

26/06/2018

Encadrants :

Jean-Louis Dessalles

Pierre-Alexandre Murena

Giovanni Sileno

Contraste et catégorisation

Etudiants :

Aurélien Blicq

Antoine Bellami

Clément Bonet

Benoît Malézieux

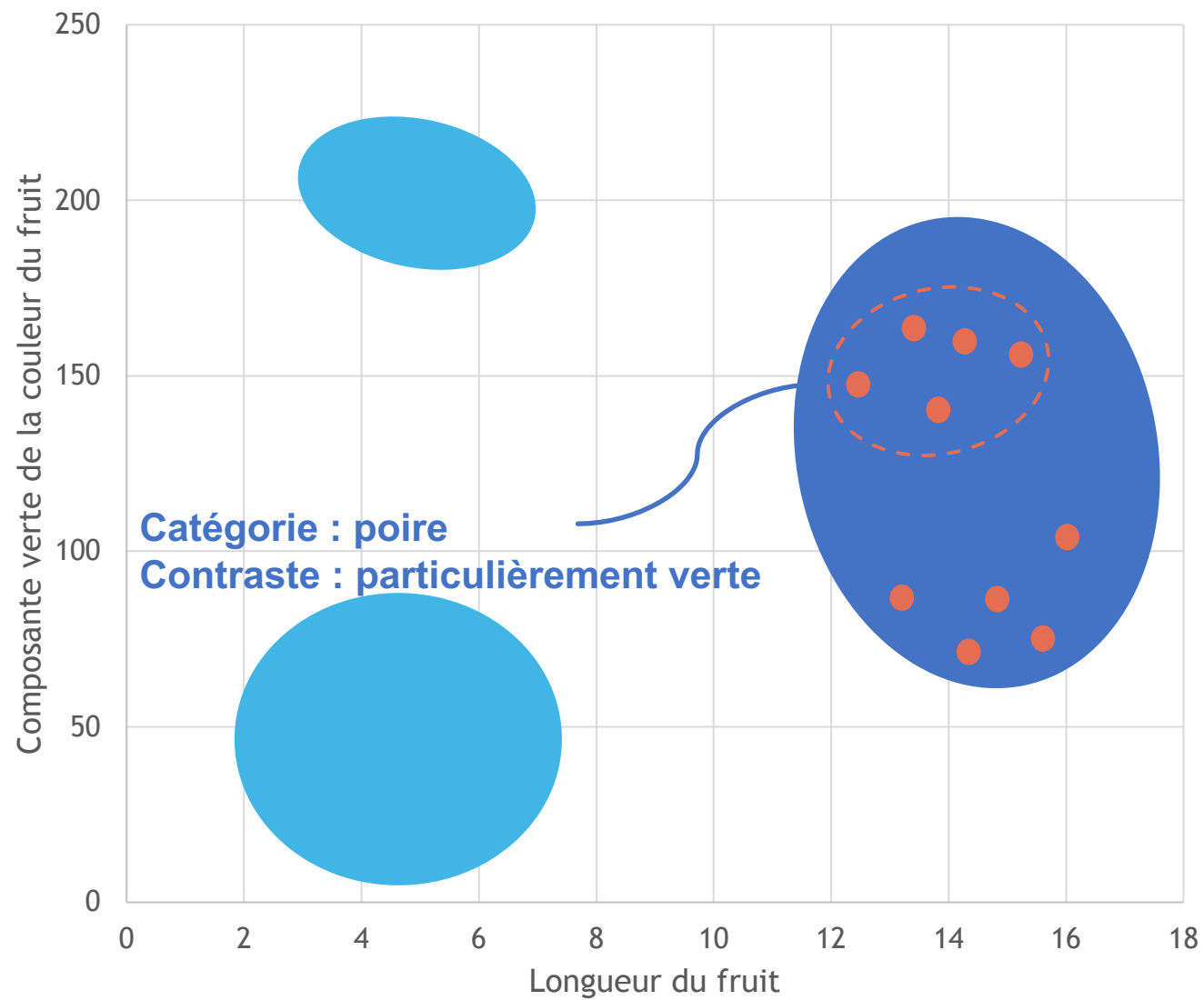
Louis Penet de Monterno

Bastien Vagne

Le problème

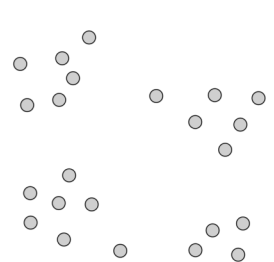
- ▶ Comment différencier deux objets d'une même catégorie mais avec des caractéristiques différentes ?
- ▶ Comment repérer une exception ou une anomalie tout en plaçant l'objet dans la bonne catégorie ?

