cdot \mathcal{V} \cdot \frac{\sqrt{\sigma_2^2}}{\sqrt{\sigma_1^2}}) = -\frac{1}{2} \cdot \sigma_1^2 (x-\mu_2)^2 - -\frac{1}{2} \cdot \sigma_2^2 (x-\mu_1)^2 \tag{2}$$

$$= \sigma_1^2 (x^2 - 2\mu_2 x + \mu_2^2) - \sigma_2^2 (x^2 - 2\mu_1 x + \mu_1^2) \tag{3}$$

$$= (\sigma_1^2 - \sigma_2^2) \cdot x^2 + (-\mu_2 \sigma_1^2 + \mu_1 \sigma_2^2) 2x + \sigma_1^2 \mu_2^2 - \sigma_2^2 \mu_1^2 \tag{4}$$

$$0 = (\sigma_1^2 - \sigma_2^2) \cdot x^2 + (-\mu_2 \sigma_1^2 + \mu_1 \sigma_2^2) 2x + \sigma_1^2 \mu_2^2 - \sigma_2^2 \mu_1^2 + 2 \cdot \sigma_2^2 \cdot \sigma_1^2 \cdot ln(\mathcal{P} \cdot \mathcal{V} \cdot \frac{\sqrt{\sigma_2^2}}{\sqrt{\sigma_1^2}}) \tag{5}$$

1.1.4 pq-Formel

Benötigt einen Ausdruck der Form $x^2 + px + q = 0$. Um die pq-Formel $x_{1/2} = -\frac{p}{2} \pm \sqrt{(\frac{p}{2})^2 - q}$ anzuwenden muss der Term mit $/ \div (\sigma_1^2 - \sigma_2^2)$ umgeformt werden

1.2 Zurückweisung

$$p(C_1|x) \geq \theta$$

$$p(C_1|x) = \frac{p(C_1) \cdot p(x|C_1)}{p(x)}$$

$$p(x) = p(C_1) \cdot p(x|C_1) + p(C_2) \cdot (x|C_2)$$

$$\frac{p(C_1) \cdot p(x|C_1)}{p(C_1) \cdot p(x|C_1) + p(C_2) \cdot (x|C_2)} \geq \theta$$

Dieser Ausdruck hat die Form $\frac{A}{A+B}$ was umgeformt werden kann in $\frac{1}{1+\frac{B}{A}}$

$$\frac{1}{1 + \frac{p(C_2) \cdot (x|C_2)}{p(C_1) \cdot p(x|C_1)}} \geq \theta$$

$$1 \geq \theta (1 + \frac{p(C_2) \cdot (x|C_2)}{p(C_1) \cdot p(x|C_1)}) \tag{6}$$

$$\frac{1}{\theta} \geq 1 + \mathcal{P}' \cdot \frac{p(x|C_2)}{p(x|C_1)} \tag{7}$$

$$\frac{1}{\theta} - 1 \geq \mathcal{P}' \cdot \frac{\frac{1}{\sqrt{2\pi\sigma_2^2}} \cdot e^{\frac{-\frac{1}{2}(x-\mu_2)^2}{\sigma_2^2}}}{\frac{1}{\sqrt{2\pi\sigma_1^2}} \cdot e^{\frac{-\frac{1}{2}(x-\mu_1)^2}{\sigma_1^2}}} \tag{8}$$

$$\frac{1}{\theta} - 1 \geq \mathcal{P}' \cdot \frac{\sqrt{2\pi\sigma_1^2}}{\sqrt{2\pi\sigma_2^2}} \cdot e^{\frac{-\frac{1}{2}(x-\mu_2)^2}{\sigma_2^2} - \frac{-\frac{1}{2}(x-\mu_1)^2}{\sigma_1^2}} \tag{9}$$

Nun kann analog zu 1 der Term weiter umgeformt werden zu

$$0 = (\sigma_1^2 - \sigma_2^2) \cdot x^2 + (-\mu_2 \sigma_1^2 + \mu_1 \sigma_2^2) 2x + \sigma_1^2 \mu_2^2 - \sigma_2^2 \mu_1^2 + 2 \cdot \sigma_1^2 \cdot \sigma_2^2 \cdot ln((\frac{1}{\theta} - 1) \cdot \mathcal{P}' \cdot \frac{\sqrt{\sigma_2^2}}{\sqrt{\sigma_1^2}})$$

