TANAGRA

- Origine, architecture
- Interface
- Accès aux données
- Définir les traitements
- Bibliothèque
- Évaluation et comparaisons
- Expérimentations
- Performances
- Exploration graphique

Université de Lyon 2

http://chirouble.univ-lyon2.fr/~ricco/tanagra/

Culture « machine learning », stat. et stat. exploratoire

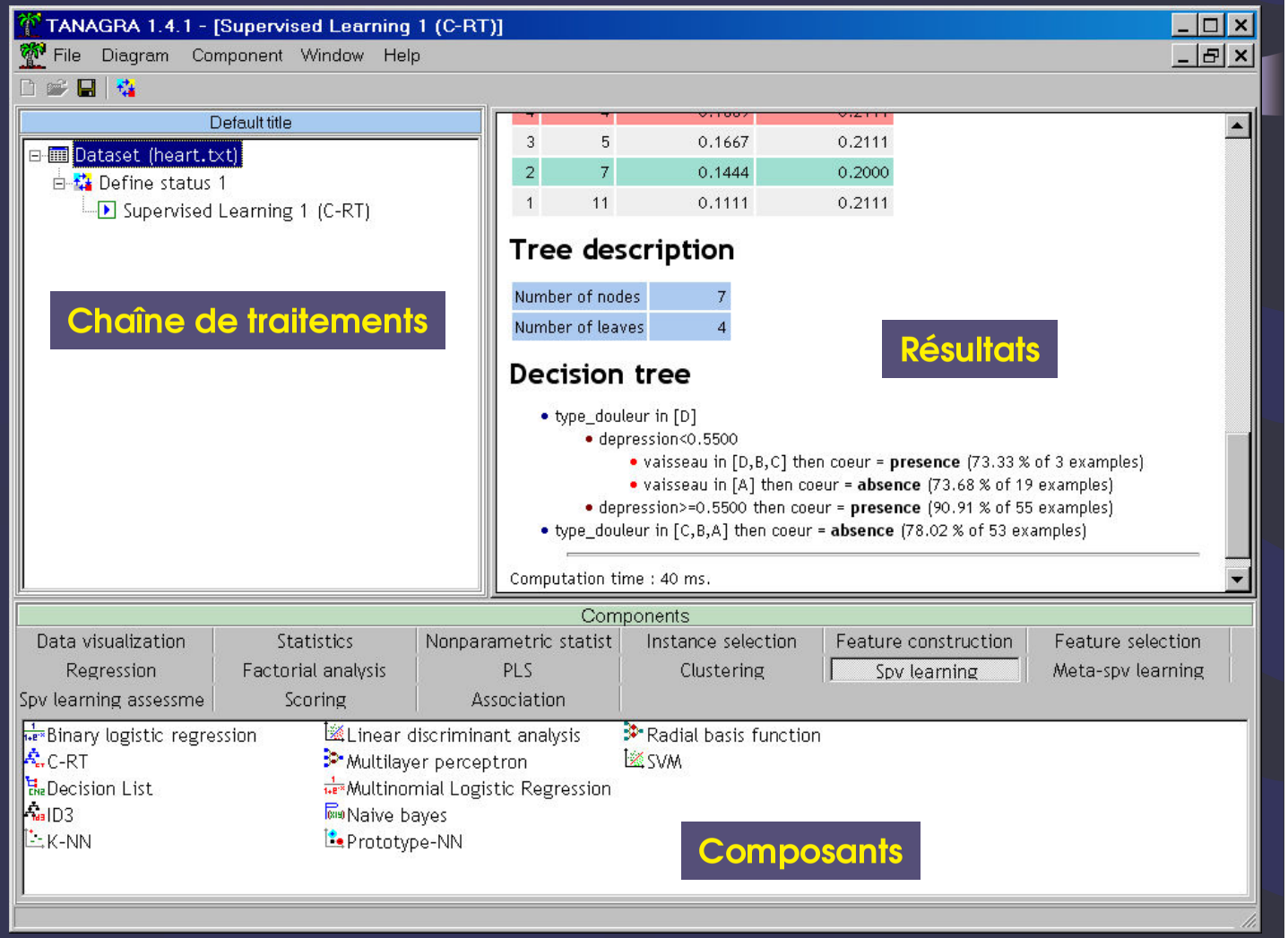
Code source libre en DELPHI (Licence privative – 1.4.1)

