Klausur

APPLIED MACHINE LEARNING FUNDAMENTALS

Data Science, WWI 19DS B, DHBW Mannheim

Matrikelnummer:

7. Mai 2021, 12:30 Uhr - 13:30 Uhr

Hinweise:

1. Bitte überprüfen Sie, ob Sie **alle 13 Aufgabenblätter** (ausgenommen des Deckblatts) erhalten haben **bevor** Sie mit der Bearbeitung der Klausur beginnen. Bitte wenden Sie sich an die Prüfungsaufsicht, falls Ihr Druckexemplar unvollständig sein sollte.
2. Vergessen Sie nicht, Ihre Matrikelnummer auf der Klausur anzugeben. **Bitte verwenden Sie nicht Ihren Namen (Anonymisierung)!**
3. Notieren Sie Ihre Antworten direkt auf den Aufgabenblättern. Benutzen Sie gegebenenfalls die leeren Rückseiten oder die letzte Seite dieser Klausur.
4. Es steht Ihnen frei, die Fragen entweder auf Englisch oder auf Deutsch zu beantworten. Bitte übersetzen Sie keine technischen Begriffe, um Verwirrung zu vermeiden.
5. Die Klausur besteht aus 60 Punkten und ist **innerhalb von 60 Minuten** zu lösen. Die maximal zu erreichenden Punkte pro Aufgabe sind jeweils angegeben. Nutzen Sie sie als Hinweis darauf, wie umfangreich Ihre Antworten sein sollten.
6. Bitte schalten Sie alle Kommunikationsgeräte aus. Die folgenden Hilfsmittel sind erlaubt: ❶ Nicht programmierbarer Taschenrechner ❷ Zweiseitig handbeschriebenes Cheat Sheet
7. **Verstöße gegen die Prüfungsordnung werden als Täuschungsversuch gewertet!**

