

## Machine Learning

In den Aufgaben 1 und 2 dürfen Sie technische Hilfsmittel (Matlab, Numpy, WolframAlpha, Excel, ...) zur Berechnung der Lösungen benutzen. **Achten Sie darauf, dennoch ihren Lösungsweg nachvollziehbar zu dokumentieren und relevante Zwischenergebnisse anzugeben.**

Aufgaben 3 und 4 sind **Verständnisaufgaben** und sollen **ohne technische Hilfsmittel** gelöst werden. Übermäßige Nutzung technischer Hilfsmittel kann zu Punktabzug bei diesen Aufgaben führen.

**Aufgabe 1.** Diskriminanzanalyse

7 P.

Gegeben sind die folgenden Trainingsdaten:

$$\{(x_i, y_i)\} = \{(0, 4), (1, 2), (2, 0), (3, 0), (4, 2), (5, 3), (6, 5), (7, 8), (8, 9), (9, 11), (10, 13), (11, 15), (12, 16), (13, 15), (14, 14), (15, 13)\}.$$

Wir nehmen an, dass die Daten sich durch ein probabilistisches Modell der Form  $p(y|x, \theta) = \mathcal{N}(\mathbf{w}^\top \phi(x), \sigma^2)$  beschreiben lassen. Schätzen Sie den Parameter  $\mathbf{w}$  folgendermaßen.

- (a) Mittels Maximum-Likelihood Schätzung, wobei  $\phi : x \mapsto (1, x)^\top$ .
- (b) Mittels Maximum-a-posteriori Schätzung mit einer Gaußschen Priorverteilung über  $\mathbf{w}$  mit Mittelwert  $\mathbf{0}$  und Kovarianzmatrix  $\mathbf{I}/4$ . Nehmen Sie dazu an, dass  $\sigma^2 = 5$  und  $\phi : x \mapsto (1, x, x^2, x^3)^\top$ .
- (c) Bestimmen Sie für beide Modelle die Residuenquadratsumme und die Vorhersage für den Wert  $x = -1$ . Geben Sie außerdem die Steigung der Funktion  $x \mapsto \mathbf{w}^\top \theta(x)$  an der Stelle  $x = 8$  an.

**Aufgabe 2.** Lineare Regression

9 P.

Gegeben sind die folgenden Trainingsdaten:

$$\{(x_i, y_i)\} = \{(1., 1.), (2., 2.), (3., 3.), (1.5, 2.5), (2.5, 1.5), \quad \text{Klasse A} \\ (-1., 1.), (-2., 2.), (-3., 3.), (-1.5, 2.5), (-2.5, 1.5), \quad \text{Klasse B} \\ (0., 0.), (0., -1.), (0., -2.), (-1., -1.), (1., -1.) \quad \text{Klasse C}\}.$$

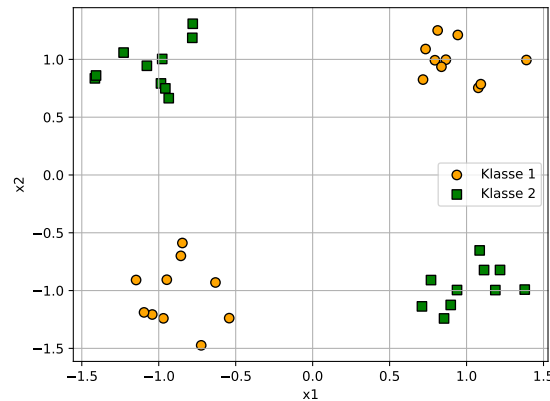
Fitten Sie mithilfe der Maximum-Likelihood Methode die folgenden Klassifizierungsmodelle an die Daten, unter der Annahme, dass  $p(y = c|\theta) = \text{Cat}(y|\theta)$  kategorisch verteilt ist.

- (a) Ein Gaußsches Diskriminanzanalyse Modell.
- (b) Ein Lineares Diskriminanzanalyse Modell.
- (c) Skizzieren Sie für beide Modelle die geschätzten Verteilungen  $p(\mathbf{x}|y, \theta)$ .
- (d) Bestimmen Sie die Entscheidungsregionen des Linearen Diskriminanzanalyse Modells und zeichnen Sie sie in Ihre Skizze ein.
- (e) Bestimmen für beide Modelle die Klassenvorhersage der Modelle für den Punkt  $(1, 0.5)$ .

**Aufgabe 3. Verrauschtes XOR-Problem**

7 P.

Gegeben sind die folgenden Trainingsdaten.



Geben Sie an, ob die folgenden Modelle alle Trainingsdaten korrekt zu klassifizieren. Begründen Sie Ihre Antworten.

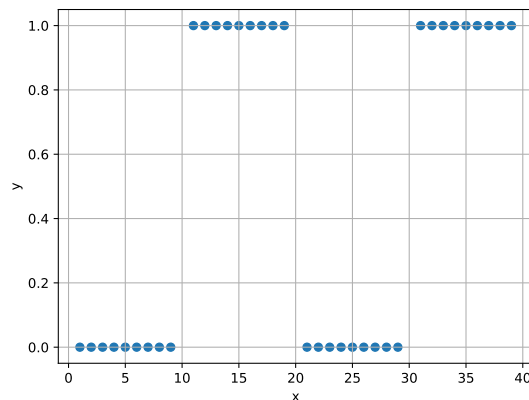
- (a) Ein Entscheidungsbaum der maximalen Tiefe zwei (also maximal 4 Regionen), der mit dem standard greedy Algorithmus gefittet wird (wie auf Blatt 11).
- (b) Ein logistisches Regressionsmodell dessen Parameter mit der Maximum-Likelihood Methode geschätzt werden.
- (c) Gaußsche Diskriminanzanalyse (GDA) dessen Parameter mit der Maximum-Likelihood Methode geschätzt werden.

Hinweis: Ein ausreichend genauer Näherungswert der Maximum-Likelihood Schätzungen kann visuell bestimmt werden.

**Aufgabe 4. Periodisches Signal**

7 P.

Gegeben sind die folgenden Trainingsdaten auf die ein Regressionsmodell der Form  $f(x, \theta) = y$  gefittet werden soll.



Fitten Sie auf die Daten entweder ein Lineares Regressionsmodell mit kubischen Features (Polynom vom Grad 3) oder einen Regressionsbaum der Tiefe 3. Nutzen Sie zum Fitten die Maximum-Likelihood Methode (falls Lineare Regression) bzw. den standard greedy Algorithmus (falls Regressionsbaum).