Universität Salzburg | Florian Graf

Machine Learning

Übungsblatt **1** | 15 Punkte

Aufgabe 1. 4 P.

Die multivariate Normalverteilung $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ in $\mathbb{R}^d$ mit Mittelwert $\boldsymbol{\mu} \in \mathbb{R}^d$ und Kovarianzmatrix $\Sigma \in \mathbb{R}^{d\times d}$ ist definiert über die Dichte

$$\rho : \mathbb{R}^2 \to [0, \infty) \ , \qquad \mathbf{x} \mapsto \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

Das heißt, für normalverteilte Zufallsvariablen $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ gilt $\mathbb{P}[\mathbf{X} \in A] = \int_A \rho(\mathbf{x}) d\mathbf{x}$.

Im folgenden sei $d = 2$, d.h. wir betrachten die zweidimensionale Normalverteilung mit $\boldsymbol{\mu} = \begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix}$ und $\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}$.

(a) Wir betrachten die Standardnormalverteilung $\mathcal{N}(\mathbf{0}, \mathbf{I})$, d.h. $\boldsymbol{\mu} = \mathbf{0} = \begin{pmatrix}0 \\ 0\end{pmatrix}$ und $\Sigma = \mathbf{I} = \begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}$.

Zeigen Sie dass $\mathbb{P}[\mathbf{X} \in \mathbb{R}^2] = 1$. (Tipp: Berechnen Sie das Integral in Polarkoordinaten.)

(b) Berechnen Sie den Erwartungswert $\mathbb{E}[\mathbf{X}] = \begin{pmatrix}\mathbb{E}[X_1] \\ \mathbb{E}[X_2]\end{pmatrix}$ einer standard-normalverteilten Zufallsvariable $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) = \begin{pmatrix}X_1 \\ X_2\end{pmatrix}$, wobei $\mathbb{E}[X_i]$ definiert ist als $\mathbb{E}[X_i] = \int_{\mathbb{R}} \int_{\mathbb{R}} x_i \rho(x_1, x_2)\, dx_1\, dx_2$.

(c) Berechnen Sie die Kovarianzen $\mathrm{Cov}(X)_{ij} = \mathbb{E}\left[(X_i - \mathbb{E}[x_i])(X_j - \mathbb{E}[x_j])\right]$.

(d) Bonus: Nun sei $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ für beliebiges $\boldsymbol{\mu} \in \mathbb{R}^2$ und positiv definites $\Sigma \in \mathbb{R}^{2\times2}$. Zeigen Sie $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$ und $\mathrm{Cov}(\mathbf{X}) = \Sigma$.

(e) Zeichnen Sie die Niveaulinien der Wahrscheinlichkeitsdichte ρ, d.h. die Mengen $L_c := \{\mathbf{x} \in \mathbb{R}^2 : \rho(\mathbf{x}) = c\}$ für verschiedene Werte $c > 0$ für

- $\mathcal{N}(\mathbf{0}, \mathbf{I})$ und
- $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ mit $\boldsymbol{\mu} = \begin{pmatrix}1 \\ 1\end{pmatrix}$ und $\Sigma = \begin{pmatrix}1 & 0 \\ 0 & 2\end{pmatrix}$.

Nutzen Sie diese um eine Stichprobe der Verteilungen zu skizzieren.

Aufgabe 2. 6 P.

Gegeben sei das folgende Wahrscheinlichkeitsmodell

$$p(y = c|\mathbf{x}, \boldsymbol{\theta}) = \frac{p(\mathbf{x}|y = c, \boldsymbol{\theta}) p(y = c|\boldsymbol{\theta})}{\sum_{c'} p(\mathbf{x}|y = c', \boldsymbol{\theta}) p(y = c'|\boldsymbol{\theta})} \ , \tag{1}$$

wobei $c \in \{c_1, c_2\}$, $p(\mathbf{x}|y = c, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_c, \Sigma_c)$ und $p(y = c_1|\boldsymbol{\theta}) = \lambda$.

(a) Wir weisen einer Beobachtung $\mathbf{x} \in \mathbb{R}^2$ die Klasse c zu, falls $p(y = c|\mathbf{x}, \boldsymbol{\theta}) = \max_{c' \in c_1, c_2} p(y = c'|\mathbf{x}, \boldsymbol{\theta})$. Zeigen Sie, dass die Entscheidungsgrenze (decision boundary) $G := \{\mathbf{x} \in \mathbb{R} : p(y = c_1|\mathbf{x}, \boldsymbol{\theta}) = p(y = c_2|\mathbf{x}, \boldsymbol{\theta})\}$ durch eine quadratische Gleichung der Form

$$\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c = 0. \tag{2}$$

bestimmt ist und leiten Sie Formeln für $\mathbf{A}$, $\mathbf{b}$ und c her.

