

LLMs from Scratch: Assignment 1A

计科31 刘思远

2025.10.12

Task 1

Problem 1.1 (a)The minimum and maximum amount of bytes is both n .

(b)The minimum and maximum amount of bits is both $3n$.

(c)The byte sequence is

[226, 128, 156, 230, 136, 145, 232, 131, 189, 229, 144, 158, 228, 184, 139, 231, 142, 187, 231, 146, 131, 232, 128, 140, 228, 184, 141, 228, 188, 164, 232, 186, 171, 228, 189, 147, 227, 128, 130, 73, 32, 99, 97, 110, 32, 101, 97, 116, 32, 103, 108, 97, 115, 115, 44, 32, 105, 116, 32, 100, 111, 101, 115, 32, 110, 111, 116, 32, 104, 117, 114, 116, 32, 109, 101, 46, 226, 128, 157]

The length of the byte sequence is 79.

Problem 1.2 (a) The outputs are as follows:

Do you like this course? No. Feel boring?\x08\x08\x08\x08\x08\x08\x08\x08\x08\x08\x

Do you like this course? YES!!!!!!!!!!!!!!

(b) It is the backspace character. It moves the cursor one position backwards, and the next character will override the previous character.

Problem 1.3 It's 1. is a string satisfying the requirement. It will be pre-tokenized as [“It”, “”, “SS”, “1.”], [“It”, “”, “SS”, “1”, “.”], [“It”, “S”, “1”, “.”], respectively, depending on the pre-tokenization method.