

Machine Learning

Übungsblatt 5

20 Punkte

Aufgabe 1. LDA

12 P.

- (a) Nennen Sie die grundlegende Verteilungsannahme, die der Gaußschen Diskriminanzanalyse zugrunde liegt und aufgrund welcher Kriterien Beobachtungen klassifiziert werden.
- (b) Nennen Sie die spezifischen Verteilungsannahmen und Modellparameter der folgenden Modelle:
- Quadratische Diskriminanzanalyse (QDA)
 - Lineare Diskriminanzanalyse (LDA)
 - Naive Bayes
- (c) Zeigen Sie, dass im Falle von nur zwei Klassen, die LDA-Entscheidungsregionen halbräume sind, d.h. dass Sie durch eine Gleichung der Form $\mathbf{w}^\top \mathbf{x} > c$ bzw. $\mathbf{w}^\top \mathbf{x} < c$ beschrieben werden können. Bestimmen Sie außerdem eine Formel für $\mathbf{w} \in \mathbb{R}^d$ und $c \in \mathbb{R}$ in Abhängigkeit der Parameter des LDA Modells.
- (d) Zeigen Sie, dass im Falle von nur zwei Klassen, die QDA-Entscheidungsgrenze die Lösung einer quadratischen Gleichung der Form $\mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{b}^\top \mathbf{x} + c = 0$ ist. Hierbei ist $\mathbf{A} \in \mathbb{R}^{d \times d}$ eine Matrix, $\mathbf{b} \in \mathbb{R}^d$ ein Vektor und $c \in \mathbb{R}$ ein Skalar. Bestimmen Sie außerdem eine Formel für \mathbf{A} , \mathbf{b} und c in Abhängigkeit der Modellparameter.

Gegeben sind nun die folgenden Beobachtungen.

$$\begin{array}{c|cccc|cccc} x_1 & \sqrt{3} & -\sqrt{3} & \sqrt{5} & -\sqrt{5} & 2 + \sqrt{5} & 2 - \sqrt{5} & 2 + \sqrt{3} & 2 - \sqrt{3} \\ x_2 & \sqrt{3} & -\sqrt{3} & -\sqrt{5} & \sqrt{5} & 2 + \sqrt{5} & 2 - \sqrt{5} & 2 - \sqrt{3} & 2 + \sqrt{3} \\ y & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \end{array}$$

Fitten Sie ein QDA Modell an die Daten. Gehen Sie dabei folgendermaßen vor.

- (e) Fitten Sie ein QDA Modell an die Daten. Nutzen Sie dazu die aus der Vorlesung / dem Lehrbuch bekannten Formeln für die Maximum-Likelihood Parameter Schätzungen.
- (f) Skizzieren Sie die Verteilungen $p(x|y = c)$ in dem Sie mehrere Niveaumengen der entsprechenden Dichtefunktionen handschriftlich in ein Koordinatensystem einzeichnen. Achten Sie dabei besonders auf die Form der Niveaumengen.
- (g) Setzen Sie die Modellparameter in die Gleichung für die Entscheidungsgrenze aus Teilaufgabe (d) ein. Lösen Sie die Gleichung nach x_2 (also nach der zweiten Koordinate von \mathbf{x}).
- (h) Beschreiben Sie die Form der Entscheidungsgrenze und zeichnen Sie sie in ihre Skizze aus Teilaufgabe (f) ein.

Aufgabe 2. LDA als Dimensionsreduktion

8 P.

Gegeben seien klassenweise Verteilungen $X_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, $k \in \{1, 2\}$ mit jeweiligen Erwartungswerten $\boldsymbol{\mu}_k \in \mathbb{R}^d$ und Kovarianzmatrizen $\boldsymbol{\Sigma}_k \in \mathbb{R}^{d \times d}$.

- (a) Es sei $\mathbf{v} \in \mathbb{R}^d$. Bestimmen Sie die Verteilungen der univariaten Zufallsvariablen $Y_1 = \mathbf{v}^\top X_1$ und $Y_2 = \mathbf{v}^\top X_2$. Was ist die geometrische Interpretation der Y_k ?

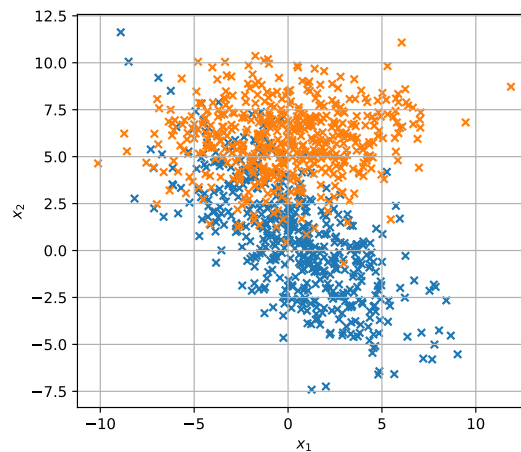
Wir betrachten nun die Größe

$$J(\mathbf{v}) = \frac{\mathbf{v}^\top (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \mathbf{v}}{\mathbf{v}^\top (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \mathbf{v}}.$$

- (b) Was ist der Zusammenhang zwischen $J(\mathbf{v})$ und Y_1, Y_2 ? Was ist die Interpretation von $\mathbf{w} = \arg \max_{\mathbf{v} \in \mathbb{R}^d} J(\mathbf{v})$?

- (c) Bestimmen Sie die Maximierer \mathbf{w} .
Hinweis. Für alle $a \in \mathbb{R}$ ist $J(a\mathbf{v}) = J(\mathbf{v})$.
- (d) Falls die Verteilungen der X_k unbekannt sind, hätten wir keinen Zugriff auf deren Parameter und würden stattdessen (Maximum-Likelihood) Schätzungen verwenden. Vergleichen Sie die resultierenden Formel für \mathbf{w} mit der Entscheidungsgrenze eines LDA Modells, das mithilfe der Maximum-Likelihood Methode an die Daten angepasst wurde.

Die folgende Abbildung zeigt ein 2-dimensionales Beispiel mit $p(\mathbf{x}|y = i) = \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ und $p(y = 1) = p(y = 2)$, wobei $\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 6 \end{pmatrix}$, $\boldsymbol{\Sigma}_1 = \begin{pmatrix} 11 & 0 \\ 0 & 4 \end{pmatrix}$, und $\boldsymbol{\mu}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\boldsymbol{\Sigma}_2 = \begin{pmatrix} 10 & -7 \\ -7 & 10 \end{pmatrix}$



- (e) Bestimmen Sie ein $\mathbf{w} = \arg \max_{\mathbf{v}} J(\mathbf{v})$ und zeichnen Sie die Wahrscheinlichkeitsdichten der resultierenden Zufallsvariablen $\mathbf{w}^\top X_1$ und $\mathbf{w}^\top X_2$ in ein Koordinatensystem ein. Machen Sie das gleiche für $\mathbf{w} = \begin{pmatrix} 1 & 0 \end{pmatrix}^\top$ und $\mathbf{w} = \begin{pmatrix} 0 & 1 \end{pmatrix}^\top$ und vergleichen Sie die Abbildungen.