Aufgabe	**1**	**2**	**3**	**4**	**5**	**gesamt**
Punkte	/14	/14	/10	/8	/14	/60

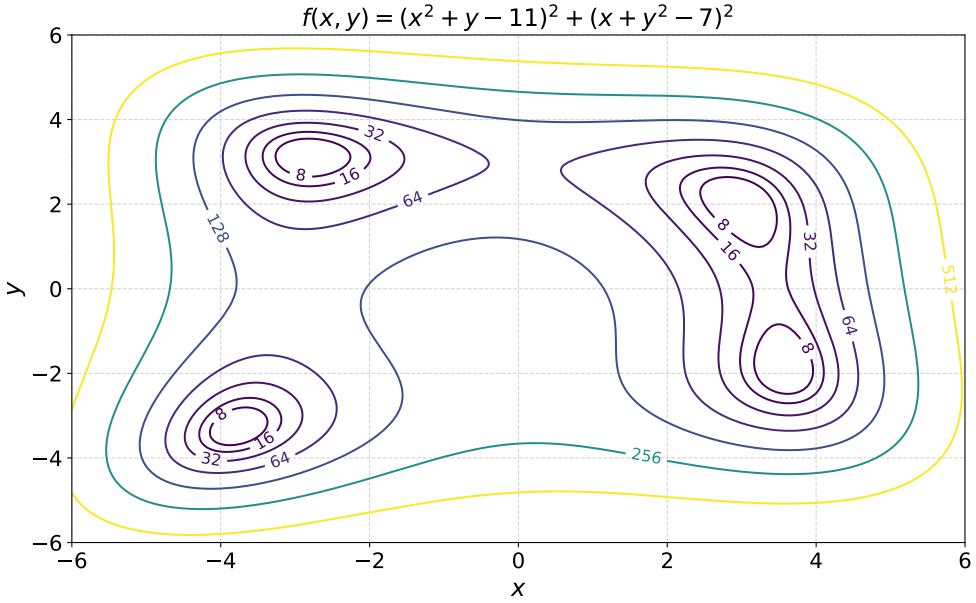
1 Optimierung mit Gradient Descent

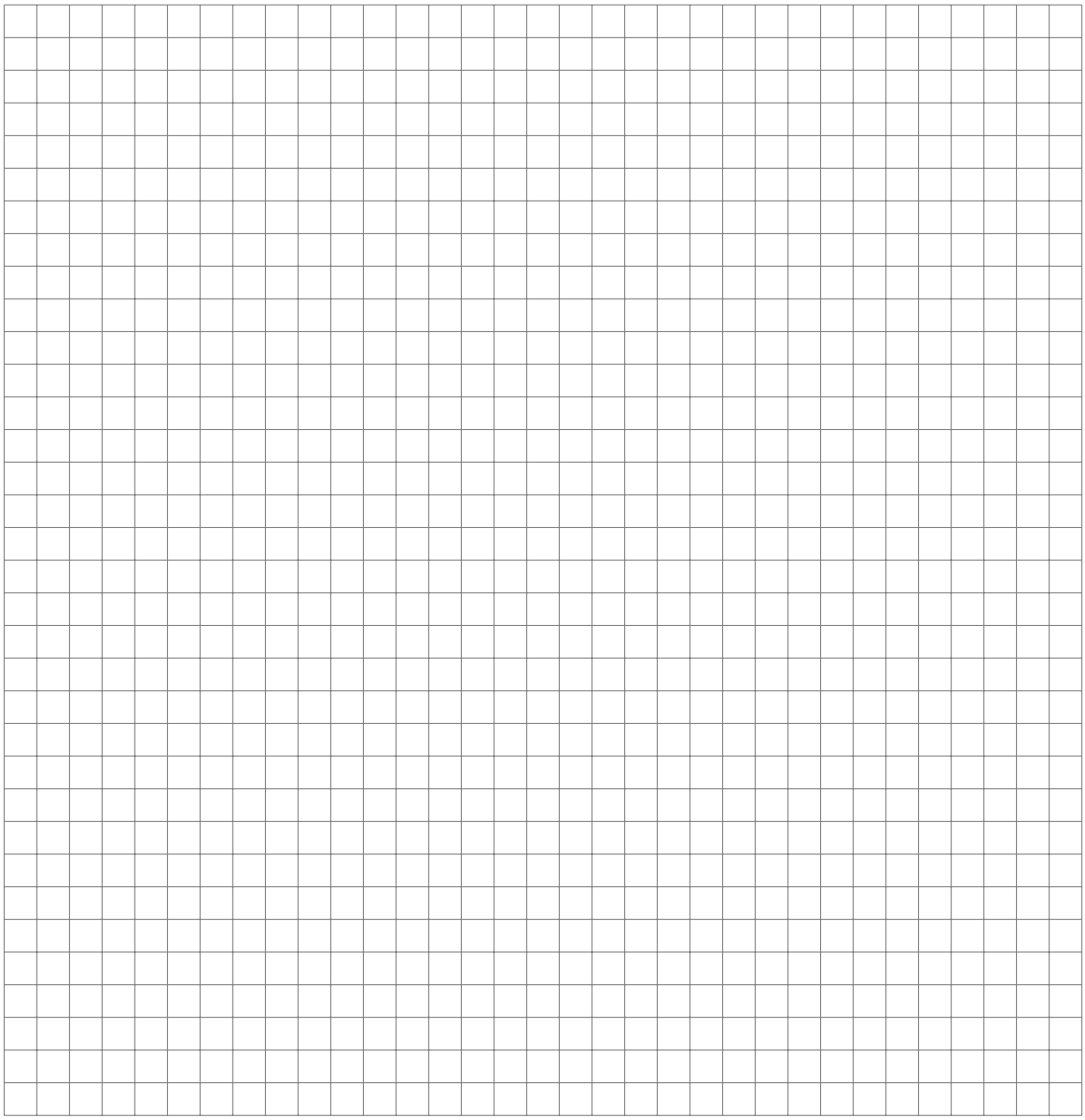
1.1 Sei $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$

eine reellwertige Funktion. Führen Sie zwei Iterationen des *Gradient Descent* Algorithmus durch. Nehmen Sie hierzu an, dass die Lernrate $\alpha = 0.02$ beträgt. Starten Sie mit $x_0 = y_0 = 0$, wobei der Index für die Anzahl durchgeführter Iterationen steht. Geben Sie die Werte für x_2 und y_2 an. Wie lautet der zugehörige Funktionswert $f(x_2, y_2)$?

