

TP Visualisation de donnée

Visualisation exploratoire en d3.js

1 Point de vue technique

Vous allez développer une application *html/SVG/javascript* utilisant la librairie *d3.js* afin de réaliser une visualisation interactive permettant d'explorer un jeu de données multivarié.

Récupérez dans un premier temps un jeu de donnée adapté (au moins une quinzaine d'attributs) sur le dépôt *machine learning repository* : <https://archive.ics.uci.edu/ml/datasets.php>.

Un ensemble de documentation sur *d3.js* vous est proposé sur madoc (de façon non limitative!).

2 Application à réaliser

2.1 Contexte

L'application que vous allez réaliser nécessite un interpréteur *html/SVG*. La cible choisie sera le navigateur *Chrome/Chromium* de *Google*. *SVG* vous permettra de réaliser relativement facilement des graphiques vectoriels, *javascript* permettra d'automatiser la génération de votre graphique, notamment avec la librairie *d3.js*, enfin l'interface graphique d'interaction pourra être réalisée en *html5*

2.2 Réalisation

Vous devez créer une application comprenant 3 visualisations contenus sur une même page web. Deux seront dédiées à la visualisation globale du jeu de donnée et une à la visualisation détaillée de la distribution d'une dimension.

Les visualisations globales correspondront à une *matrice Scatterplot* et à une visualisation à *coordonnées parallèles*. La visualisation détaillée d'une dimension correspondra à la superposition sur un même graphique de son histogramme et de la courbe de densité à noyau gaussien.

Vous devrez également permettre à l'utilisateur d'interagir avec votre visualisation afin de sélectionner dans une des 3 visualisations des individus qui seront mis en évidence de manière synchronisée (par de la couleur par exemple) dans les autres visualisations.

Matrice Scatterplot La *matrice Scatterplot* doit permettre à l'utilisateur de pouvoir comparer visuellement les dimensions deux à deux. attention à choisir une taille de graphique adaptée au nombre de dimensions de votre jeu de donnée, tout en restant suffisamment grande pour être visualisable.

La sélection des données doit pouvoir être faite à l'intérieur d'un *Scatterplot* et la mise en surbrillance doit concerner tous les *Scatterplots* de manière simultanée.

Coordonnées parallèles Il ne sera pas nécessaire dans la visualisation en coordonnées parallèle de permettre à l'utilisateur de modifier l'ordre des attributs. En revanche l'utilisateur doit pouvoir sélectionner des individus sur un attributs, afin de les mettre en surbrillance dans ce graphique et de manière simultanée dans les autres graphiques.

Histogramme et courbe de densité Il est nécessaire de permettre à l'utilisateur de sélectionner la dimension à visualiser dans ce graphique. Vous devez proposer une méthode simple, potentiellement à partir des autres visualisation, pour effectuer cette tâche.

Les deux visualisations (histogramme et courbe de densité à noyau gaussien) doivent être superposées sur un même graphique. Vous devez permettre à l'utilisateur de modifier la largeur des colonnes de l'histogramme afin de pouvoir choisir interactivement le meilleur compromis de largeur. Et d'une manière symétrique, vous devez permettre à l'utilisateur de modifier la valeur de lissage (largeur de votre noyau gaussien) pour adapter la courbe de densité.

Ces deux paramètres doivent pouvoir être modifiés graphiquement, chacun par un slider (html5 par exemple), en bornant les extrémités à des valeurs raisonnables.