Site WEB avec de nombreux didacticiels

Exécution stand-alone sous Windows (Exécutable)

TANAGRA

- Origine, architecture
- Interface
- Accès aux données
- Définir les traitements
- Bibliothèque
- Évaluation et comparaisons
- Expérimentations
- Performances
- Exploration graphique



Transcription de la notion de traitements alternatifs sur les mêmes données
Rôle central du composant « Define Status »

TANAGRA

- Origine, architecture
- Interface
- Accès aux données
- Définir les traitements
- Bibliothèque
- Évaluation et comparaisons
- Expérimentations
- Performances
- Exploration graphique

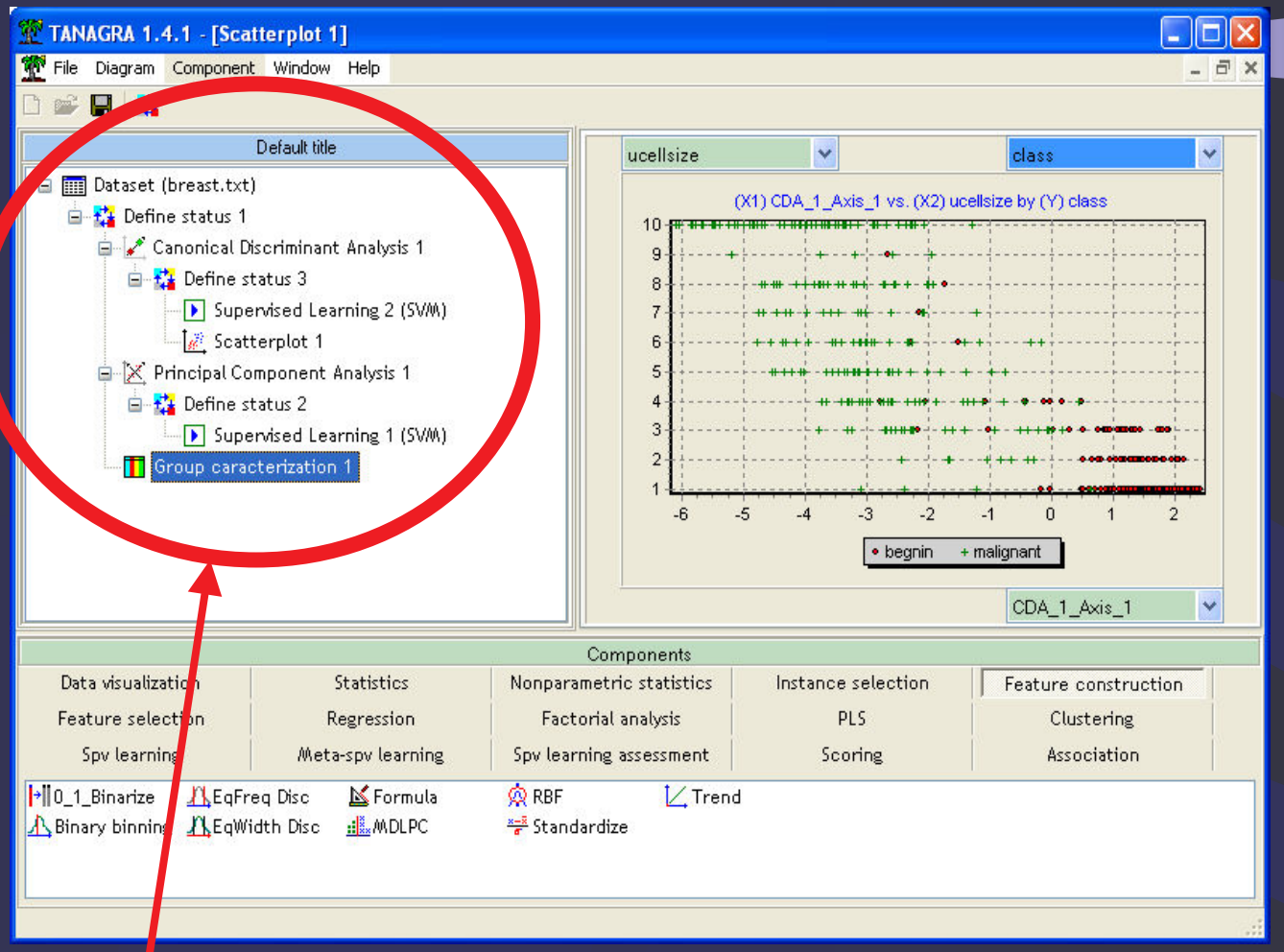
Fichier texte, avec séparateur tabulation (export tableur par exemple)

```
sep_length»	sep_width»	pet_length»	pet_width»	type¶