Runden Sie alle Ergebnisse auf zwei Nachkommastellen. **(8 p)**





1.2 Handelt es sich bei obiger Funktion um eine konvexe Funktion? Begründen Sie Ihre Antwort. **(2 p)**

1.3 Welche Aussagen bezüglich des *Gradient Descent* Algorithmus sind korrekt? **(2 p)**

- [ ] Der stochastische Gradient ist am genauesten, jedoch aufwändig zu berechnen.
- [ ] Die zu optimierende Funktion muss differenzierbar sein.
- [ ] Der *Gradient Descent* Algorithmus ist ein analytisches Verfahren.
- [ ] Eine kleine Lernrate (z. B. $\alpha = 0.001$) begünstigt die Konvergenz des Algorithmus.
- [ ] Die Lernrate sollte deutlich größer als eins gewählt werden.
- [ ] *Gradient Descent* findet in jedem Fall das globale Optimum.

1.4 Angenommen, Sie möchten ein Maximum einer gegebenen Funktion bestimmen. Wie müsste die entsprechende Update-Regel für einen *Gradient Ascent* Algorithmus aussehen? **Hinweis:** Auf deutsch nennt man dieses Verfahren Gradientenaufstieg. **(2 p)**

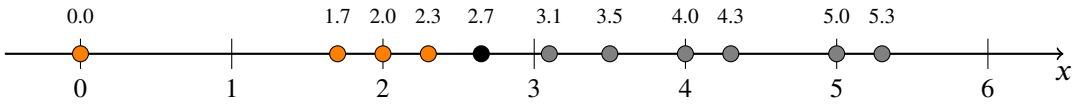
Maximal erreichbare Punkte für Aufgabe 1: **14 Punkte**

