

# *Лабораториска Вежба 3*

## Ненадгледувано Учење

### 1. Симнете го податочното множество.

<http://archive.ics.uci.edu/ml/datasets/Mice+Protein+Expression>

Поради тоа што целта на лабораториската е ненадгледувано учење, треба да ги тргнете првиот и последните 4 атрибути, и да работите само со 77-те протеински експресии и сами да пробате да ги најдете кластерите кои ги сугерираат класите.

### 2. Запознавање со податочното множество и претпроцесирање.

Визуелизирајте го податочното множество во 2 или 3 димензии користејќи PCA.

### 3. Кластерирање.

Искористете 2 од методите за кластерирање кои ги имаме учено и видете кои се резултатите кои се добиваат. Дали резултатите се слични?

- a) Gaussian Mixture Models
- b) K-Means
- c) Hierarchical Clustering
- d) DBSCAN

### 4. Визуелизација на резултатите.

Визуелизирајте ги резултатите во 2 или 3 димензии со една боја на секој кластер, повторно користејќи некој од методите како PCA или t-SNE.

На курсот задолжително поставете го `groupb` фајлот со решение.