5.10»	3.50»	1.40»	0.20»	Iris-setosa¶
4.90»	3.00»	1.40»	0.20»	Iris-setosa¶
4.70»	3.20»	1.30»	0.20»	Iris-setosa¶
4.60»	3.10»	1.50»	0.20»	Iris-setosa¶
5.00»	3.60»	1.40»	0.20»	Iris-setosa¶
5.40»	3.90»	1.70»	0.40»	Iris-setosa¶
4.60»	3.40»	1.40»	0.30»	Iris-setosa¶
5.00»	3.40»	1.50»	0.20»	Iris-setosa¶
4.40»	2.90»	1.40»	0.20»	Iris-setosa¶
4.90»	3.10»	1.50»	0.10»	Iris-setosa¶
5.40»	3.70»	1.50»	0.20»	Iris-setosa¶
4.80»	3.40»	1.60»	0.20»	Iris-setosa¶
4.80»	3.00»	1.40»	0.10»	Iris-setosa¶
4.30»	3.00»	1.10»	0.10»	Iris-setosa¶
5.80»	4.00»	1.20»	0.20»	Iris-setosa¶
5.70»	4.40»	1.50»	0.40»	Iris-setosa¶
5.40»	3.90»	1.30»	0.40»	Iris-setosa¶
5.10»	3.50»	1.40»	0.30»	Iris-setosa¶
5.70»	3.80»	1.70»	0.30»	Iris-setosa¶
5.10»	3.80»	1.50»	0.30»	Iris-setosa¶
```

