



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

CENTRUM FÜR INFORMATIONS- UND SPRACHVERARBEITUNG
STUDIENGANG COMPUTERLINGUISTIK



KLAUSUR ZUM BACHELORMODUL
„PROBEKLAUSUR VORLESUNG COMPUTERLINGUISTISCHE ANWENDUNGEN“
PROBEKLAUSUR,
DR. BENJAMIN ROTH
KLAUSUR AM

VORNAME:

NACHNAME:

MATRIKELNUMMER:

STUDIENGANG: B.Sc. Computerlinguistik, B.Sc. Informatik, Magister
 anderer:

Die Klausur besteht aus **7 Aufgaben**. Die Punktzahl ist bei jeder Aufgabe angegeben. Die Bearbeitungsdauer beträgt **45 Minuten**. Bitte überprüfen Sie, ob Sie ein vollständiges Exemplar erhalten haben. Tragen Sie die Lösungen in den dafür vorgesehenen Raum im Anschluss an jede Aufgabe ein. Falls der Platz für Ihre Lösung nicht ausreicht, benutzen Sie bitte **nur** die ausgeteilten Zusatzblätter! Verwenden Sie einen dokumentenechten Kugelschreiber oder Füller, **keine** Bleistifte. Es sind **keine Hilfsmittel** zugelassen, außer ein **selbst von Hand beschriebenes DIN A4 - Blatt**. Geben Sie Programmcode immer in **Python** an. **Sie können Fragen auf Englisch bearbeiten**. Bitte tragen Sie **zuerst**, d.h., bevor Sie die Aufgaben lösen, auf **allen** Seiten Ihren Namen ein und füllen Sie die Titelseite aus.

| Aufgabe | mögliche Punkte | erreichte Punkte |
|--|-----------------|------------------|
| 1. Evaluierung von Klassifikatoren | 4 | |
| 2. Perzeptron | 6 | |
| 3. Objektorientierung | 5 | |
| 4. Unsupervised vs Supervised Learning | 3 | |
| 5. NLTK and Lexical Information | 6 | |
| 6. WordNet | 3 | |
| 7. POS Tagging | 3 | |
| Summe | 30 | |
| Note | | |

Einwilligungserklärung (optional)

Hiermit stimme ich einer Veröffentlichung meines Klausurergebnisses in der Veranstaltung „PROBEKLAUSUR Vorlesung Computerlinguistische Anwendungen“ vom unter Verwendung meiner Matrikelnummer im Internet zu.

Datum: _____ Unterschrift: _____

NAME:

Aufgabe 1 **Evaluierung von Klassifikatoren**

Gegeben ein binärer Klassifikator für die Klassen True und False.

- (a) Geben Sie Formel zur Berechnung von Precision, Recall und F1-Measure an (für Klasse True). Erklären Sie alle verwendeten Variablen.

- (b) Geben Sie Formel zur Berechnung der Accuracy an. Erklären Sie alle verwendeten Variablen.

4 PUNKTE

NAME:

Aufgabe 2 Perzeptron

Gegeben eine Trainingsinstanz mit Merkmalsvektor \vec{x} und Label y . Vor dem Trainingsschritt hat der Perzeptron-Klassifikator den Gewichtsvektor $\vec{w}^{(t)}$. Die Werte von \vec{x} , y und $\vec{w}^{(t)}$ sind unten angegeben.

- Welchen Gewichtsvektor $\vec{w}^{(t+1)}$ hat der Klassifikator nach dem Trainingsschritt?
- Was ist die Vorhersage des Klassifikators **vor** und **nach** dem Trainingsschritt?

$$x = \begin{bmatrix} 1 \\ 2.5 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$y = \text{False}$$