Nun muss für $p(C_2|x)$ dies Analog geschehen

$$\frac{p(C_2) \cdot p(x|C_2)}{p(C_1) \cdot p(x|C_1) + p(C_2) \cdot (x|C_2)} \geq \theta$$

$$\frac{1}{1 + \frac{p(C_1) \cdot p(x|C_1)}{p(C_2) \cdot p(x|C_2)}} \geq \theta$$

$$\frac{1}{\theta} \geq 1 + \mathcal{P} \cdot \frac{p(x|C_1)}{p(x|C_2)}$$

$$\frac{1}{\theta} - 1 \geq \mathcal{P} \cdot \frac{\sqrt{2\pi\sigma_2^2}}{\sqrt{2\pi\sigma_1^2}} \cdot e^{\frac{-\frac{1}{2}(x-\mu_1)^2}{\sigma_1^2} - \frac{-\frac{1}{2}(x-\mu_2)^2}{\sigma_2^2}}$$

$$0 = (\sigma_2^2 - \sigma_1^2) \cdot x^2 + (-\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2) 2x + \sigma_2^2 \mu_1^2 - \sigma_1^2 \mu_2^2 + 2 \cdot \sigma_1^2 \cdot \sigma_2^2 \cdot ln((\frac{1}{\theta} - 1) \cdot \mathcal{P} \cdot \frac{\sqrt{\sigma_1^2}}{\sqrt{\sigma_2^2}})$$

2 Lineare Klassifikatoren

Klassifizieren eines Vektors V wird durch Wenn $V = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ so kann die Gleichung gegen null gesetzt werden und dieses Ergebniss ist die Diskriminanzfunktion.

$$-w^{*T} \cdot V - w_0 \tag{10}$$

Das Signum des Ergebnisses gibt die Klassifizierung an.

2.1 Fisher Diskriminante

Brerechnen des Klassenmittelpunktes μ der Klasse i

$$\mu_i = \frac{1}{|i|}(\sum_{k \in C_1} x_k)$$

Berechnen der Kovarianz zwischen den Klassen

$$S_B = (\mu_2 - \mu_1)(\mu_2 - \mu_1)^T$$

Berechnen der Gesamtkovarianz innerhalb der Klassen

$$S_W = \sum_{k \in C_1} (x_k - \mu_1)(x_k - \mu_1)^T + \sum_{k \in C_2} (x_k - \mu_2)(x_k - \mu_2)^T$$

$$w^* = S_w^{-1}(\mu_2 - \mu_1)$$

$$w_0^* = w^{*T}\frac{1}{N}(N_1\mu_1 + N_2\mu_2)$$

wobei N die Anzahl aller Klassifizierten Vektoren und N_i die Anzahl der Vektoren die in der Klasse i ist

Über 10 kann nun die Diskriminanzfunktion bestimmt oder Punkte Klassifiziert werden.

2.2 Lineare Kleinste Quadrate

Erweitern von w in höhere Dimension

Berechnung der Diskriminanzfunktion w^*

$$w^* = (XX^T)^{-1}Xt^T$$

Über 10 kann nun die Diskriminanzfunktion bestimmt oder Punkte Klassifiziert werden.

3 Perceptron

Überführung des Inputs in höher Dimensionales Problem $x \in R^{D+1}$

$$x = \begin{pmatrix} 1 & 1 & 1 & 1 & ... \\ x_1 & x_2 & x_3 & x_4 & ... \\ y_1 & y_2 & y_3 & y_4 & ... \\ \vdots \end{pmatrix}$$

3.1 Bestimmung der Diskriminanzfunktion

zum Zeitpunkt t

$$\begin{aligned} g &= w^T x + w_0 \cdot 1 \\ &= \begin{pmatrix} w_1 & w_2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + w_0 \cdot 1 \\ &= \begin{pmatrix} w_0 & w_1 & w_2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} \\ 0 &= w(t)^T * \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} \end{aligned} \tag{11}$$

Durch einsetzen kann auch die Klassifikation eines Punktes bestimmt werden.

3.1.1 Bestimmung der nächsten Diskriminanzfunktion

$$w(t+1) = w(t) - \rho(t)\frac{\partial}{\partial w}J(w)|_{w=w(t)}$$

$$\frac{\partial}{\partial w}J(w)|_{w=w(t)} = \sum_{k\in Y} -t_k \cdot x_k$$

wobei Y die Menge aller falsch klassifizierten Punkte beschreibt.

