

## Un nouvel apprentissage pour du clustering

Vous devez rendre votre projet (code + document – questions, instructions de déploiement, description de l'algorithme, étude de la complexité et des performances) pour le 3 Janvier 2018 minuit. Les langages dédiés au Web et les applications Web sont interdits. Toute heure de retard sera pénalisée pour -1 point sur la note finale. Tout rendu doit être individuel, tout travail rendu collectif sera sanctionné par 0. Le rendu se fera par mail à lcrepin@ludivinecrepin.fr.

### Principe : distance de Hamming

On vous fournit ces exemples non classifiés :

Exemple	Couleur	Noyaux	Flagelles	Membrane
1	claire	2	2	fine
2	claire	1	2	fine
3	foncée	2	2	fine
4	claire	2	1	fine
5	claire	2	2	épaisse
6	claire	1	1	épaisse
7	foncée	2	2	épaisse
8	foncée	1	1	fine
9	foncée	1	1	épaisse
10	foncée	2	1	épaisse

Afin de les regrouper de manière intelligente, on vous demande de faire un clustering avec la distance de Hamming.

La distance de Hamming entre deux instances  $x$  et  $y$  est donnée par :  $d(x, y) = \text{valeur absolue}(\{a \in A : x(a) \neq y(a)\})$ .

Pour simplifier, la distance de Hamming donne le nombre de différences entre deux éléments. Par exemple, prenons deux chaînes de caractères : "toto" et "titi" : la distance de Hamming est de 2 car deux caractères diffèrent entre les deux chaînes.

### Questions

- Calculer la distance de Hamming pour tous les éléments du tableau d'exemples.
- A partir des distances de Hamming, classer les exemples en deux clusters tels que la distance de Hamming interne (entre éléments d'un même cluster) est toujours plus petite que la distance de Hamming externe (entre éléments de clusters différents).
- Proposer un programme pour diviser un lot d'exemples tel que celui présenté en n cluster.