Exemple

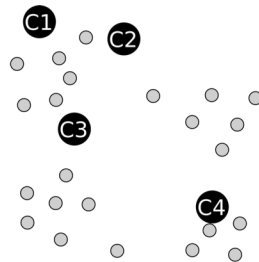


Méthode Offline

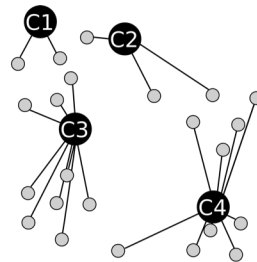
- 1ere étape : catégorisation avec l'algorithme K-Means ou Gaussian Mixture Model



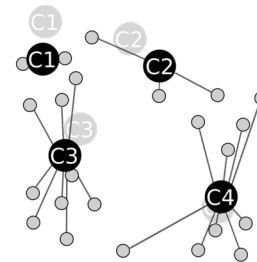
0a. Données d'entrée



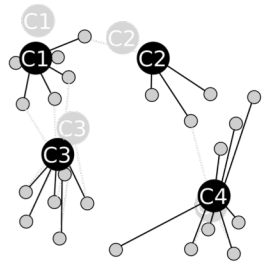
0b. initialisation



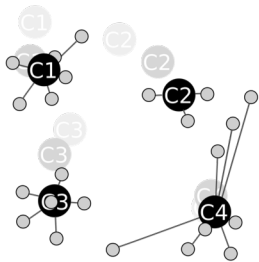
1a. assignation



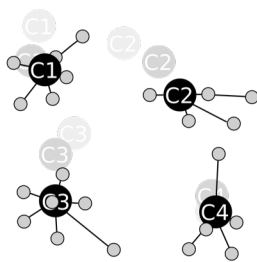
1b. calcul des points moyens



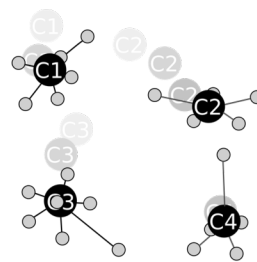
2a. assignation



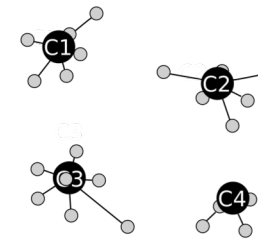
2b. calcul des points moyens



3a. assignation



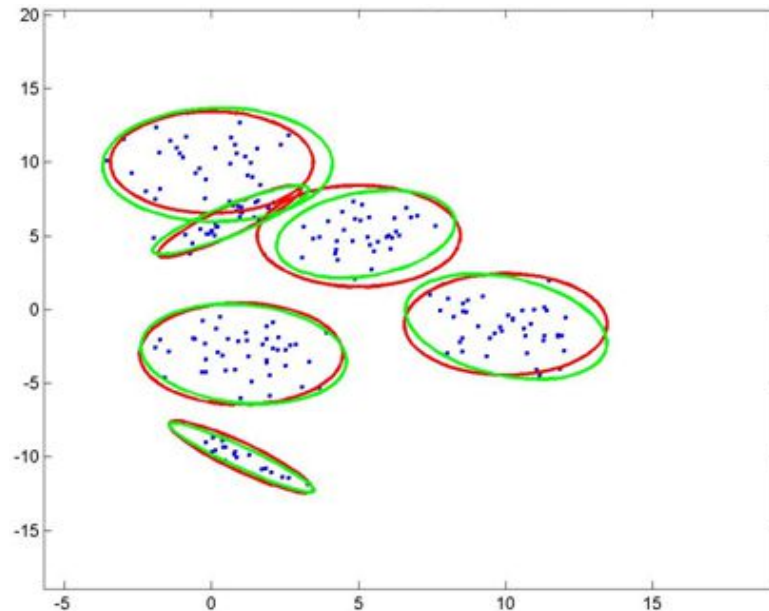
3b. calcul des points moyens



4a. assignation
clusters stables (fin)

