

Input

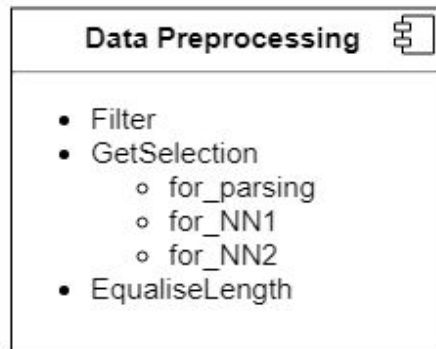
1. data.txt (.fasta etc)

First line -- all the classes presented in data that should be used in classification splitted by ','. Then sequences with metadata splitted by "\n". Each meta begins with '>' or another special symbol that is not presented anywhere else in file

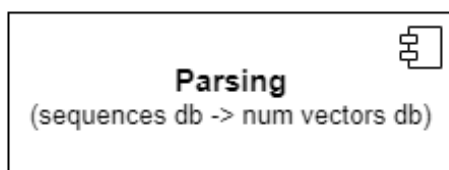
```
class1,class2 ...  
>meta1  
sequence1  
>meta2  
sequence2  
...
```

2. grammar.yrd

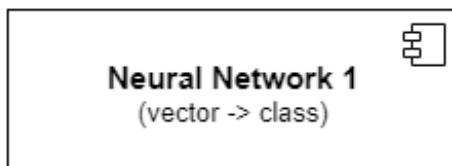
Modules description



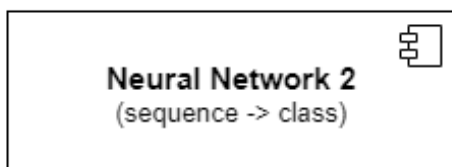
- Filter -- user-defined, optional function, e.g. remove unclassified organisms
- GetSelection -- proportional selection of the existing data samples including the train:validation:test splitting
- EqualiseLength -- cut input sequences or add special symbol til the same length



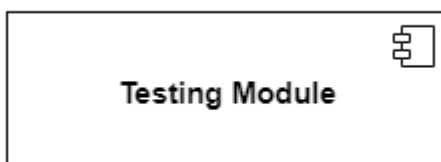
Console utility that converts sequences database to numerical vectors database.
Input parameters: input file, grammar, output file



Establishes a correspondence between the numerical vector and the class of the sequence this vector was obtained from



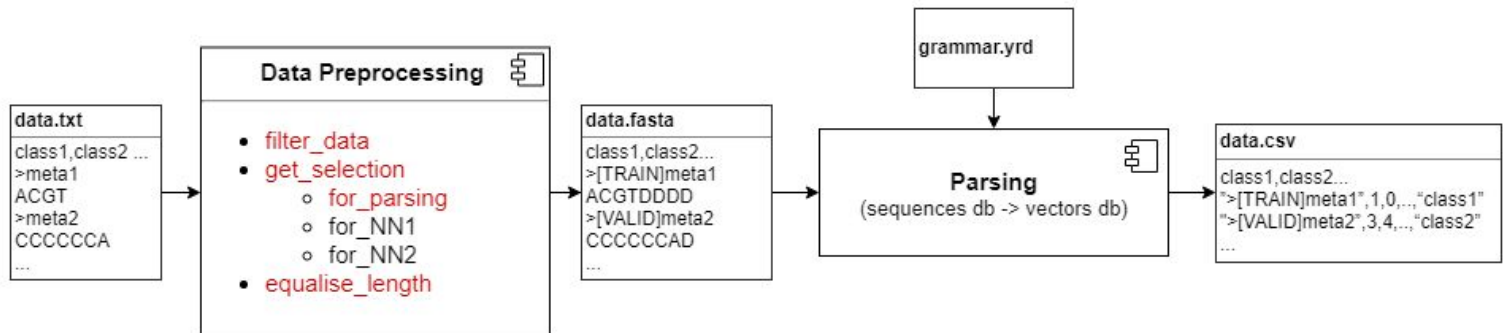
Establishes a correspondence between the sequence and its class using the weights of NN1 in the following way: sequence -> vector -> class



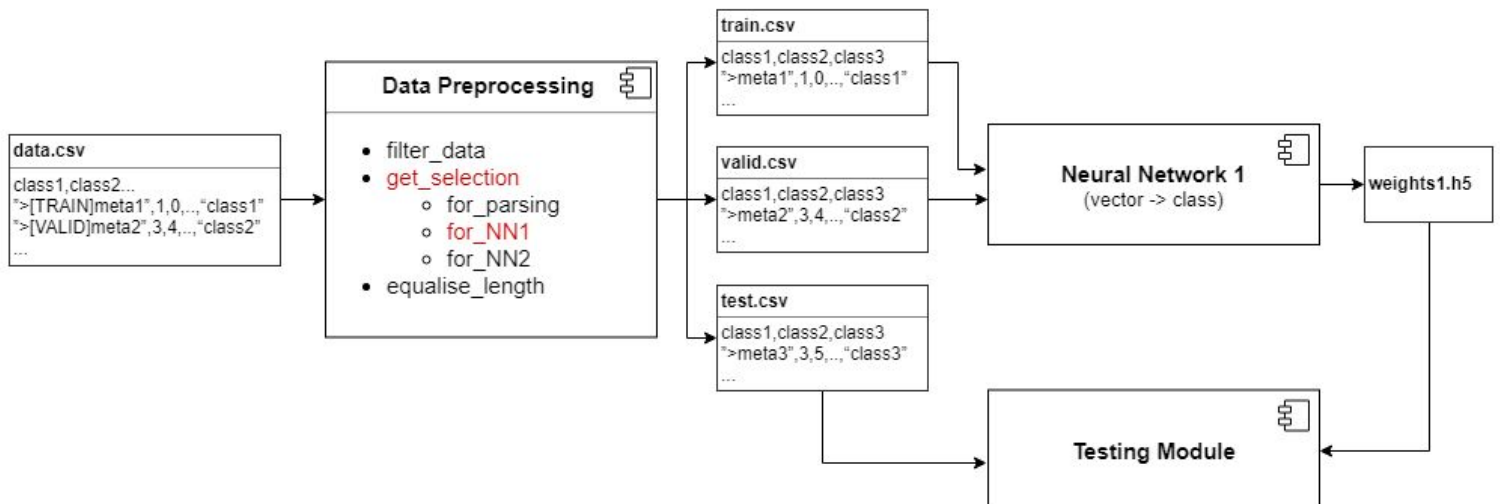
Neural networks test automation...?

Workflow

1. Parsing



2. Neural Network 1 training and testing



3. Neural Network 2 training and testing