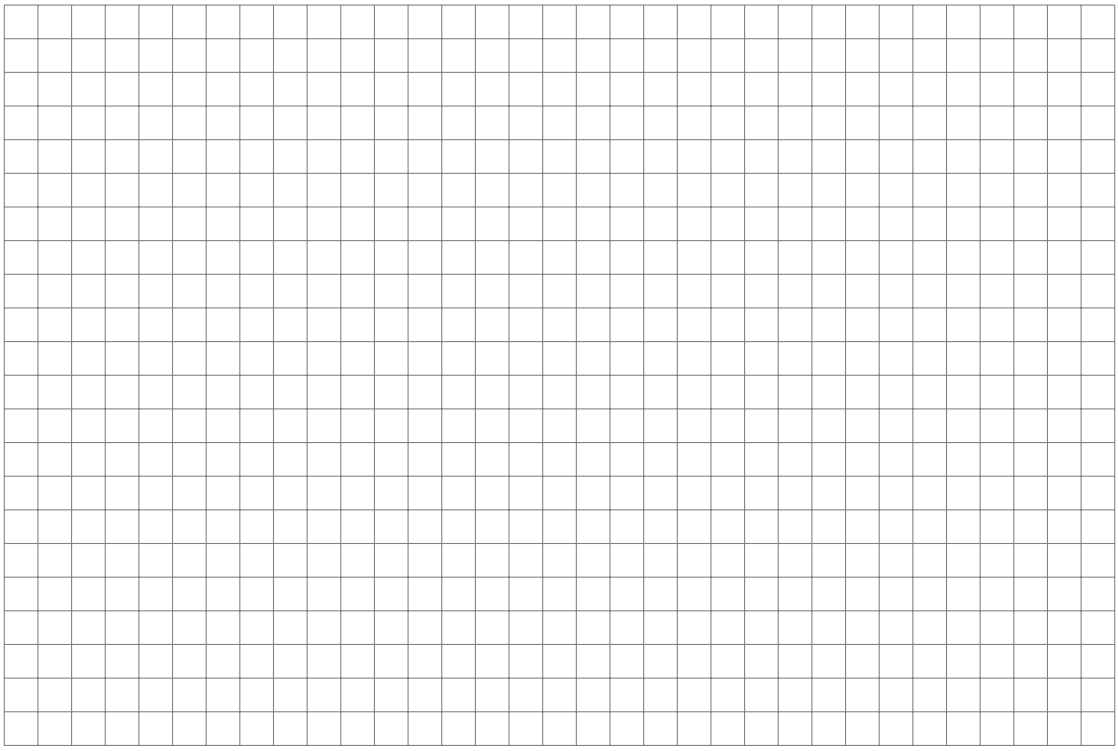
2 Bayes'sche Entscheidungstheorie

2.1 Ihnen liegt der unten abgebildete eindimensionale Datensatz vor. Dieser besteht aus zwei Klassen (orange und grau). Schätzen Sie für jeweils beide Klassen eine Gauß'sche Normalverteilung (zur Erinnerung: $\mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$) aus den Daten, d. h. jeweils den Mittelwert μ sowie die Varianz σ^2.

Wie hoch ist die Wahrscheinlichkeit, dass der unbekannte Punkt (schwarz) mit $x = 2.7$ zu Klasse orange bzw. Klasse grau gehört? Nutzen Sie die berechneten Wahrscheinlichkeiten, um für diesen Punkt eine Klassifikation anzugeben.

Runden Sie alle Ergebnisse auf zwei Nachkommastellen. **(7 p)**





2.2 Für was steht die Abkürzung i. i. d. und was bedeutet sie? **(2 p)**

2.3 Wann ist ein Klassifizierer *Bayes optimal*? **(2 p)**

2.4 Geben Sie Bayes' Theorem wieder und beschreiben Sie kurz dessen Komponenten. **(3 p)**

Maximal erreichbare Punkte für Aufgabe 2: **14 Punkte**

3 Evaluation von Machine Learning Modellen

3.1 Sie wenden ein trainiertes Logistic Regression Modell auf einen Testdatensatz an. Dieser enthält zehn Beispiele, die entweder zu Klasse $\oplus$ oder Klasse $\ominus$ gehören. Das Modell gibt die untenstehenden Wahrscheinlichkeiten (für Klasse $\oplus$) aus. Als Schwellenwert (*threshold*) sei 0.4 gegeben. Bestimmen Sie die Klassifikationen und ergänzen Sie die Konfusionsmatrix. **(4 p)**

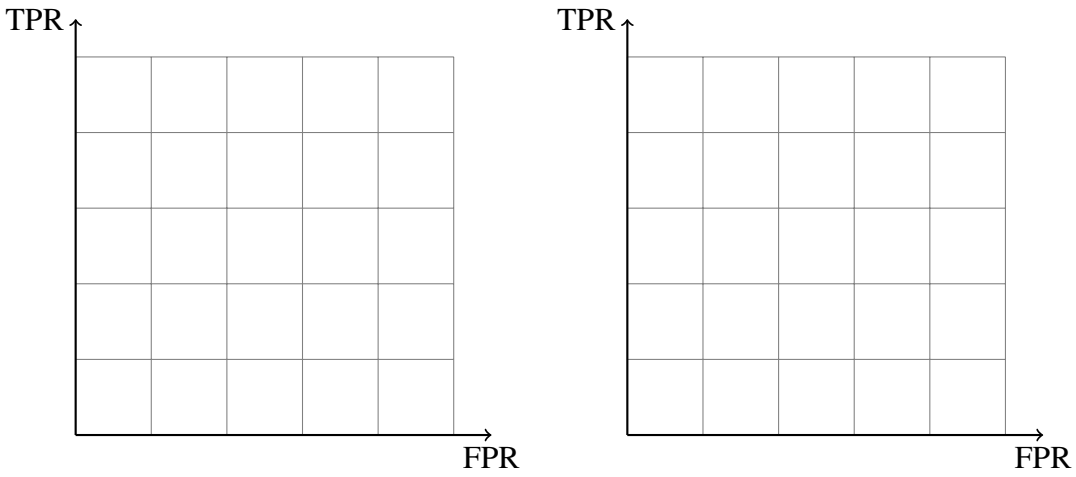
Beispiel	**Gold Label**	**Wahrscheinlichkeit**	**Vorhersage**
1	$\oplus$	0.95	
2	$\oplus$	0.30	
3	$\ominus$	0.35	
4	$\ominus$	0.10	
5	$\oplus$	0.80	
6	$\oplus$	0.55	
7	$\ominus$	0.25	
8	$\ominus$	0.75	
9	$\ominus$	0.05	
10	$\oplus$	0.20	