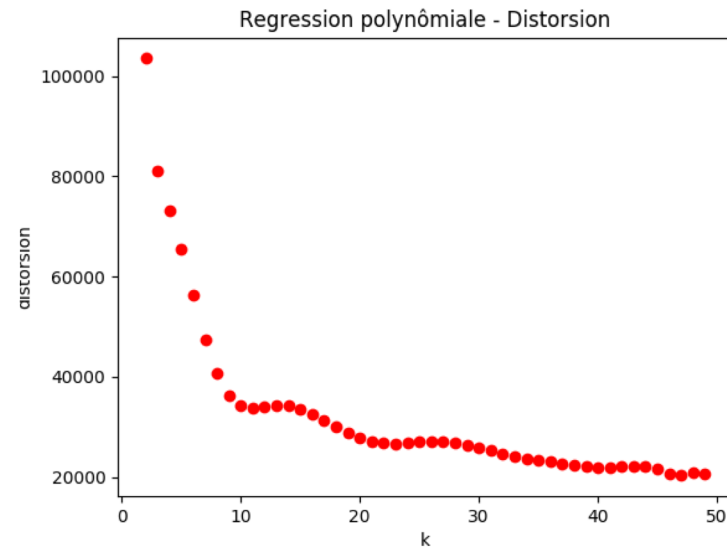
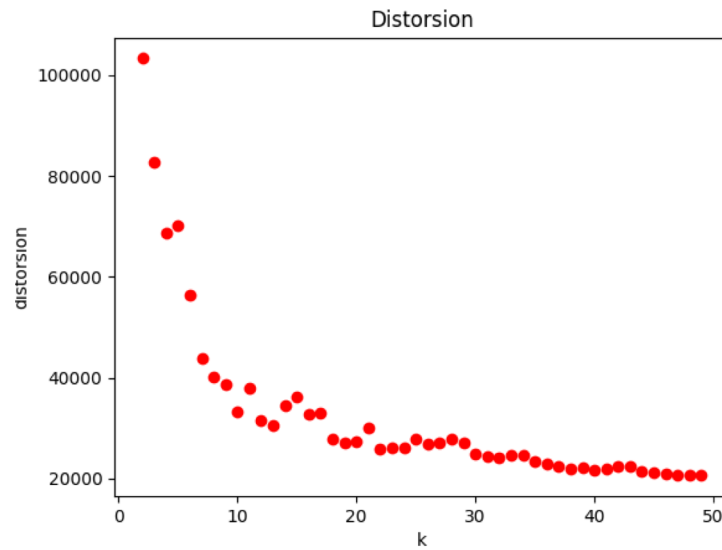
Méthode Offline

- ▶ GMM fait correspondre un ensemble de gaussiennes aux données, tandis que k-means identifie plus simplement le « plus proche » centroïde



