

Machine Learning

Übungsblatt 4

25 Punkte

Aufgabe 1.

4 P.

Nennen Sie die grundlegende Verteilungsannahme, die der Gaußschen Diskriminanzanalyse zugrunde liegt. Welche spezifischen Annahmen entsprechen den folgenden Modellen:

- (a) Quadratische Diskriminanzanalyse (QDA)
- (b) Lineare Diskriminanzanalyse (LDA)
- (c) Naive Bayes

Aufgabe 2.

7 P.

- (a) Bestimmen Sie die Entscheidungsgrenze eines zwei-Klassen GDA Modells unter der Annahme, dass $p(y = 1) = p(y = 2)$, $p(\mathbf{x}|y = 1, \boldsymbol{\mu}_1) = \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{1})$ und $p(\mathbf{x}|y = 2, \boldsymbol{\mu}_2) = \mathcal{N}(\boldsymbol{\mu}_2, \mathbf{1})$.

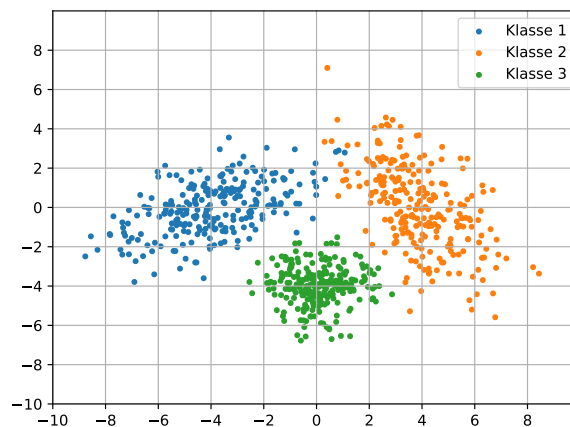


Abbildung 1: Daten mit zwei Features. Für jede Klasse sind 250 Beobachtungen gegeben.

- (b) Wir betrachten nun die Daten aus Abbildung 1. Bestimmen Sie die Entscheidungsgrenzen eines GDA Modells unter den Annahmen $p(y = 1) = p(y = 2) = p(y = 3)$ und $p(\mathbf{x}|y = c, \boldsymbol{\mu}_c) = \mathcal{N}(\boldsymbol{\mu}_c, \mathbf{1})$ für alle $c \in \{1, 2, 3\}$. Zeichnen Sie die Entscheidungsregionen handschriftlich in Abbildung 1 ein. Erklären Sie Ihr Vorgehen.
- (c) Wir modellieren die Daten nun mittels QDA. Schätzen Sie hierfür die Parameter des Modells per Augenmaß. Dabei genügt es, wenn Sie die Kovarianzmatrizen $\boldsymbol{\Sigma}_c$ bis auf einen Faktor genau schätzen, d.h. $\hat{\boldsymbol{\Sigma}}_c \approx t\boldsymbol{\Sigma}_c$ für ein $t > 0$.

Aufgabe 3.

7 P.

Wir betrachten nun den Datensatz, der in Abbildung 2 dargestellt ist.

- (a) Lassen sich die (Trainings-)daten mittels eines LDA Modells zu 100% korrekt klassifizieren. Wie ist es im Falle eines QDA Modells? Begründen Sie Ihre Antwort.
- (b) Schätzen Sie die Parameter des QDA Modells per Augenmaß und berechnen Sie die Entscheidungsgrenze. Wie in Aufgabe 3 sind die genauen Werte der Schätzungen für $\boldsymbol{\Sigma}_c$ unwesentlich, aber die grundsätzliche Form sollte stimmen.

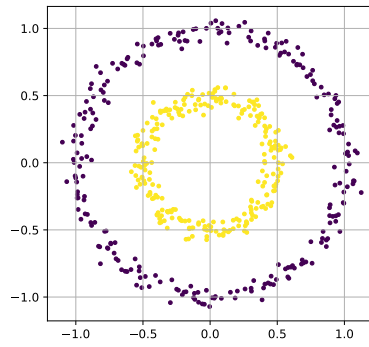


Abbildung 2: Daten mit zwei Features. Für jede Klasse sind 250 Beobachtungen gegeben.

- (c) Anstelle von 250 Beobachtungen pro Klasse haben wir nun 100 Beobachtungen der Klasse violett und 400 Beobachtungen der Klasse gelb. Wie ändert sich die Entscheidungsgrenze des QDA Modells qualitativ (keine Rechnung erforderlich).

Aufgabe 4.

7 P.

Wir betrachten nun den Datensatz, der in Abbildung 3 dargestellt ist.

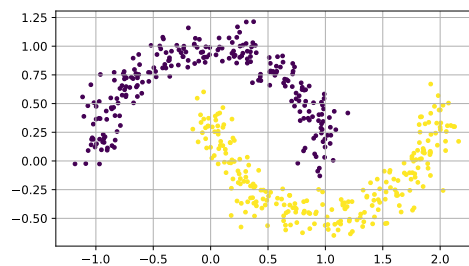


Abbildung 3: Daten mit zwei Features. Für jede Klasse sind 250 Beobachtungen gegeben.

- (a) Lassen sich die (Trainings-)daten mittels eines LDA Modells zu 100% korrekt klassifizieren. Wie ist es im Falle eines QDA Modells? Begründen Sie Ihre Antwort.
- (b) Der Maximum-Likelihood Schätzer der Modellparameter des QDA Modells ergibt $\hat{\Sigma}_{\text{violett}} \approx \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{10} \end{pmatrix}$, $\hat{\mu}_{\text{violett}} \approx \begin{pmatrix} 0 \\ 0.6 \end{pmatrix}$ und $\hat{\mu}_{\text{gelb}} \approx \begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$. Berechnen Sie die Entscheidungsgrenze des QDA Modells und zeichnen Sie sie handschriftlich in Abbildung 3 ein.
- (c) Anstelle von 250 Beobachtungen pro Klasse haben wir nun 100 Beobachtungen der Klasse violett und 400 Beobachtungen der Klasse gelb. Bestimmen Sie die neue Entscheidungsgrenze und zeichnen Sie diese ein.