D'autres accès sont disponibles : format WEKA (arff), EXCEL 97&2000

TANAGRA

- Origine, architecture
- Interface
- Accès aux données
- Définir les traitements
- Bibliothèque
- Évaluation et comparaisons
- Expérimentations
- Performances
- Exploration graphique



Plusieurs traitements en parallèle
Combinaison de traitements

TANAGRA

Origine, architecture

Interface

Accès aux données

Définir les traitements

Bibliothèque

Évaluation et comparaisons

Expérimentations

Performances

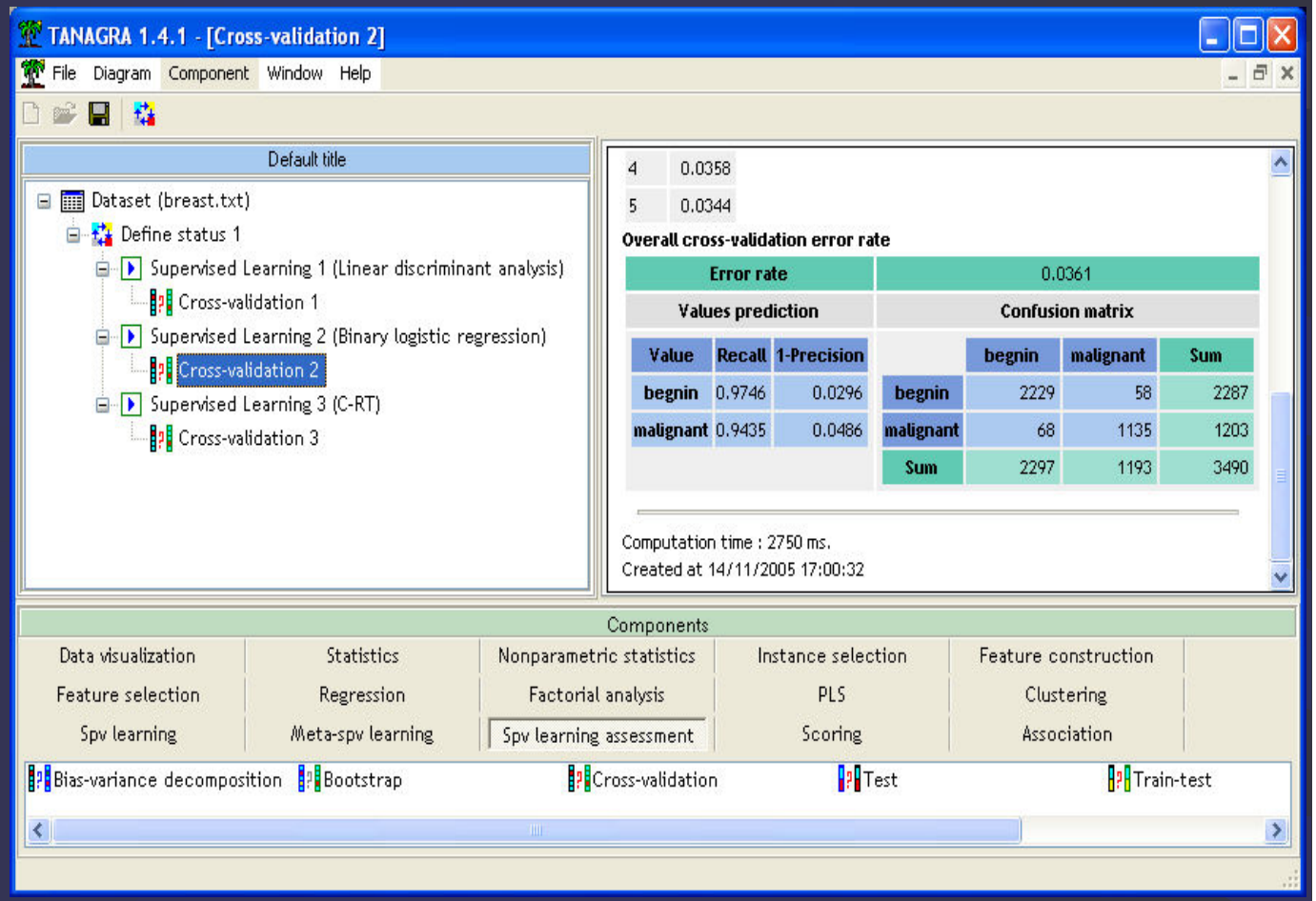
Exploration graphique

- Description rapide des données et stat. descriptives ☺☺
- Traitement des données manquantes ☹
- Restrictions sur les individus (échantillon, condition) ☺☺
- Tests statistiques ☺☺
- Sélection des variables pour le supervisé ☺☺☺
- Construction de variables ☺☺
- Apprentissage supervisé (et régression) ☺☺☺
- Apprentissage non-supervisé ☺☺
- Règles d'association ☺
- Méthodes factorielles ☺☺
- Séries temporelles ☹

Un mix des cultures !

