# Klausur

## Applied Machine Learning Fundamentals

Data Science, WWI 20DS B, DHBW Mannheim
Matrikelnummer:

4. Februar 2022, 09:00 Uhr - 10:00 Uhr

---

**Hinweise:**

1. Bitte überprüfen Sie, ob Sie **alle 12 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** | **gesamt** |
|---|---|---|---|---|---|
| **Punkte** | /20 | /15 | /11 | /14 | /60 |

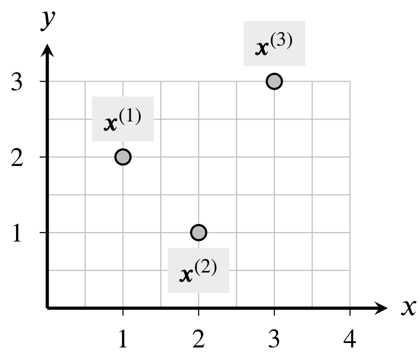
# 1 Lineare Regression

1.1 Ihnen liegt ein aus drei Trainingsbeispielen bestehender Datensatz $\mathcal{D}_{\text{train}}$ vor. Dieser Datensatz sei gegeben durch $\mathcal{D}_{\text{train}} = \{(1,2),(2,1),(3,3)\}$. Jedes Trainingsbeispiel ist ein Tupel der Form $(x, y)$, wobei $x$ das Feature, und $y$ das dazugehörige Label darstellt.

Bestimmen Sie mittels eines **analytischen** Verfahrens die Parameter der Regressionsfunktion $h_\theta(x) = \theta_0 + \theta_1 x$ (d. h. **ohne** Verwendung von Basisfunktionen). Zeichnen Sie die Regressionsgerade in das untenstehende Diagramm ein.

**Hinweise:**

- Allgemeine Lösungsformel: $\boldsymbol{\theta} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{y}$
- Sei $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$ invertierbar. Es ist $A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ **(8 p)**

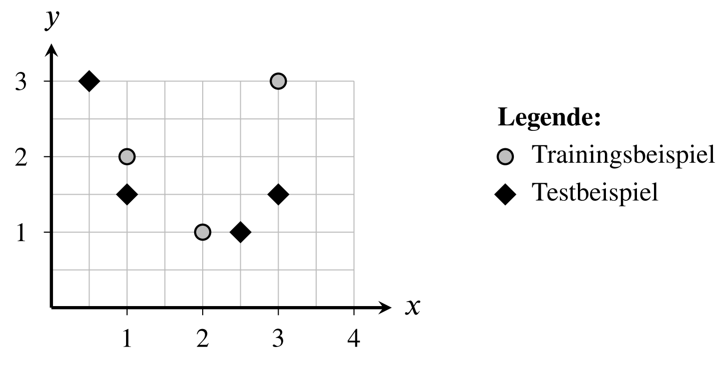


1.2 Für den Test Ihres in Aufgabe 1.1 trainierten Regressionsmodells steht Ihnen ein separater Testdatensatz $\mathcal{D}_{\text{test}}$ zur Verfügung. Es ist $\mathcal{D}_{\text{test}} = \{(0.5, 3), (1, 1.5), (2.5, 1), (3, 1.5)\}$, wobei jedes Tupel von derselben Form wie in Aufgabe 1.1 ist.

Berechnen Sie den RMSE (*Root Mean Square Error*) des Modells auf diesem Datensatz.