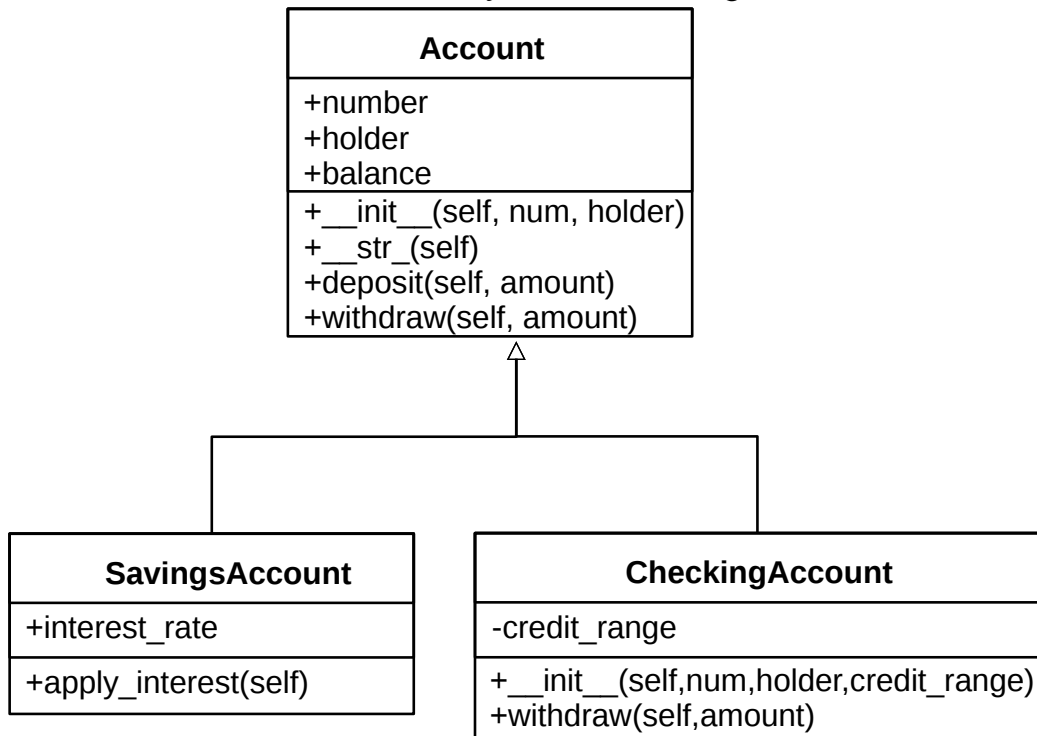
$$w^{(t)} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -2 \\ -0.5 \end{bmatrix}$$

6 PUNKTE

NAME:

Aufgabe 3 Objektorientierung

Gegeben die UML-Klassenhierarchie unten. `annesAccount` ist ein Instanzobjekt von `SavingsAccount`, und `stefansAccount` ist ein Instanzobjekt von `CheckingAccount`.



Geben Sie für die folgenden Aufrufe jeweils an, ob sie definiert sind, und wenn ja, in welcher Klasse die aufgerufene Methode definiert wurde:

- `SavingsAccount(2, "Anne")`
- `CheckingAccount(5, "Stefan", 300)`
- `str(annesAccount)`
- `str(stefansAccount)`
- `annesAccount.deposit(200)`
- `stefansAccount.deposit(500)`
- `annesAccount.withdraw(300)`
- `stefansAccount.withdraw(300)`
- `annesAccount.apply_interest()`
- `stefansAccount.apply_interest()`

5 PUNKTE

NAME:

Aufgabe 4 **Unsupervised vs Supervised Learning**

(a) Was ist der Unterschied zwischen überwachten (supervised) und unüberwachten (unsupervised) maschinellen Lernen?

(b) Ist K-means ein unsupervised oder supervised Algorithmus? Begründen Sie Ihre Antwort.

3 PUNKTE

NAME:

Aufgabe 5 NLTK and Lexical Information

(a) Definieren Sie folgende Begriffe:

(a) Token

(b) Type

(c) Collocation

(d) Bigram

(b) Gegeben folgender Programmcode:

```
1 import nltk
2
3 text = nltk.corpus.genesis.words("english-kjv.txt")
4 bigrams = nltk.bigrams(text)
5 cfd = nltk.ConditionalFreqDist(bigrams)
6
7 print(list(cfd["living"]))
8 >>> ['creature', 'thing', 'soul', '.', 'substance', ',']
9
10 print(list(cfd["living"].values()))
11 >>> [7, 4, 1, 1, 2, 1]
12
13 result = cfd["living"].max()
```

Was wird in der Zeile 13 berechnet?

Was ist der Inhalt der Variable result?

6 PUNKTE

NAME:

Aufgabe 6 WordNet

(a) Erklären Sie kurz die Idee des Lesk-Algorithmus.

(b) Die unten angegebene Tabelle zeigt 2 Bedeutungen von dem Wort "bank". In welcher Bedeutung wird dieses Wort laut Lesk-Algorithmus im Satz "Where do you bank in this town?" benutzt? Begründen Sie Ihre Antwort.

| Sense | Definition |
|------------------------|--|
| Synset('bank.v.03') | do business with a bank or keep an account at a bank |
| Synset('deposit.v.02') | put into a bank account |

3 PUNKTE

NAME:

Aufgabe 7 **POS Tagging**

Gegeben die Hypothese: Ein Satz endet niemals mit einer Präposition (preposition).
Beschreiben Sie, wie Sie diese Hypothese mit Hilfe von NLTK verifizieren können.

3 PUNKTE