Méthode Offline

- Choix du nombre de clusters par la méthode du coude



Méthode Offline

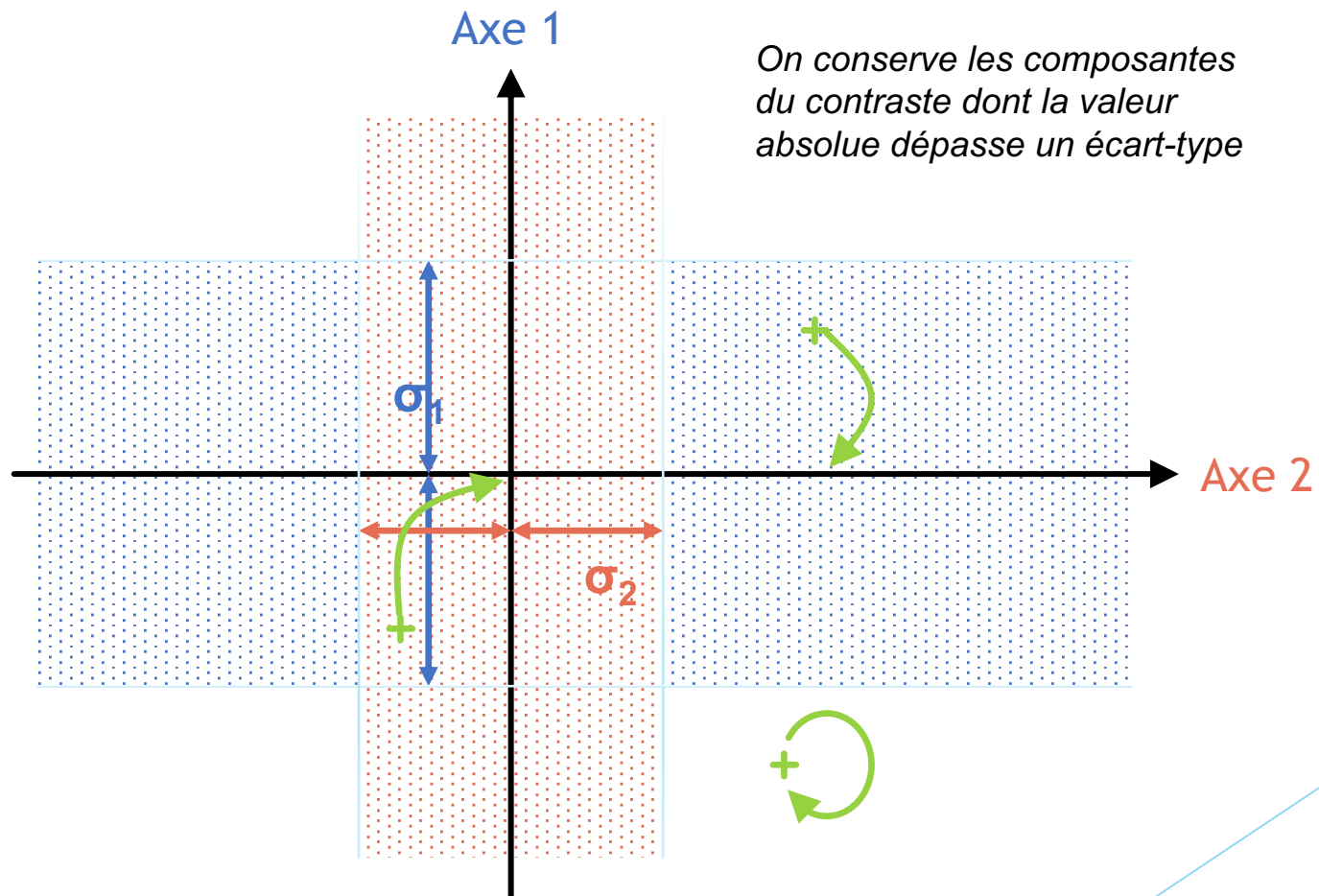
- ▶ 2e étape : calcul du contraste
- ▶ On effectue la différence vectorielle de chaque point avec son centre puis on divise par l'écart-type :

$$\frac{x - \mu}{\sigma}$$

- ▶ Technique du *sharpening* : on ne garde que les valeurs exceptionnelles du contraste, c'est-à-dire les composantes plus grandes qu'une certaine valeur

