Exercises Language Modeling II

Question 1

Scenario: Predicting Tumor Types

A machine learning model is trained to classify patients' tumor types into two categories based on their medical scans:

- **Class A**: Benign (Non-cancerous)
- **Class B**: Malignant (Cancerous)

The model is tested on a dataset of **100 patients**, and the actual diagnoses are compared to the model's predictions. Out of the 100 patients:

- **60 patients** actually have benign tumors, and **40 patients** have malignant tumors.
- The model correctly identifies **50** of the benign cases.
- The model incorrectly classifies **10** benign cases as malignant.
- Out of the patients with malignant tumors, the model correctly identifies **35** cases as malignant.

Generate a Confusion Matrix and compute accuracy, recall, precision, F1 score.

Question 2

A language model assigns the following probabilities to a sequence of words in a sentence:

Sentence: "The cat sat on the mat"

Word	Probability
"The"	0.1
"cat"	0.05
"sat"	0.2
"on"	0.15
"the"	0.1
"mat"	0.05

Compute the **perplexity** of the language model for this sentence.