Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
SNLP Summer Term 2025

Exercise Sheet 1

Python and probability basics, Zipf's and Mandelbrot's Law

Deadline: 23.04.2025 23:59

Guidelines: You are expected to work in a group of 2-3 students. While submitting the assignments, please make sure to include the following information for all our teammates in each of your PDF files/python scripts:

Name:

Student ID (matriculation number):

Email:

Your submissions should be zipped as **Name1_id1_Name2_id2_Name3_id3.zip** when you have multiple files. For assignments where you are submitting a single file, use the **same naming convention** without creating a zip. For any clarification, please reach out to us on the **CMS Forum**. These instructions are mandatory. If you are not following them, tutors can decide not to correct your exercise.
Please note:

- Ex 1.1 and 1.2 are written assignments, please submit a pdf (written using Latex) with the **names, matriculation IDs and emails** of all team members for this part. In case you are not familiar with Latex, clearly written handwritten submissions are also accepted, but we strongly encourage pdfs written using Latex.
- Ex 1.3 and 1.4 are programming assignments, you can write your code in the supplied notebooks and submit them. Don't forget to put in your **names, matriculation IDs and emails** in the given sections.
- Submit the pdfs and notebooks together in a zip file in CMS. No need to resubmit any datasets.

Exercise 1.1 - Probability Basics (1+1+0.5=2.5 points)

Here are some notations used in this exercise:

- S: Sample Space
- Uni-gram: One token/letter. Eg: a,b,c
- Bi-gram: Two tokens/letters. Eg: $(a, b), (b, c), (c, d)$
- $p(x, y)$: Probability of x followed by y
- $pR(x)$: Probability of x being the right hand bi-gram member. Eg: bi-grams like $(z, x), (a, x), (x, x)$

- $pL(x)$: Probability of x being the left hand bi-gram member. Eg: bi-grams like $(x, y), (x, p), (x, n)$

Let $S = \{a, b, c\}$ and p be the joint distribution on a sequence of two events (i.e. on $S \times S$, ordered). Given the values for:

- $p(a, a) = 0.25$,
- $p(c, c) = 0.25$,
- $p(b, a) = 0.125$,
- $p(b, b) = 0$,
- $p(a, c) = 0.25$,
- $pL(a)$ [unigram probability of a as a left-hand bigram member] = .5,
- $pR(b)$ [unigram probability of b as the right-hand bigram member] = 0.125

Based on the above information, compute:

a) A missing bigram probabilities. Ensure the total probability over $S \times S$ sums to 1. If you cannot compute the probability, explain why? (3-5 sentences)

b) Determine whether any pairs of consecutive events (x, y) are independent (i.e., $p(x, y) = pL(x) \cdot pR(y)$).

c) Is it enough to compute $p(b|c)$ (i.e., the probability of seeing b if we already know that the preceding event generated c)? Justify your answer.

Exercise 1.2 - Zipf's Law (0.5 + 0.5 + 0.5 = 1.5 points)

Please answer the following questions in 2-3 sentences:

a) What is Zipf's Law?

b) Does every kind of language (natural, man-made, programming) follow Zipf's Law?

c) What are the limitations of Zipf's Law?

Exercise 1.3 - Python Basics (0.5 + 0.5 + 0.5 + 0.5 = 2 points)

See attached notebook

Exercise 1.4 - Zipf's and Mandelbrot's Law (1 + 0.5 + 0.5 + 0.5 + 0.5 = 4 points)

See attached notebook