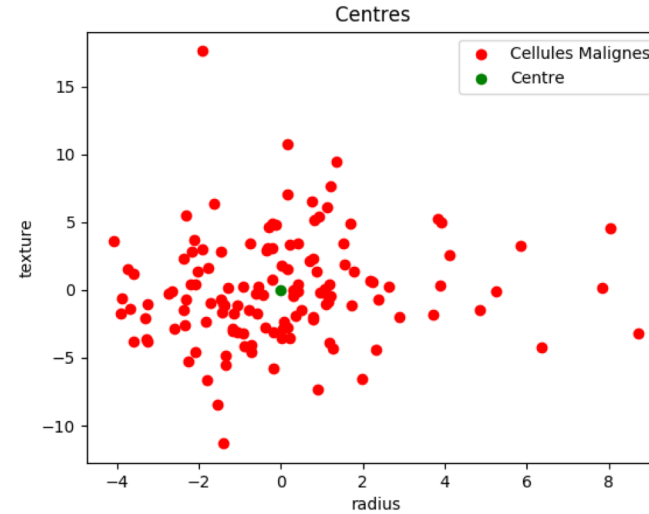
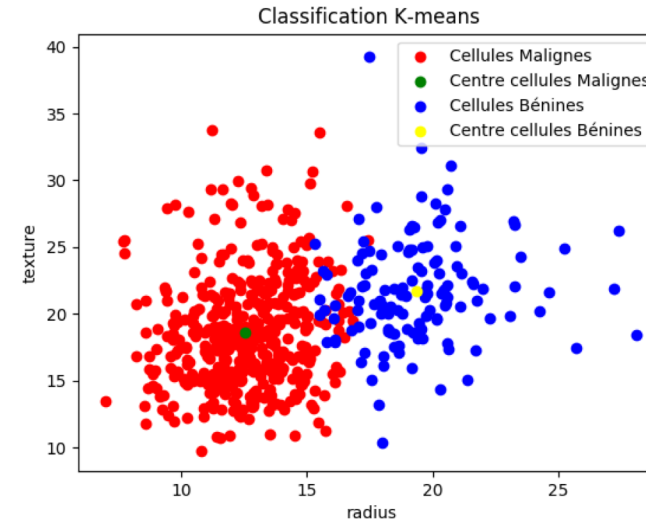
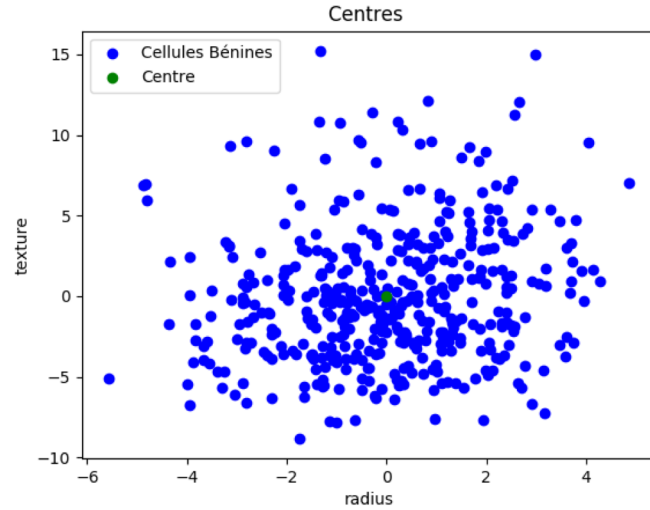
Méthode Offline

