

Report - 02 (black_crow)

Implementation Links

K-mer:

<https://colab.research.google.com/drive/1CQ8boEiaRBzZhbXzSsPLKQmSncSRRMhI?usp=sharing>

Recursive Feature Implementation:

https://colab.research.google.com/drive/1txQCivTJYs9vAesQW0szyoylPZ6F01y_?usp=sharing

Description

Anika Tabassum:

K-mer:

K-mer is the substring of any length k in a sequence. This is the all possible length of sequence k . Counting the kmer means to calculate the substring from a string of k length.

For example,

If we have a DNA sequence like TGCCTACTAAGAAACC and $k = 4$, then TGCC is a kmer. Other kmer of this sequence are GCCT, CCTA, CTAC, TACT etc. After finding kmer we need to find how many times a kmer occurs in the targeted sequence. For finding overlap, we need to find the suffix or prefix of a k -mer. Kmer is an efficient way to deal with the huge amount of DNA sequence. But it is very important to select the size of a kmer as we need unique and also effective kmers.

Recursive Feature elimination:

Recursive feature elimination (RFE) is a feature selection method that removes the insignificant feature. RFE is a wrapper-style function choice algorithm that additionally makes use of filter-based function choice internally. RFE only chooses features that are applicable for the prediction. It selects the features with the largest or smallest score. This is an iterative process until the desired number of features are achieved. RFE can be used for feature selection for both classification and regression predictive modeling. Number of features that are going to be selected are unknown. Cross validation is performed with RFE to find the optimum number of features. From different feature subsets, best features with the highest values are selected.

Shayed Ashraf :

K-mer:

k-mer means all the possible substrings of length k that are contained in a string. Generally k-mer is used in Bioinformatics just as a nucleotide sequence of a certain length. For instance a dinucleotide is a kmer where $k=2$.

When we talk about all kmers to talk about all the possible sequences of that length. So for example, when $K=2$ all the possible kmers are: AA AT AC AG TA TT TC TG CA CT CC CG GA GT GC GG

K is usually bigger than 2, so we can talk about all 4mers (256 of them), all 6mers (4096 of them), all 7mers (16,384 of them) etc.

Recursive Feature elimination:

RFE is a popular feature selection algorithm. When we fit the RFE in model it removes the weakest feature (or features) until the specified number of features reached. Features are ranked by the model's attributes, and by recursively eliminating a small number of features per loop, RFE attempts to eliminate dependencies and collinearity that may exist in the model. RFE used in regression and classification model to predict.

Ashak Mahmud:

K-mer:

Usually, K-mer means a sequence's subsequences of length k . And k-mer is normally used in Bioinformatics and related sectors.

For example:

Let $K\text{-mer}=3$

sequence=GGGCAGGGACTTCAGTTTCCGCCACAACAGGTATTCGCATCAAAAAAG

then,

GGG is a k-mer and also GGC, GCA, CAG, AGG etc. are also k-mer examples.

Recursive Feature elimination:

Recursive feature elimination (RFE) is a very well known method which is used in feature selection. It fits a model and removes the weakest feature until it found the specified number of features which it wanted to find. RFE only fetches the feature which is applicable for the prediction.

Nasif Ishtiaque Islam:

K-mer:

Counting K-mer is an essential technique in many bioinformatics methods. It also helps on error corrections of sequence reads. We can understand our dataset better using K-mer. Different combinations of data gives us different insights about the dataset. Space capacity is a matter of concern in K-mer. So, in our dataset we implemented 4-mers on our primary implementation.

Recursive Feature elimination:

Recursive feature elimination is used to remove the weakest feature from the dataset. At first it ranked all the dataset according to their importance and then removed them recursively. On our dataset we'll use recursive feature elimination on our next step.