Universität Salzburg	Florian Graf
Machine Learning	
Übungsblatt **3**	30 Punkte

Aufgabe 1. *Mehrdimensionale Normalverteilung – Teil II* 14 P.

Die d-dimensionale standard Normalverteilung $\mathcal{N}(\mathbf{0},\mathbf{1})$ ist die Wahrscheinlichkeitsverteilung für Zufallsvektoren $\mathbf{X} = (X_1, \dots X_d)$, wobei alle Komponenten voneinander unabhängig sind und einer eindimensionalen standard Normalverteilung folgen, d.h., $X_i \sim \mathcal{N}(0,1)$.

(a) Bestimmen Sie die Dichtefunktion $f(x_1, \dots, x_d)$ von $\mathbf{X}$.

(b) Bestimmen Sie die Niveaumenge der Dichtefunktion zum Niveau $(2\pi)^{-d/2}e^{-\frac{1}{2}}$, d.h., die Menge $\{(x_1, \dots, x_d) \in \mathbb{R}^d : f(x_1, \dots, x_d) = (2\pi)^{-d/2}\}$. Welchem geometrischen Objekt entspricht sie?

Es sei nun $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit $\boldsymbol{\mu} \in \mathbb{R}^d$. In diesem Fall ist die Dichte von $\mathbf{Y}$ gegeben durch

$$g(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{y}-\boldsymbol{\mu})\right) \quad .$$

(c) Zeigen Sie, dass die Definition von g nur Sinn macht, falls $\boldsymbol{\Sigma}$ positiv definit ist, d.h., falls $\mathbf{x}^\top\boldsymbol{\Sigma}\mathbf{x} > 0$ für alle $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^d$. Zeigen Sie dazu, dass

  (i) $\boldsymbol{\Sigma}$ ist invertierbar, dies gilt z.B., falls $\ker(\boldsymbol{\Sigma}) = \{\mathbf{0}\}$ (d.h. falls $\boldsymbol{\Sigma}\mathbf{x} = \mathbf{0}$, dann ist $\mathbf{x} = \mathbf{0}$).

  (ii) $(\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{y}-\boldsymbol{\mu}) > 0$ für alle $\mathbf{y} \neq \boldsymbol{\mu}$., bzw., dass $\boldsymbol{\Sigma}^{-1}$ positiv definit ist.

  (iii) $\det(\boldsymbol{\Sigma}) > 0$, dies gilt z.B., falls alle Eigenwerte von $\boldsymbol{\Sigma}$ positiv sind.

(d) Es sei nun $\mathbf{Y} = \mathbf{A}\mathbf{X}$. Zeigen Sie, dass $\boldsymbol{\Sigma}$ positiv definit ist, falls $\ker(\mathbf{A}^\top) = \{\mathbf{0}\}$.

Es sei nun $d = 2$ und $\mathbf{A} = \begin{pmatrix}\sigma_1\mathbf{v}_1 & \sigma_2\mathbf{v}_2\end{pmatrix} \in \mathbb{R}^2$, wobei $\sigma_1, \sigma_2 > 0$ und die Spaltenvektoren $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ orthonormal sind (d.h., es gilt $\mathbf{v}_1^\top\mathbf{v}_1 = \mathbf{v}_2^\top\mathbf{v}_2 = 1$ und $\mathbf{v}_1^\top\mathbf{v}_2 = 0$).

(e) Bestimmen Sie die inverse Matrix $\mathbf{A}^{-1}$.

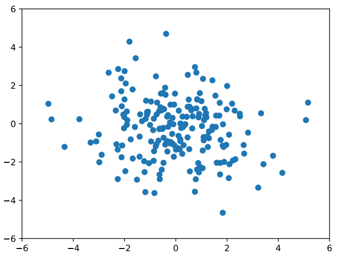
(f) Bestimmen Sie die Niveaumenge $\{\mathbf{y} \in \mathbb{R}^2 : g(\mathbf{y}) = \frac{e^{-\frac{1}{2}}}{2\pi\sigma_1\sigma_2}\}$.
Hinweis: Machen Sie den Ansatz $\mathbf{y} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2$.