► Exemple de sharpening



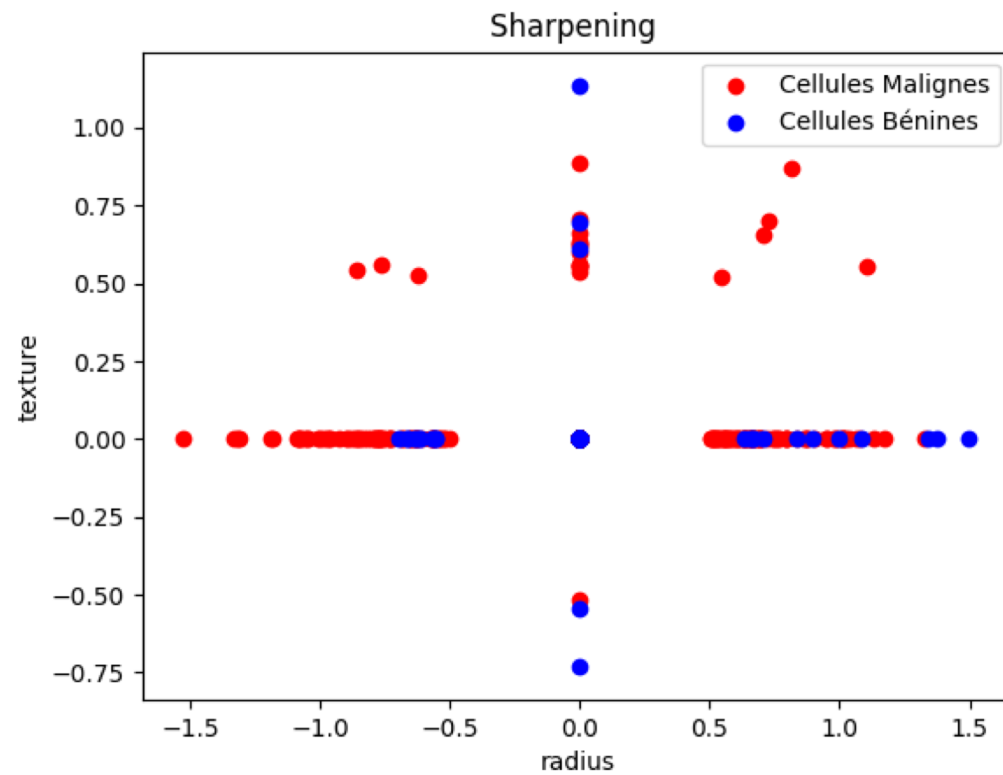
Méthode Offline

► Exemple de données contrastées

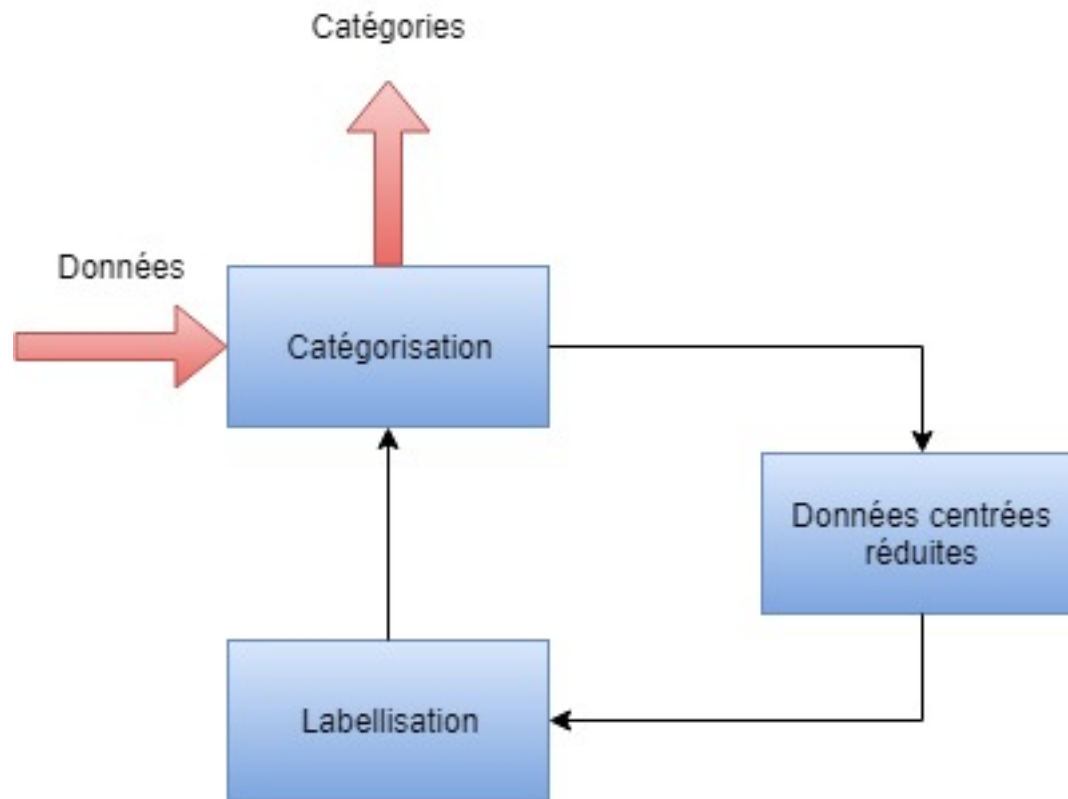


Méthode offline

► Sharpening



Implémentation



Implémentation

- On utilise une base de données artificielle de fruits

	fruit	longueur	largeur	r	v	b	sucré	eau	fibres
0	abricot#petit~	5.992987	4.607582	230.000000	126.000000	48.000000	2.000000	87.000000	1.800000
1	abricot#	7.000000	5.000000	232.098033	117.722015	51.817661	2.000000	87.000000	1.800000
2	abricot#	7.415552	5.144938	235.226165	121.031608	60.406919	5.785021	80.749193	1.749528
3	abricot#long~	7.703690	5.428736	230.000000	126.000000	48.000000	2.000000	87.000000	1.800000
4	abricot#plus rouge~plus vert~	7.000000	5.000000	244.327768	143.660484	54.050655	2.000000	87.000000	1.800000

Résultats

- Ce tableau illustre notre algorithme de description d'un nouvel élément. On associe ainsi à un nouveau fruit une espèce et un adjectif.