4 Abstandsmaße

d_1 = manhatten Abstand. Definiert durch $d_1 = |(x_1 - x_2)| + |(y_1 - y_2)|$
d_2 = euklidischer Abstand. Definiert durch $d_2 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

5 Lösungen für Probeklausur

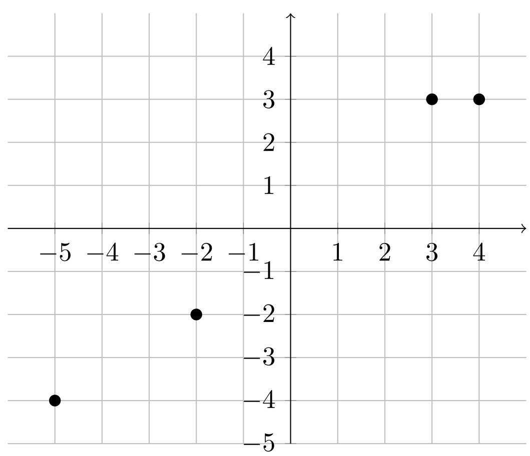
5.1 ProbeKlausur Muster SS17

5.1.1 Aufgabe 1

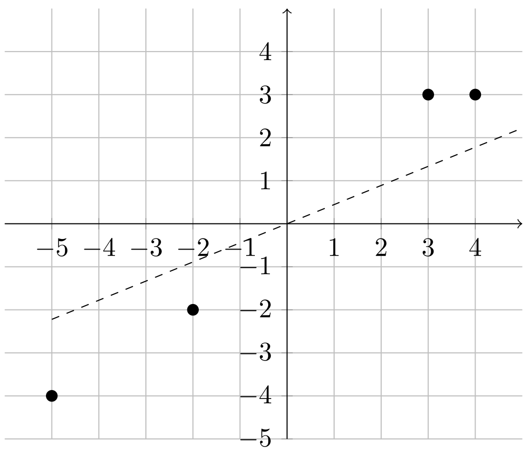
- a:
- b: $x_1 = 1,07449, x_2 = 8,06836$
- c: $x_1 = 0,590907, x_2 = 8,55195$
- d: $x_1 = 1,00744, x_2 = 8,13541$
- e:
  $p(C_1|x) : x_1 = 0,8805, x_2 = 8.26236$
  $p(C_2|x) : x_1 = 1,2799, x_2 = 7,8629$

5.1.2 Aufgabe 2

- a:



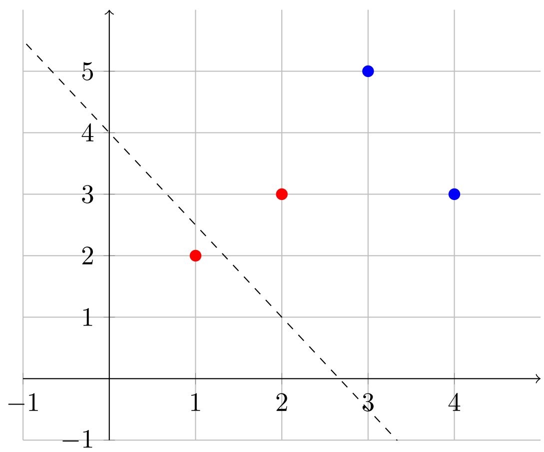
- b:
  $\mu_1 = \begin{pmatrix} -3.5 \\ -3 \end{pmatrix}$ $\mu_2 = \begin{pmatrix} 3.5 \\ 3 \end{pmatrix}$
  $S_w = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix}$
  $w^* = \begin{pmatrix} 4 \\ -9 \end{pmatrix}$ $w_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$



- c:
  $w^* = \begin{pmatrix} 0 \\ \frac{8}{27} \\ -\frac{2}{3} \end{pmatrix}$ $w_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
- d:
  $w^* \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} - w_0$
  Fischer: $= -1$
  KLQ: $= -\frac{2}{277}$

5.1.3 Aufgabe 3

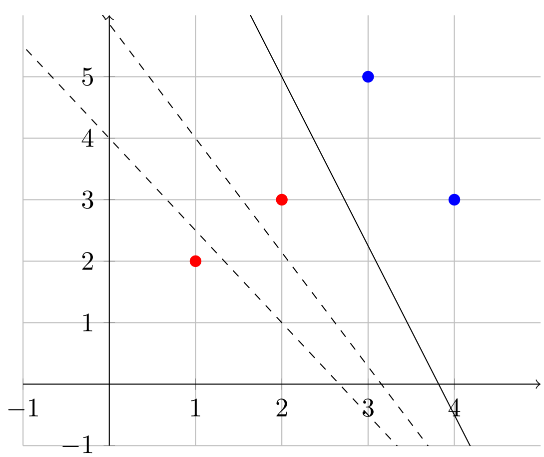
- a:



- b:

