

X4BI040 - Bioinformatique structurale niveau III***X4BI040 - Structural Bioinformatics level III*****Projet/Project****Sessions 2-4**

Le projet que vous réaliserez consistera à classer des motifs structuraux de 5 résidus consécutifs selon la technique de classification non-supervisée utilisant l'algorithme de Kohonen.

The project you will undertake will consist of classifying structural patterns of 5 consecutive residues according to the unsupervised classification technique using the Kohonen algorithm.

Les données de départ vous seront fournies et représentent les angles dièdres (ϕ et ψ) autour des 5 carbones alpha constitutifs d'un grand nombre de pentapeptides censées représenter un bloc protéique tel que défini par de Brevern et al (2000) comme vue en cours. Votre projet consistera à découvrir les éventuels sous catégories de motifs structuraux pour ce bloc protéique.

The starting data will be constructed by you which will represent the dihedral angles (ϕ and ψ) around the 5 alpha-carbons derived from a large number of pentapeptides supposed to represent a protein block as defined by Brevern et al (2000) and as seen in the class of B. Offmann. Your project will consist of discovering any subcategories of structural motifs for this protein block.

Vous mettrez en œuvre l'algorithme de Kohonen selon la méthodologie décrite dans l'article de Weld et al, 1996 pour classer les données.

You will implement the Kohonen algorithm according to the methodology described in the article by Weld et al, 1996 to classify the data.

A cette fin, vous définirez une carte de topologie plane de 5 x 5 "neurones". Dans le cadre d'un plan d'expérience, vous fixerez arbitrairement une valeur de départ de taux d'apprentissage (η) et de rayon d'apprentissage (r) mais vous étudierez d'autres valeurs de départ. Vous étudierez aussi l'impact du nombre de cycle de passage sur la classification.

To this end, you will define a map with a plane topology composed of 5 x 5 "neurons". As part of an experiment plan, you will arbitrarily set a starting value for the learning rate (η) and the learning radius (r) but you will also study other starting values. You will also investigate the impact of the number of learning cycles (c) on the classification.

Enfin, une fois l'apprentissage effectuée, vous fournirez une représentation graphique de la carte ainsi obtenue et vous préciserez les statistiques associées à chacun des classes (celles qui sont les mieux représentées vs celles qui sont plus rares).

Finally, once the learning has taken place, you will provide a graphical representation of the map thus obtained and specify the statistics associated with each class (those that are best represented vs. those that are rarer).

On vous conseille de réaliser ce projet entièrement sous R. Néanmoins, vous pourrez réaliser ce projet à l'aide d'un tout autre langage informatique.

You are advised to execute this project entirely under R. Nevertheless, you can carry out this project with the help of a completely different computer language of your choice.

Vous fournirez un rapport en anglais rédigé à l'aide de LaTeX, ne dépassant pas 6 pages (police de caractère de 11 pts et simples interlignes), comprenant une introduction, un descriptif de la méthodologie et du plan d'expérience, une description des résultats avec illustrations et une discussion.

You will provide a written report in English using LaTeX, not exceeding 6 pages (font size of 11 pts and single spaced), including an introduction, a description of the methodology and the plan of experience, a description of the results with illustrations and a discussion.

You will also submit in a tar.gz file, the code you used and that you have documented.