PDF REPORT

EXO 1-Part 1 : Data distribution and the law of large numbers

Our 2-dimensional variable Z = (X, Y) represent the height `X` and weight `Y` of individuals in a population.

X is a continuous random variable representing height with a normal distribution of mean 170 cm and standard deviation of 10 cm. Y is a continuous random variable representing weight with a normal distribution of mean 70kg and standard deviation of 10kg.

The mean and standard deviation correspond to :

```
mean_x, stdDerivation_x
mean_y, stdDerivation_y
```

We sample n = 1000 points and we also save this sample has a plot in the images folder. File name is `samples_*.png`.

We calculate the empirical average using numpy, plot the result and save it in the images folder. The name of the file is `empirical_average_*.png`.

Sébastien Collard

EXO 2-Part 2 : meteorological data : dimensionality reduction and visualization

In order to implement my dimensionality reduction method, I choose the linear method.

To do the linear method I used the method PCA in the sklearn librairies.

Load the data an label -> use PCA -> plot in 2D and 3D

Which dimension, between 2 and 3, seems to allow to predict the label based on the projected components only ?

The dimensions wich seems to allow to predict the label is the one in which we can best see the difference between the two classes so I think is the 3D version even if we can clearly see the difference between them in the 2D version.

Elliot Khalife

EXO 3-Part 3 : company clustering customers

For this exercise we use the the following clustrenig methods : - K-Means - Agglomerative And for the following heuristic to choose a relevant number of cluster : - Elbow - Silhouette

You will find in the images folder all the plot we generate to choose the most relevant number of cluster for each heuritics and clustering methods.

We determined that for the K-Means methods and the Elbow heuritics the optimal number of cluster is 4, and for the Silhouette heuristic is 6.

We determined that for the Agglomerative methods and the Elbow heuritics the optimal number of cluster is 4, and for the Silhouette heuristic is 6.

We find that the K-Means clustering method and the Silhouette heuristic are a better choice in this situation because K-Means uses both inter- and intra-group distances in its scoring function, whereas the elbow method only uses intra-group distances.

Sébastien Collard

EXO 4-part 4 : exploitation/exploration compromise

If the agent is at the leftmost position, move right.

If the agent is at the rightmost position, move left.

Otherwise, choose the action that leads to the highest known reward and if the rewards are equal for both positions it choose randomly with wright or left.

To implent this policy I tried to maximise the reward, so move towards highest known reward or stay if already at the hightest reward.

Elliot Khalife

Made By :

- Elliot Khalife
- Sébastien Collard