Konfusionsmatrix		**gold**	
		$\oplus$	$\ominus$
vorhergesagt	$\oplus$		
	$\ominus$		

3.2 Begünstigt der oben gewählte Schwellenwert (*threshold*) von 0.4 eher *False Positives* oder eher *False Negatives*? **(1 p)**

3.3 Betrachten Sie erneut den Datensatz aus Aufgabe 3.1. Zeichnen Sie die ROC-Kurve (*Receiver Operating Characteristic*) und berechnen Sie die Fläche unter der Kurve (AUC). **Hinweis:** Der ROC-Raum besteht aus 25 Kästchen. Wie ist der Klassifizierer angesichts dieses Wertes zu beurteilen?

Das rechte Koordinatensystem ist für den Fall, dass das Zeichnen der ROC-Kurve nicht auf Anhieb funktioniert. **(4 p)**



3.4 Welche Aussage bezüglich *Bias* und *Variance* ist korrekt? **(1 p)**

- Modelle mit einem hohen *Bias* tendieren zu *Overfitting*.
- Eine hohe *Variance* kann durch mehr Trainingsbeispiele reduziert werden.
- Ein Entscheidungsstumpf hat einen niedrigen *Bias*.
- *Bias* und *Variance* haben nichts mit *Overfitting* bzw. *Underfitting* zu tun.
- Keine der Aussagen ist korrekt.

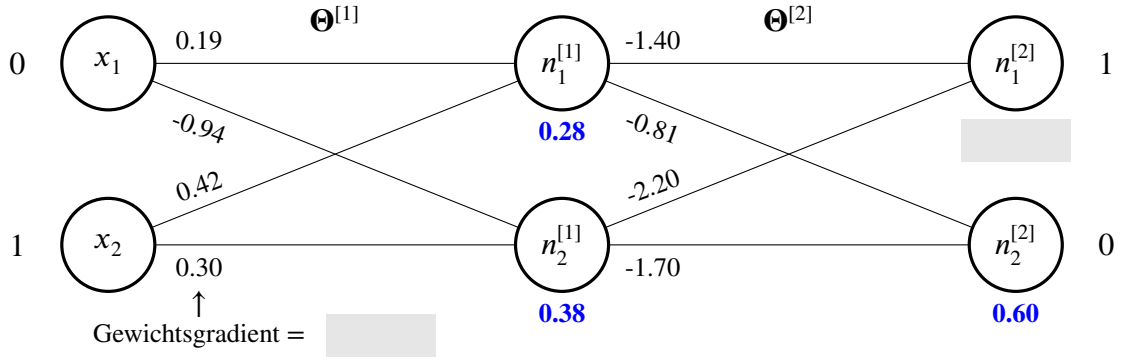
Maximal erreichbare Punkte für Aufgabe 3: **10 Punkte**