  $w(1) = \begin{pmatrix} 4.1 \\ -1.3 \\ -0.7 \end{pmatrix}$

  $w(2) = \begin{pmatrix} 4.2 \\ -1.1 \\ -0.4 \end{pmatrix}$



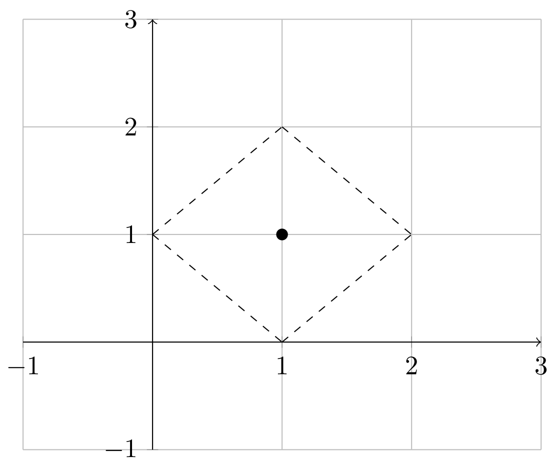
- c: Der Algorithmus endet sobald alle Punkte richtig Klassifiziert werden. Die Schrittweite bestimmt dabei wie schnell/langsam sich die Gerade im Raum bewegt. Bei einer zu großen Schrittweite kann unter Umständen keine richtige Klassifizierung gefunden werden. Wird die Schrittweite zu klein gewählt so dauert die Berechnung unnötig lang.

- d: $w(2)^T \cdot \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix} = -\frac{7}{10}$
  $sig(-\frac{7}{10})$ ist negativ deswegen würde der Punkt auf $-$ Klassifiziert werden.

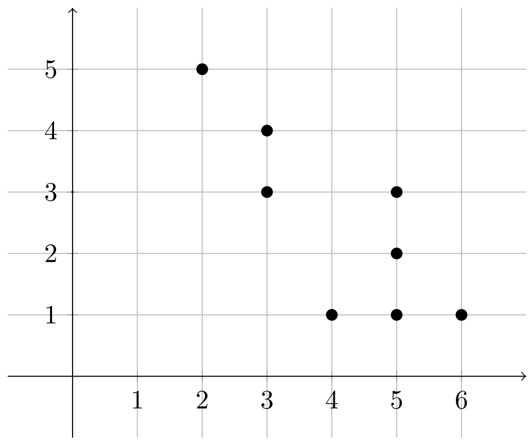
5.1.4 Aufgabe 4

5.1.5 Aufgabe 5

- a
  - $a1 = \sqrt{5}$
  - $a2 = 3$
  - $a3 = \sqrt{5}$
- b



- c d_{max}



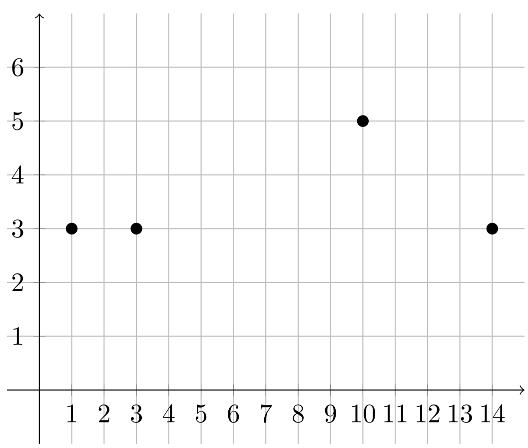
- d: 1 und zweiter Punkt tauschen

5.2 Prüfungs SS17

5.2.1 Aufgabe 1

5.2.2 Aufgabe 2

- a

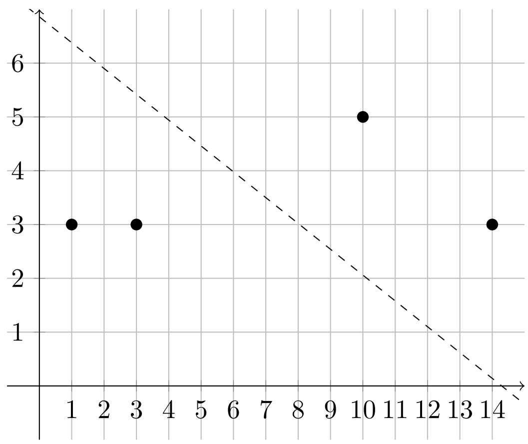


- b

  $\mu_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ $\mu_2 = \begin{pmatrix} 12 \\ 4 \end{pmatrix}$

  $S_w = \begin{pmatrix} 10 & -4 \\ -4 & 1 \end{pmatrix}$

  $w^* = \begin{pmatrix} 6 \\ 12\frac{1}{2} \end{pmatrix}$ $w_0 = 85.75$



- c

Aufgabe c: