

# Probeklausur WS 2025/2026

**Freiwillige Abgabe: 25.01.2026, 23:59**

Diese Probeklausur gilt als “Bonusübungsblatt” und kann genutzt werden, um ein vorher verpasstes oder nicht ausreichend bearbeitetes Übungsblatt auszugleichen. Dazu müssen die Lösungen zur Probeklausur bis spätestens 25.01.2026 per CMS eingereicht werden.

Auf die Probeklausur wird es **kein persönliches Feedback** geben. Für **Fragen zur Probeklausur** stehen die Fragestunde zur Klausur im Vorlesungsslot am **27.01.2026** sowie die üblichen Kanäle (CMS Forum, Email) zur Verfügung.

Die Anzahl an Aufgaben der Probeklausur ist im Allgemeinen **höher** als die der Hauptklausur (ca. 6-8 Aufgaben), da ein Überblick über mögliche Aufgabentypen gegeben werden soll. Sie stellt keine Garantie für die Art oder den Umfang der Aufgabenstellungen in der Hauptklausur dar.

**Viel Erfolg!**

Aufgabe	max Punktzahl	erreichte Punktzahl
Aufgabe 1	2	
Aufgabe 2	4	
Aufgabe 3	10	
Aufgabe 4	8	
Aufgabe 5	10	
Aufgabe 6	6	
Aufgabe 7	8	
Aufgabe 8	6	
Aufgabe 9	10	
Aufgabe 10	6	
Gesamtpunktzahl	70	

**Aufgabe 1: Wortsegmentierung (2 Punkte)**

Zerlegen Sie die folgenden deutschen Wörter in ihre morphologischen Bestandteile und benennen Sie diese sowie die ihnen zugrundeliegenden morphologischen Prozesse.

(a) versuchten

(b) Tischbeine

**Aufgabe 2: Reguläre Ausdrücke (4 Punkte)**

(a) Schreiben Sie einen regulären Ausdruck, der das Wort „dog“ sowohl klein- als auch großgeschrieben findet.

(b) Wie kann der Ausdruck verändert werden, damit auch der Plural des Wortes gefunden wird?

(c) Wie kann der Ausdruck weiter verändert werden, damit zusätzlich auch beliebige Variationen der Art „dooooooooog“ gefunden werden könnten?

(d) Wie kann der Ausdruck weiter verändert werden, damit zusätzlich **verhindert** wird, dass „dogmatic“ gematcht wird?

### Aufgabe 3: n-Gramm-Modelle (10 Punkte)

Gegeben sei ein kleines Korpus, bestehend aus den folgenden Sätzen:

$\langle s \rangle$  Er sieht ein Kaninchen auf dem Feld.  $\langle \backslash s \rangle$   
 $\langle s \rangle$  Sabine sieht aus wie ein Model.  $\langle \backslash s \rangle$   
 $\langle s \rangle$  Niemand kennt sich so gut aus wie Michael.  $\langle \backslash s \rangle$   
 $\langle s \rangle$  Zu Weihnachten möchte Lena ein Kaninchen.  $\langle \backslash s \rangle$   
 $\langle s \rangle$  Karl führt sich auf wie ein König.  $\langle \backslash s \rangle$

1. (6 Punkte) Berechnen Sie im Bigram-Modell die Wahrscheinlichkeit des folgenden Satzes:

“ $\langle s \rangle$  Er sieht aus wie ein Kaninchen  $\langle \backslash s \rangle$ ”

2. (4 Punkte) Um Modelle zu bewerten, wird oft die sogenannte Perplexität verwendet. Steht dabei eine hohe Perplexität für ein gutes oder ein schlechtes Modell? Berechnen Sie die Perplexität für den Satz aus Aufgabenteil 1.

#### Aufgabe 4: Kontextfreie Grammatik (8 Punkte)

Gegeben sei eine kontextfreie Grammatik mit den folgenden Produktions- und lexikalischen Regeln:

$S \rightarrow$	NP VP	$V \rightarrow$	find, sah, kam
$VP \rightarrow$	V NP	$N \rightarrow$	Cowboy, Pferd, Gangster, Lasso
$VP \rightarrow$	V	$P \rightarrow$	auf, mit
$VP \rightarrow$	VP PP	$D \rightarrow$	der, dem, den, die, der
$NP \rightarrow$	NP PP		
$PP \rightarrow$	P NP		
$NP \rightarrow$	D N		

(6 Punkte) Leiten Sie den Satz

*Der Cowboy fand den Gangster mit dem Lasso.*

auf zwei syntaktisch unterschiedliche Arten ab, indem Sie die beiden Strukturbäume angeben, die aus der jeweiligen Ableitung resultieren.

1. Lesart:

2. Lesart:

(2 Punkte) Beschreiben Sie, wie sich die beiden Ableitungen in ihrer Bedeutung unterscheiden.

### Aufgabe 5: Naive Bayes (10 Punkte)

Wir betrachten Sätze mit einer NP-V-NP-PP-Struktur, in denen die Präpositionalphrase PP strukturell entweder an das vorausgehende Verb (*high attachment*, 1. Beispiel) oder an das Objekt (*low attachment*, 2. Beispiel) angehängt werden kann.

1. Beispiel: “*Peter baked the cake with friends*”  
→ die PP “with friends” wird an das V “baked” gehängt
2. Beispiel: “*Peter baked the cake with almonds*”  
→ die PP “with almonds” wird an das N “cake” gehängt

In einem kleinen Korpus wurden Beispiele solcher Ambiguitäten als *low*- oder *high-attachment* annotiert. Die folgenden Merkmale des Satzes sollen jetzt als “Features” für den Naive Bayes Klassifizierer genutzt werden: das Verb V, das Objekt des Satzes N1, die Präposition P und das Substantiv N2 der Präpositionalphrase.

Nummer	V	N1	P	N2	Attachment
1	ate	bread	with	cheese	low
2	saw	bread	with	cheese	low
3	saw	documentary	on	cheese	low
4	saw	documentary	on	Saturday	high
5	saw	documentary	with	friends	high

Hinweis: Da die Wörter der verschiedenen Wortklassen nicht beliebig austauschbar sind, müssen für die vier Features V, N1, P und N2 jeweils eigene Vokabulare geführt werden, statt eines Vokabulars, das alle registrierten Wörter enthält.

1. (6 Punkte) Geben Sie die einzelnen Wahrscheinlichkeiten an, die ein Naive-Bayes-Klassifikator verwenden würde, um vorausszusagen, ob im Satz “*Peter ate bread on Saturday*” high oder low attachment vorliegt. Geben Sie dabei insbesondere auch die Formeln an!
2. (2 Punkte) Wie klassifiziert der Naive-Bayes-Klassifikator den Satz?
3. (2 Punkte) Beschreiben Sie das Problem, auf das Sie beim Beantworten der vorherigen Aufgabe stoßen. Wie könnte dieses Problem behoben werden?

### Aufgabe 6: Precision & Recall (6 Punkte)

Die folgende Tabelle zeigt die Evaluation eines Klassifikators, der in einer „smarten Katzenklappe“ eingesetzt wird, nachdem dieser intensiv getestet wurde. Diese Katzenklappen sollen nur die eigenen Katzen ins Haus lassen und für alle anderen Tiere verschlossen bleiben.

	echte eigene Katze	echtes anderes Tier
klass. als eigene Katze	90	20
klass. als anderes Tier	10	380

- (a) (1 Punkt) Der F-Score für die Klasse „anderes Tier“ beträgt ca. 0,96. Bedeutet das, dass der Klassifikator besonders gut ist? Begründen Sie.
- (b) (4 Punkte) Berechnen Sie den F-Score für die Klasse „eigene Katze“.
- (c) (1 Punkt) Halten Sie es bei diesem System für die Klasse „eigene Katze“ für wichtiger auf Präzision oder auf Recall zu optimieren? Begründen Sie.

### Aufgabe 7: Logistische Regression (8 Punkte)

Gegeben seien die folgenden Features eines Klassifizierers, der auf der logistischen Regression basiert:

$$\begin{aligned}x_1 &= \text{email enthält das Wort } \textit{Bitcoin} \\x_2 &= \text{email enthält das Wort } \textit{sofort} \\x_3 &= \text{email enthält Fragezeichen}\end{aligned}$$

Die jeweiligen Gewichte sind aktuell:

$$w_1 = 3, w_2 = 1, w_3 = -0.5, b = -2.$$

Berechnen Sie basierend auf der folgenden Email das Ergebnis des Klassifizierers sowie einen Trainingsschritt:

„Diversifizieren Sie Ihr Portfolio mit Bitcoin! Probieren Sie es sofort selbst aus und starten Sie mit 250 EUR Kapital!“

Die korrekte Klasse dieser Email ist „Spam“ ( $y=1$ ).

1. (1 Punkt) Berechnen Sie  $z$ .
2. (3 Punkte) Wird der Klassifizierer diese Email als Spam erkennen oder nicht?
3. (1 Punkte) Berechnen Sie den Cross-entropy Loss für diese Trainingsemail.
4. (1 Punkte) Geben Sie die Formel an, mit der die neuen Gewichte  $w_1, w_2, w_3, b$  nach der Verarbeitung dieser Email berechnet werden und nennen Sie die Bezeichnungen der einzelnen Variablen.
5. (2 Punkt) Erläutern Sie, was es mit der Schrittgröße  $\eta$  auf sich hat. Welche Auswirkungen kann die Wahl einer zu großen bzw. zu kleinen Schrittgröße haben?

### Aufgabe 8: Distributionelle Semantik (6 Punkte)

Gegeben sind die auf Kookkurrenz basierenden Kontext-Vektoren der folgenden Wörter eines bestimmten Textes „cherry“, „strawberry“ and „digital“.

	Kontextwörter				
	computer	data	result	pie	sugar
<b>cherry</b>	2	8	9	442	25
<b>strawberry</b>	0	0	1	60	19
<b>digital</b>	1670	1683	85	5	4

1. (5 Punkte) Bestimmen Sie anhand der Methode cosine-similarity, welche zwei Wörter sich am meisten bzw. am wenigsten ähneln.
2. (1 Punkt) Diskutieren Sie, ob die cosine-similarity zweier Vektorrepräsentationen negative Werte annehmen kann.



### Aufgabe 9: Semantische Relationen (10 Punkte)

Das Wort „Flügel“ ist mehrdeutig. Geben Sie für zwei der Bedeutungen jeweils eine **Definition** und ein **Synset** an, sowie jeweils ein **Hypernym**, ein **Hyponym** und wahlweise ein **Holonym oder Meronym**. Geben Sie an, ob Sie sich für ein Holonym oder ein Meronym entschieden haben.

*Flügel<sub>1</sub>*

Definition:

Synset:

*Flügel<sub>1</sub>* ist Hypernym von:

*Flügel<sub>1</sub>* ist Hyponym von:

*Flügel<sub>1</sub>* ist Meronym / Holonym von:  
(zutreffendes bitte einkreisen)

*Flügel<sub>2</sub>*

Definition:

Synset:

*Flügel<sub>2</sub>* ist Hypernym von:

*Flügel<sub>2</sub>* ist Hyponym von:

*Flügel<sub>2</sub>* ist Meronym / Holonym von:  
(zutreffendes bitte einkreisen)

### Aufgabe 10: Theoriefragen (6 Punkte)

1. (2Punkte) Nennen Sie wesentliche Unterschiede zwischen Architekturen von Encoder- und Decoder-Sprachmodellen!
2. (2Punkte) Welche Rolle nehmen die sogenannten *keys*, *queries* und *values* in attention-basierten Modellen ein?
3. (2 Punkte) Erläutern Sie den Unterschied zwischen *pretraining* und *in-context learning* von Transformer-basierten Sprachmodellen.