4 Neuronale Netze

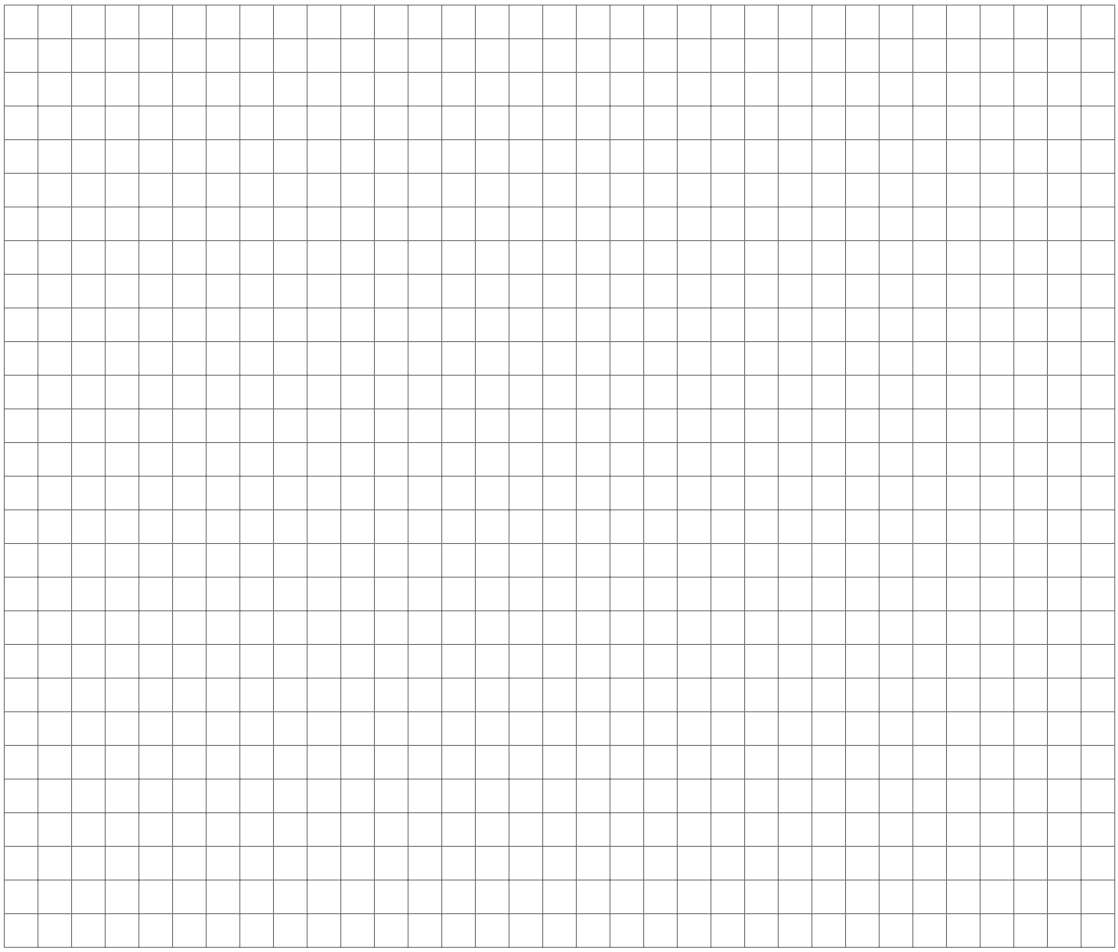
4.1 Unter welchen Umständen konvergiert der *Perceptron* Lernalgorithmus? **(1 p)**

4.2 Sie trainieren das unten abgebildete neuronale Netz mit dem stochastischen *Gradient Descent* Algorithmus. Das Trainingsbeispiel in der momentanen Iteration ist durch den Featurevektor $\boldsymbol{x} = (0, 1)$ gegeben. Das dazugehörige Label ist *one-hot* kodiert und lautet $\boldsymbol{y} = (1, 0)$. Als Fehlerfunktion wird der quadratische Fehler (*squared error*) benutzt. **Hinweis:** $\mathcal{J}' = 2 \cdot (z_{n_k^{[2]}} - y_k)$.