**Hinweis:** Legen Sie der Berechnung $\theta = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$ zugrunde. **(3 p)**

1.3 Die folgende Grafik visualisiert den Trainingsdatensatz $\mathcal{D}_{\text{train}}$ und den Testdatensatz $\mathcal{D}_{\text{test}}$. Wie müssen Sie das Modell aus Aufgabe 1.1 angesichts dieser Grafik bewerten? Unter welchem Phänomen leidet Ihr Modell? Welche Maßnahme können Sie ergreifen, um es zu verbessern? **(3 p)**



1.4 Welche Aussagen zur Regularisierung eines Regressionsmodells sind korrekt? **(2 p)**

- ☐ Regularisierung reduziert die Gefahr von Underfitting.
- ☐ Regularisierung reduziert die Gefahr von Overfitting.
- ☐ Der Regularisierungsparameter $\lambda$ sollte so groß wie möglich gewählt werden.
- ☐ Die Lösungsformel für *Ridge*-Regression lautet $\boldsymbol{\theta} = (\boldsymbol{X}^\intercal \boldsymbol{X} + \lambda \boldsymbol{I})^{-1} \boldsymbol{X}^\intercal \boldsymbol{y}$.
- ☐ Regularisierung schließt die Verwendung von Basisfunktionen aus.

1.5 Sie erfahren, dass Sie in Kürze einen **weitaus größeren Datensatz** für das Training Ihres Modells erhalten werden. Ein Kollege weist Sie darauf hin, dass die Berechnung einer analytischen Lösung für sehr große Datensätze **zu aufwändig** ist.

Welches Verfahren können Sie stattdessen anwenden? Beschreiben Sie stichpunktartig dessen Funktionsweise. **(4 p)**

*Maximal erreichbare Punkte für Aufgabe 1:* **20 Punkte**

# 2 Bayes'sche Entscheidungstheorie

2.1 Welcher Annahme ist es geschuldet, dass man einen *Naïve Bayes* Klassifikator als **naiv** bezeichnet? Gehen Sie auch darauf ein, warum es in der Regel sinnvoll ist, diese Annahme zu treffen. **(3 p)**

2.2 Sie sind in einem Versandhaus angestellt. Ihre Aufgabe ist es, einen Klassifikator zu trainieren, welcher anhand vorgegebener Kundenmerkmale entscheidet, ob den betreffenden Kunden ein Kauf auf Rechnung gestattet werden kann. Sie entscheiden sich für einen *Naïve Bayes* Klassifikator.

Heute Morgen ging bei Ihnen eine Bestellung eines Kunden mit den folgenden Merkmalen ein: $\boldsymbol{x} = [\textit{Neukunde}, \textit{hoch}, \textit{niedrig}, \textit{Ausland}]$. Berechnen Sie alle notw. Wahrscheinlichkeiten anhand des gegebenen Datensatzes, um eine Klassifikation vorzunehmen. **(6 p)**

| **Kundenart (KA)** | **Zahl.geschw. (ZG)** | **Kauffrequenz (KF)** | **Herkunft (H)** | **Rechnung (R)** |
|---|---|---|---|---|
| Neukunde | hoch | niedrig | Inland | ja |
| Neukunde | niedrig | niedrig | Ausland | nein |
| Neukunde | niedrig | hoch | Inland | nein |
| Normalkunde | niedrig | hoch | Inland | ja |
| Neukunde | hoch | niedrig | Inland | nein |
| Normalkunde | hoch | hoch | Ausland | ja |
| Normalkunde | hoch | hoch | Inland | ja |

2.3 Im Laufe des Tages erhalten Sie eine weitere Bestellung. Die Merkmale des Kunden lauten $\boldsymbol{x} = [\textit{Stammkunde}, \textit{niedrig}, \textit{hoch}, \textit{Ausland}]$. Auf welches Problem stoßen Sie und wie können Sie es umgehen?

**Hinweis:** Sie müssen hier keine Berechnungen vornehmen. **(2 p)**

2.4 Geben Sie Bayes' Theorem wieder und beschreiben Sie kurz dessen Komponenten. Warum kann man den Nenner zum Zwecke der Klassifikation ignorieren? **(4 p)**

*Maximal erreichbare Punkte für Aufgabe 2:* **15 Punkte**

# 3 Principal Component Analysis

3.1 Nennen Sie zwei Anwendungsbereiche der PCA (*Principal Component Analysis*). **(2 p)**

3.2 Welche Aussagen bzgl. der PCA (*Principal Component Analysis*) sind **falsch**? **(2 p)**

- [ ] Das Verfahren gehört dem Bereich des überwachten Lernens (*supervised*) an.
- [ ] Der Algorithmus versucht die Varianz der Daten weitestgehend zu beseitigen.
- [ ] Die Daten werden linear transformiert.
- [ ] Alle Hauptachsen (*principal components*) stehen senkrecht aufeinander.
- [ ] Alle Aussagen sind richtig.

