Raymond Brown

DSP439

Exam 4

May 8, 2025

1. In this assessment, I used lists and dictionaries as structures to store the k-mers, and both structures were relevant throughout the entire script. Lists were used to handle file imports and exports and to store import data as sequences to be passed through the counting logic. Dicts were used both to store and count, due to their structured nature. I implemented a nested dict to count both k-mers and their nearest following characters.
2. Due to the way I handled the counting logic (by the sliding window method and checking if substring length >= k), I did not have to explicitly account for first k-mers as they were already accounted for in the first conditional statement or by the window looping iteration itself. For last k-mers, I implemented a final step (last\_kmer = seq[-k:]) to count just once so there is no risk of a double count.
3. Avoiding overcounting is also supported by the counting logic where each k-mer is initialized explicitly to erase the possibility of them accidentally accounting for the same subsequent/follower. Additionally, each function acting independently of each other leaves little room for such errors.