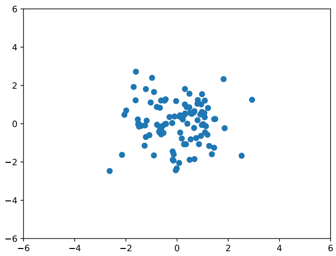
(g) Zeichnen Sie die Niveaumenge L und die Vektoren $\sigma_1\mathbf{v}_1$, $\sigma_2\mathbf{v}_2$ in ein Koordinatensystem ein für den Fall $\mathbf{v}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix}2\\1\end{pmatrix}$, $\mathbf{v}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix}-1\\2\end{pmatrix}$, $\sigma_1^2 = 5$, $\sigma_2^2 = 1$.

(h) Zeichnen Sie 2 weitere Niveaumengen ein. Nutzen Sie die Niveaumengen um eine Stichprobe zu skizzieren, die mit hoher Wahrscheinlichkeit 20 unabhängigen Beobachtungen von $\mathbf{Y}$ entspricht. (Hierfür ist keine Rechnung notwendig)

Aufgabe 2. *Zweidimensionale Normalverteilung – Beispiele* 3 P.

Die folgenden Abbildungen zeigen jeweils eine Stichprobe einer zweidimensionaler Normalverteilung.





Ordnen Sie jede Stichprobe einer der folgenden Wahrscheinlichkeitsverteilungen zu. Begründen Sie Ihre Wahl.

$\mathcal{N}(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix})$, $\mathcal{N}(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}2 & -1\\-1 & 5\end{pmatrix})$, $\mathcal{N}(\begin{pmatrix}0\\2\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix})$,

$\mathcal{N}(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}3 & 0\\0 & 3\end{pmatrix})$, $\mathcal{N}(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}6 & -3\\-3 & 15\end{pmatrix})$, $\mathcal{N}(\begin{pmatrix}0\\2\end{pmatrix}, \begin{pmatrix}3 & 0\\0 & 3\end{pmatrix})$.

Aufgabe 3. *MLE – Bernoulli Verteilung* 5 P.

Gegeben seien Beobachtungen $x_1, \dots, x_n, \in \{0, 1\}$. Wir nehmen an, dass diese von einer Bernoulliverteilten Zufallsvariable $X \sim \text{Ber}(\lambda)$ stammen, d.h. $\mathbb{P}[X = 1] = \lambda$ und $\mathbb{P}[X = 0] = 1 - \lambda$. Schätzen Sie den Parameter λ mit der Maximum-Likelihood Methode. Überprüfen Sie, dass tatsächlich ein globales Maximum vorliegt.

Aufgabe 4. *MLE – Normalverteilung* 8 P.

Gegeben seien Beobachtungen $x_1, \dots, x_n, \in \mathbb{R}$.

(a) Wir nehmen an, dass die Beobachtungen von einer Normalverteilten Zufallsvariable $X \sim \mathcal{N}(\mu, \sigma^2)$ stammen. Schätzen Sie die Parameter μ und σ^2 gemeinsam mittels der Maximum-Likelihood Methode. Überprüfen Sie, dass tatsächlich ein globales Maximum vorliegt.

Hinweis: In diesem Fall ist die log-Likelihoodfunktion eine Funktion $\log L : \mathbb{R}^2 \to \mathbb{R}^+$. Eine notwendige Bedingung an die lokalen Maxima $(\hat{\mu}, \hat{\sigma}^2)$ einer solchen Funktion ist, dass der Gradient

$$\nabla \log L(\mu, \sigma^2) := \begin{pmatrix} \frac{\partial \log L(\mu,\sigma^2)}{\partial \mu} \\ \frac{\partial \log L(\mu,\sigma^2)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \quad \text{für } \mu = \hat{\mu} \text{ und } \sigma^2 = \hat{\sigma}^2 \ ,$$

und dass die Hessematrix

$$H_{\log L}(\mu, \sigma^2) := \begin{pmatrix} \frac{\partial^2 \log L(\mu,\sigma^2)}{\partial \mu \partial \mu} & \frac{\partial^2 \log L(\mu,\sigma^2)}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 \log L(\mu,\sigma^2)}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 \log L(\mu,\sigma^2)}{\partial \sigma^2 \partial \sigma^2} \end{pmatrix} \quad \text{an der Stelle } (\mu, \sigma^2) = (\hat{\mu}, \hat{\sigma}^2) \text{ positiv definit ist.}$$