3.3 Sei $\mathbf{\Sigma}$ eine reelle Kovarianzmatrix. Welche Aussagen über $\mathbf{\Sigma}$ sind richtig? **(3 p)**

- [ ] Es gilt stets $\mathbf{\Sigma}^\intercal = \mathbf{\Sigma}$ (d. h. $\mathbf{\Sigma}$ ist symmetrisch).
- [ ] $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ mit $m \neq n$.
- [ ] $\mathbf{\Sigma}$ ist immer eine quadratische Matrix.
- [ ] Die Hauptdiagonale von $\mathbf{\Sigma}$ enthält ausschließlich nicht-negative Einträge.
- [ ] $\mathbf{\Sigma}$ ist zu sich selbst invers: $\mathbf{\Sigma} = \mathbf{\Sigma}^{-1}$.
- [ ] Es gilt stets $\mathbf{\Sigma}^\intercal = -\mathbf{\Sigma}$ (d. h. $\mathbf{\Sigma}$ ist antisymmetrisch).

3.4 Ihnen liegt ein 4-dimensionaler Datensatz vor. Sie wurden damit beauftragt, die Dimensionalität dieses Datensatzes **so weit wie möglich** zu reduzieren. Allerdings soll der reduzierte Datensatz **nicht weniger als 80 % der ursprünglichen Varianz** enthalten.

Die Kovarianzmatrix $\boldsymbol{\Sigma}$ und die zugehörige Singulärwertzerlegung $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\intercal$ haben Sie in einem ersten Schritt bereits berechnet. Es ist:

$$\underbrace{\begin{pmatrix} 4 & -2 & 3 & -1 \\ -2 & 4 & -3 & 1 \\ 3 & -3 & 5 & 0 \\ -1 & 1 & 0 & 5 \end{pmatrix}}_{\boldsymbol{\Sigma}} \approx \underbrace{\begin{pmatrix} -0.5 & 0.7 & -0.5 & 0 \\ 0.5 & 0.7 & 0.5 & 0 \\ -0.6 & 0 & 0.7 & -0.3 \\ 0.2 & 0 & -0.2 & -1 \end{pmatrix}}_{\boldsymbol{U}} \underbrace{\begin{pmatrix} 10 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{pmatrix} -0.5 & 0.5 & -0.6 & 0.2 \\ 0.7 & 0.7 & 0 & 0 \\ -0.5 & 0.5 & 0.7 & -0.2 \\ 0 & 0 & -0.3 & -1 \end{pmatrix}}_{\boldsymbol{U}^\intercal}$$