Die Lösungsmengen von Gleichung (2) sind Kegelschnitte (Ellipsen, Parabeln, Hyperbeln) oder Geraden.

(b) Bestimmen Sie Verteilungen $\mathcal{N}(\boldsymbol{\mu}_{c_i}, \Sigma_{c_i})$, sodass G (i) eine Gerade oder (ii) ein Kreis ist. Skizzieren Sie in beiden Fällen die Verteilungen und die Entscheidungsgrenze G. Wie ändern sich hergeleiteten die Entscheidungsgrenzen qualitativ in Abhängigkeit von $p(y = c_1|\boldsymbol{\theta}) = \lambda \in (0, 1)$.

Im Folgenden sind die bedingten Verteilungen $p(\mathbf{x}|y = c, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_c, \Sigma_c)$ und $p(y = c_1|\boldsymbol{\theta}) = p(y = c_2|\boldsymbol{\theta}) = 1/2$ gegeben. Skizzieren Sie die Verteilungen, bestimmen Sie die Entscheidungsgrenze und skizzieren Sie diese ebenfalls.

(c) $\boldsymbol{\mu}_{c_1} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$, $\Sigma_{c_1} = \mathbf{I}$ und $\boldsymbol{\mu}_{c_2} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$, $\Sigma_{c_2} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1 \end{pmatrix}$.

(d) $\boldsymbol{\mu}_{c_1} = \boldsymbol{\mu}_{c_1} = \mathbf{0}$, $\Sigma_{c_1} = \mathbf{I}$ und $\Sigma_{c_2} = \begin{pmatrix} 1/10 & 0 \\ 0 & 5 \end{pmatrix}$.

(e) Es sei $\boldsymbol{\mu}_{c_1} = \mathbf{0}$ und $\Sigma_{c_1} = \mathbf{I}$. Weiter sei $\boldsymbol{\mu}_{c_2} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$ und $\Sigma_{c_2} = \begin{pmatrix} 1/a & 0 \\ 0 & 1/a \end{pmatrix}$ mit $a > 0$. Leiten Sie die Entscheidungsgrenze $G = G(a)$ in Abhängigkeit von a her. Wie verhält sich $G(a)$ in den Extremfällen $a \to 0$, $a \to 1$ und $a \to \infty$? Skizzieren Sie die Verteilungen und die Entscheidungsgrenze G in diesen 3 Fällen.

Aufgabe 3. 5 P.

Im Folgenden darf benutzt werden, dass für Beobachtungen $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ der ML-Schätzer für multivariate Normalverteilungen gegeben ist durch

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \ , \qquad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^\top \ . \tag{3}$$

Es seien Beobachtungen $\mathbf{x}_1^1, \ldots, \mathbf{x}_{n_1}^1 \in \mathbb{R}^2$ der Klasse c_1 und $\mathbf{x}_1^2, \ldots, \mathbf{x}_{n_2}^2 \in \mathbb{R}^2$ der Klasse c_2 gegeben.

(a) Leiten Sie die Maximum-Likelihood Schätzer für $\boldsymbol{\mu}_{c_i}$, Σ_{c_i} und λ her, unter der Annahme, dass die Beobachtungen durch das Wahrscheinlichkeitsmodell Gleichung (1) erzeugt wurden.

(b) Leiten Sie die Maximum-Likelihood Schätzer her, unter der Annahme, dass $\Sigma_{c_1} = \Sigma_{c_2} = \Sigma_c$.

(c) Leiten Sie die Maximum-Likelihood Schätzer her, unter der Annahme, dass $\Sigma_{c_1} = \sigma_1^2\mathbf{I}$ und $\Sigma_{c_2} = \sigma_2^2\mathbf{I}$ diagonal sind.

(d) Die Beobachtungen sind in folgender Tabelle aufgeführt. Zeichnen Sie diese in ein zweidimensionales Koordinatensystem. Berechnen Sie für die Fälle (a) bis (c) die Entscheidungsgrenzen und zeichnen Sie diese Ebenfalls ein.

Klasse 1

x	1	1	0	2	0	2
y	2	0	1	1	0	2

Klasse 2

x	-1	-1	-2	0	-2	0
y	0	-2	-1	-1	0	-2