Einige der (Prä-)Aktivierungen (siehe Tabelle) bzw. Fehlergradienten (in blau) wurden bereits ermittelt. Ergänzen Sie die noch fehlenden Werte. **(4 p)**



Neuron	Präaktivierung	Aktivierung	Aktivierungsfunktion
$n_1^{[1]}$	0.42		ReLU
$n_2^{[1]}$	0.30	0.30	ReLU
$n_1^{[2]}$	-1.25	0.22	Sigmoid
$n_2^{[2]}$		0.30	Sigmoid



4.3 Welche Aussagen bezüglich der Aktivierungsfunktionen sind korrekt? **(2 p)**

- Aktivierungsfunktionen sollten nicht linear sein.
- Die *Softmax*-Aktivierungsfunktion wird üblicherweise im letzten *Layer* eingesetzt.
- Ein Problem der *ReLU*-Funktion ist der sogenannte *Vanishing Gradient*.
- Die *ReLU*-Aktivierung wird gemäß der Formel $\min(0, x)$ berechnet.

4.4 Welche Art von neuronalen Netzen wird üblicherweise zur Klassifikation von Bildern eingesetzt? **(1 p)**

Maximal erreichbare Punkte für Aufgabe 4: **8 Punkte**

5 Gemischte Aufgaben

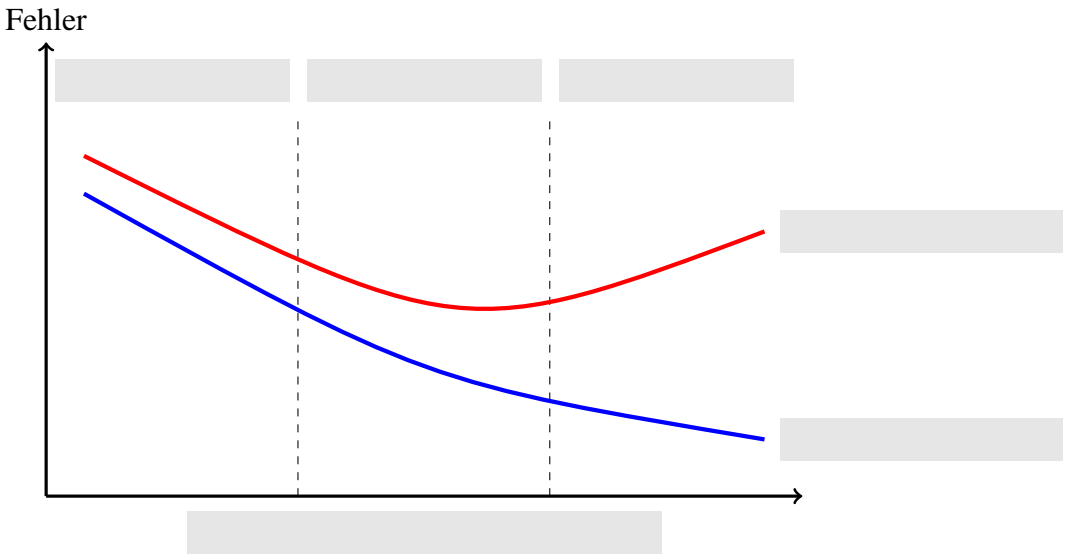
5.1 Berechnen Sie die Entropie des Datensatzes $\mathcal{D} = \{A, A, B, C, B, A, C, C, A, C\}$, welcher aus den drei Klassen A, B und C besteht. **(2 p)**

5.2 Skizzieren Sie den k-Means Algorithmus in eigenen Worten. Welche Schritte müssen durchlaufen werden? Findet der Algorithmus immer das Optimum? **(4 p)**

5.3 Nennen Sie ein Beispiel für eine Basisfunktion (*basis function*), die im Rahmen der linearen Regression verwendet werden kann. **(1 p)**

5.4 Was versteht man unter *Occam's razor* im Kontext von Machine Learning? **(2 p)**

5.5 Vervollständigen Sie die untenstehende Grafik. Nutzen Sie die folgenden Begriffe: *Underfitting*, *Overfitting*, *gutes Modell*, *Trainingsfehler*, *Testfehler*, *Modellkomplexität*. **(3 p)**



5.6 Was versteht man unter *early stopping*? **(2 p)**

Maximal erreichbare Punkte für Aufgabe 5: **14 Punkte**

Zusätzlicher Platz für Anmerkungen:

Maximal erreichbare Punkte für die Klausur: **60 Punkte**