Extrait d'un jeu de données de fruits décrits :

Longueur	Contraste longueur	Catégorie du fruit	Catégorie du contraste
5.1	+ 1.2	Fraise	Grand
12.7	- 1.7	Poire	Petit

Entrées
partielles

Sorties

Résultats

► Bilan de la première étape de catégorisation

Catégories	« tomate »	« banane »	« cerise »	« pomme »	« abricot »	« patate »	« prune »	« prune »	« poire »	« prune »
Nombre d'éléments	100	100	201	100	200	100	27	35	100	37
Éléments bien placés	100 %	100 %	49,75 %	100 %	50 %	100 %	100 %	100 %	100 %	100 %
Éléments mal placés			Poivrons : 49,75 % Prunes : 0,5 %		Carottes : 50 %					

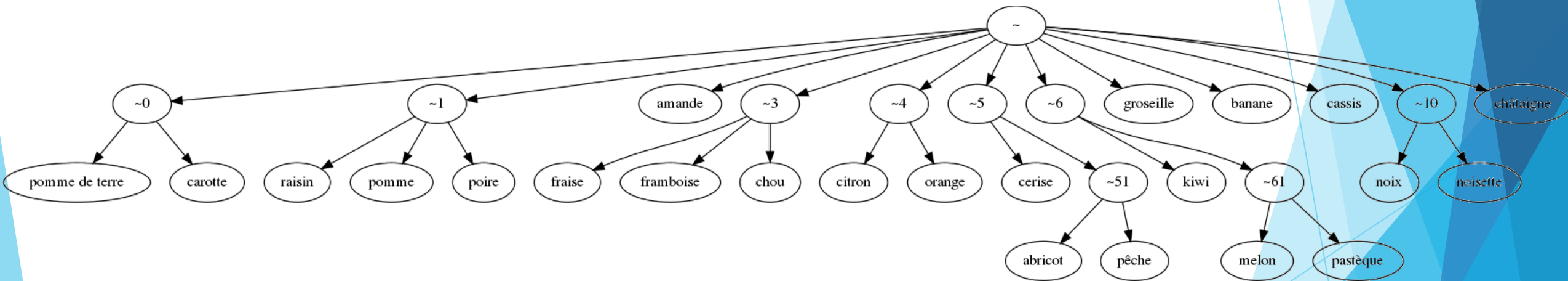
Résultats

Ce tableau synthétise les résultats obtenus par notre algorithme, chaque ligne correspondant à un cluster de contrastes

Label	Centre (en nombre d'écarts-types) (les valeurs non spécifiées sont nulles)	Interprétation	Pourcentage
['r+']	R = 2,35	Rouge	6,12 %
[]		normal	19,45 %
['longueur+', 'largeur+', 'b+', 'sucre-', 'eau+', 'fibres-']	Longueur = 0,89, largeur = 1,36, b = 1,42, sucre = -2,25, eau = 0,71, fibres = -1,83	aucune	2,19 %
[]		normal	21,10 %
['longueur+']	Longueur = 1,28	long	2,18 %
['largeur+']	Largeur = 1,63	large	4,5 %
['longueur+', 'largeur+']	Longueur = 0,93, largeur = 1,06	gros	5,89 %
['v-', 'b+', 'eau+', 'sucre+']	V = -0,63, b = 1,45, eau = 1,05, sucre = 2,02	aucune	1,64 %
['r-', 'b-', 'v-']	R = -0,83, b = -0,76, v = -0,62	Clair	6,2 %
['longueur+', 'eau-', 'sucre+', fibres-']	Longueur = 0,72, eau = -1,23, sucre = 1,42, fibres = -1,06	aucune	5,28 %
['sucre-']	Sucre = -2,31	Peu sucré	8,31 %
['eau-', 'sucre+', fibres-']	Eau = -0,60, sucre = 1,33, fibres = -1,42	aucune	4,53 %
['r-', 'v+', 'b+']	R = -0,62, v = 0,88, b = 1,04	jaune	8,42 %
['longueur+', 'largeur-']	Longueur = 1,58, largeur = -1,38	Long et fin	3,89 %

Méthode incrémentale

- Problématique différente : comment classer un petit jeu de données de manière itérative ?



Perspectives d'utilisation

- ▶ One shot learning
- ▶ Apprentissage sur un faible volume de données
- ▶ Détection d'anomalies