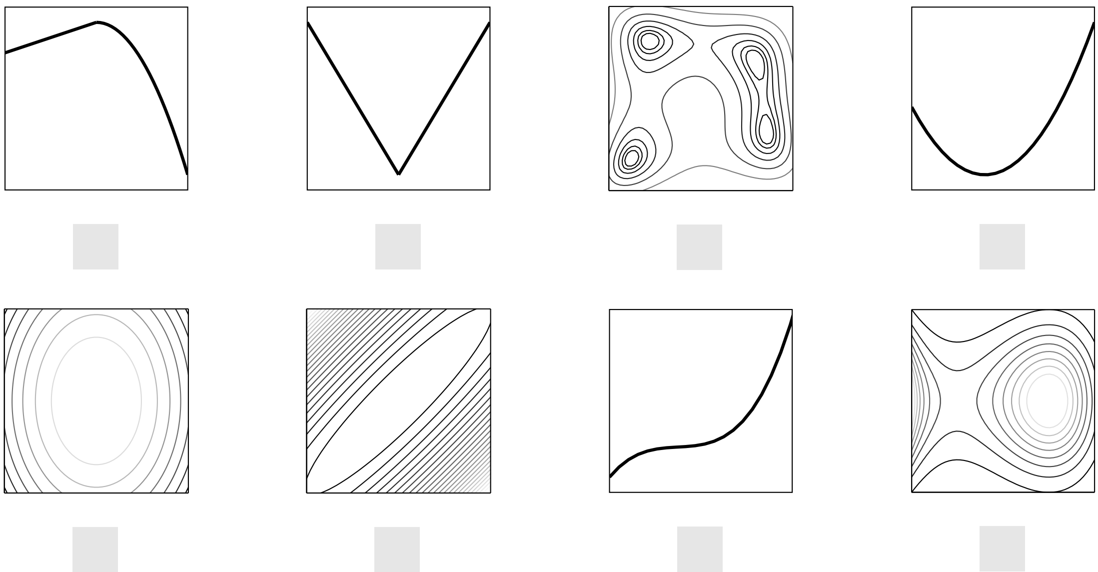
Sei $r$ die Dimensionalität des reduzierten Datensatzes. Bestimmen Sie $r$ so, dass die obige Nebenbedingung erfüllt ist. Welche Vektoren stellen die ersten $r$ Hauptachsen (*principal components*) dar? Begründen Sie Ihre Antworten. **(4 p)**

*Maximal erreichbare Punkte für Aufgabe 3:* **11 Punkte**

# 4 Gemischte Aufgaben

4.1 Bei welchen der folgenden Funktionen handelt es sich um **konvexe Funktionen**? Kreuzen Sie die entsprechenden Funktionsgraphen an.

**Hinweis zu den Contourdiagrammen:** Je größer der Funktionswert ist, desto heller ist die zugehörige Contourlinie. **(3 p)**



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

4.3 Unter welchen Bedingungen konvergiert der *Perceptron*-Lernalgorithmus? **(1 p)**

4.4 Seien $\boldsymbol{u} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \in \mathbb{R}^3$. Berechnen Sie $d_1(\boldsymbol{u}, \boldsymbol{v})$ und $d_2(\boldsymbol{u}, \boldsymbol{v})$. **(2 p)**

4.5 Sie möchten Ziffern automatisiert erkennen (10 mögliche Klassen), und verwenden dazu die logistische Regression. Da es sich um einen binären Klassifikator handelt, benutzen Sie den *One-vs-One* Ansatz.

Wie viele (binäre) Modelle müssen Sie trainieren? **(1 p)**

4.6 Sie evaluieren einen binären Klassifikator auf einem Testdatensatz. Die Tabelle zeigt sowohl die Vorhersagen des Modells, als auch die korrekten Labels. Es ist $\oplus$ die positive Klasse, $\ominus$ die negative Klasse.

Wie hoch ist der *Recall* Ihres Modells? **(2 p)**

| **Beispiel** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Vorhersage** | $\ominus$ | $\oplus$ | $\ominus$ | $\ominus$ | $\oplus$ | $\oplus$ | $\oplus$ | $\ominus$ | $\oplus$ | $\ominus$ |
| **Gold Label** | $\ominus$ | $\oplus$ | $\oplus$ | $\ominus$ | $\oplus$ | $\oplus$ | $\ominus$ | $\oplus$ | $\oplus$ | $\ominus$ |

4.7 Welche Aussagen bezüglich der Aktivierungsfunktionen in tiefen neuronalen Netzen 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.

4.8 Welche Methode können Sie anwenden, um den Hyperparameter $k$ des $k$-Means Algorithmus zu bestimmen? **(1 p)**

*Maximal erreichbare Punkte für Aufgabe 4:* **14 Punkte**

**Zusätzlicher Platz für Anmerkungen:**

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