Project-10

Write a program to read protein sequences file (prj10.fasta) and store the information a format for calculation for sequence features. Calculate following:

- Extract 9-mer overlapping peptide sequence and calculate its Shannon entropy given by equation below:

$$S = -\sum_{j=1,20} \frac{f_j}{n}\left(\log_2\left(\frac{f_j}{n}\right)\right),$$ where f_j is number of amino acids of type j

and *n* is length of peptide/protein.

  M is the number of 9-mer sequences extracted from the fasta file.

- Generate *M* number of 9-mer random peptide sequence using random number generator and calculate its Shannon entropy.

  Algorithm for generating random sequence:

  a. *Make a list of 20 amino acid sequence.*

  b. *Select any one amino acid using random function for n times. (n=9)*

- Plot box plot of Shannon entropy of 9-mer extracted from real protein sequence and for random 9-mer sequence.