



MASTER OF SCIENCE  
IN ENGINEERING

Master of Science HES-SO in Engineering  
Av. de Provence 6  
CH-1007 Lausanne

**Hes·so**

Haute Ecole Spécialisée  
de Suisse occidentale

Fachhochschule Westschweiz

University of Applied Sciences and Arts  
Western Switzerland

Master of Science HES-SO in Engineering

Orientation: Information and Communication Technologies

CONFIDENTIAL

## Thesis title

*Projet d'approfondissement effectué dans le groupe Computational Intelligence for Computational Biology de la HEIG-VD, en collaboration avec l'Institut Universitaire de Médecine Sociale et Préventive (Lausanne, Dr. Murielle Bochud) et l'Unité d'Epidémiologie Populationnelle des HUG (Genève, Dr. Idris Guessous).*

**Frédéric Montet**

frederic.montet@master.hes-so.ch





# Contents

<b>1</b>	<b>Méthodes</b>	<b>1</b>
1.0.1	Développement . . . . .	2



# 1 | Méthodes

Pour aborder la recherche de corrélations dans le dataset de stéroïdes, une approche “top-down” a été choisie. Après une étape de nettoyage des données, nous avons obtenu une vue globale sur le dataset grâce à une matrice de corrélation. Ensuite, nous nous sommes concentrés sur les corrélations partielles en utilisant les Self-Organizing Maps (SOM) et leurs composants planaires.

```
1 public class Fibonacci {
2
3     // The golden ratio  $\phi = \frac{1 + \sqrt{5}}{2}$ .
4     public static final double PHI = (1.0 + Math.sqrt(5.0)) / 2.0;
5
6     public static double fibonacci(long n) {
7         if (n < 0) throw new IllegalArgumentException();
8         return Math.floor(Math.pow(PHI, n) / Math.sqrt(5.0) + 0.5);
9     }
10
11 }
```

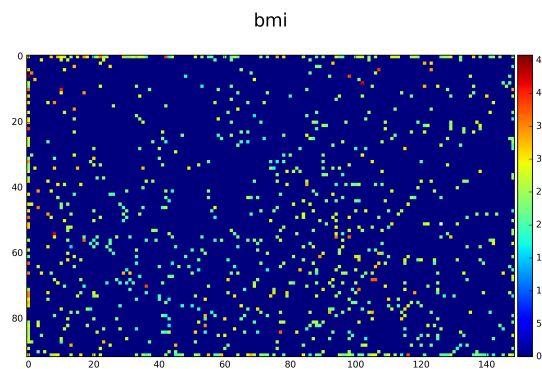


Figure 1.1: SOM colorée non lisible du fait du domaine de l'échelle sans prise en compte de la nature de la donnée à représenter et d'une colormap inadaptée

Pour avoir une vue d'ensemble des corrélations entre stéroïdes, nous avons fait une matrice des corrélations de Pearson. Étant donné la difficulté de lecture visuelle, nous avons mis en évidence les plots significatifs en vert et non significatifs en rouge. Une corrélation est considérée comme significative si sa P-valeur est inférieure à 5% corrigée selon la correction de Bonferroni. [?][?] Soit,

$$P_{max} = \frac{0.05}{(n^2 - n)/2}$$

où :  $P_{max}$  = P-valeur maximum  
 $n$  = nombre de stéroïdes

### 1.0.1 Développement

L'implémentation des outils nécessaires pour la recherche de corrélations partielles a été faite en Python 2.7.11 fourni avec la distribution Anaconda 2.5.0 (x86\_64). Ce choix a été fait car Python est le langage utilisé dans l'institut CI4CB. Plusieurs packages ont été utilisés, notamment le package kohonen qui nous a fourni une base d'implémentation.

Deux outils de visualisation de la carte de Kohonen ont été implémentés :

1. la U-Matrix en réutilisant un code existant
2. les composants planaires de chaque variables

```
1 class SOM:
2     def __init__(self, matrix, variables, params)
3     def computeUMatrix(self, kmap)
4     def constructSamplesForNeurons(self, kmap, matrix)
5     def getUMatrix(self, samples=False)
6     def getComponentPlanes(self)
7     def getErrorPlot(self)
```

Listing 1: Signatures des méthodes de la classe SOM

```
1 params = {
2     'n_iter' : 20,
3     'x' : 20,
4     'y' : 20,
5     'dimension' : len(attributes),
6     'learning_rate' : (0.2, 0.05),
7     'neighborhood_size' : (2./3*20, 1)
8 }
```

Listing 2: Exemple de contenu de la variable *params* nécessaire pour le constructeur d'un objet SOM