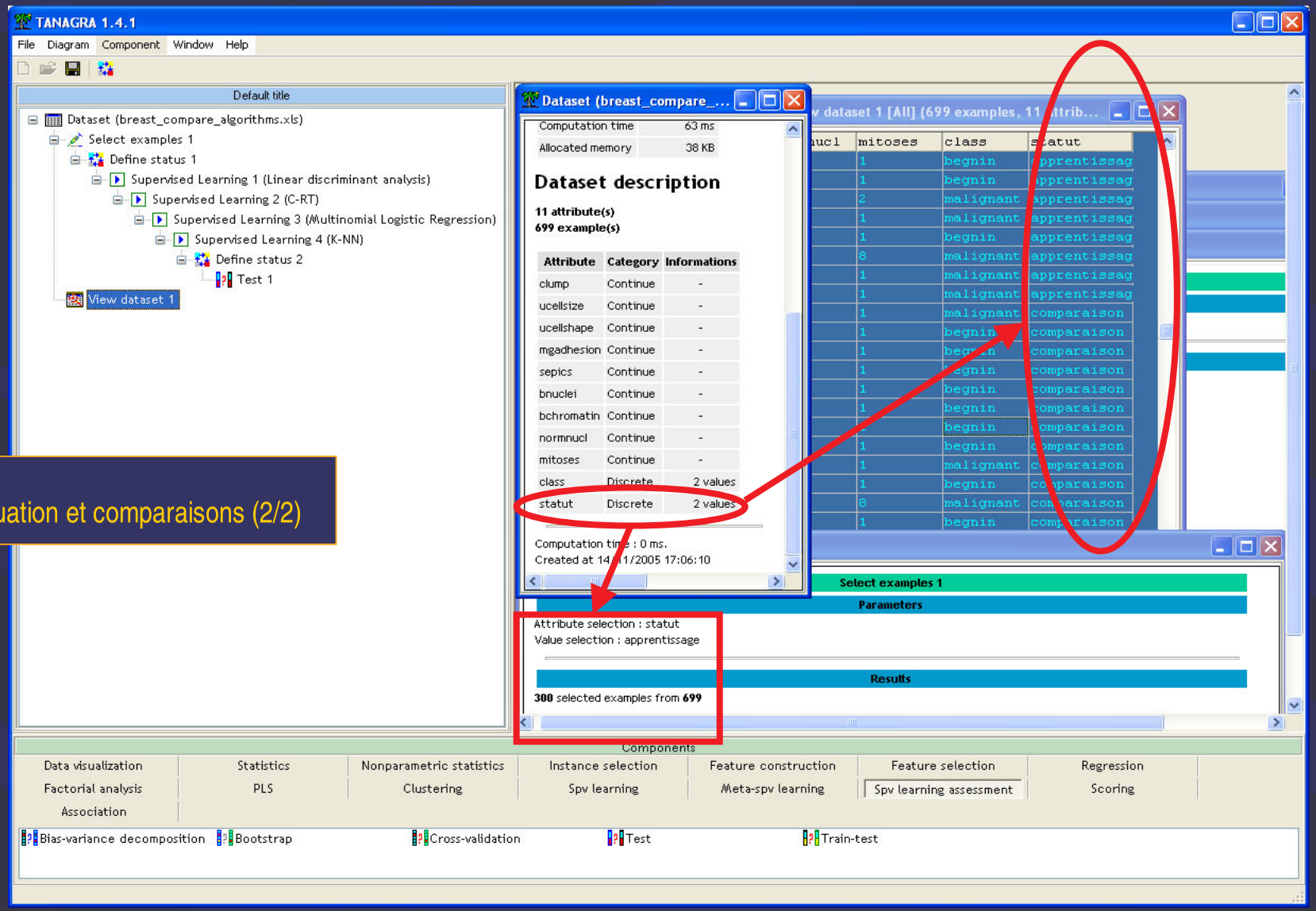
TANAGRA

- Origine, architecture
- Interface
- Accès aux données
- Définir les traitements
- Bibliothèque
- Évaluation et comparaisons (1/2)
- Expérimentations
- Performances
- Exploration graphique



(1)Visualisation séparée
(2) Mais possibilité de regrouper les résultats dans un seul fichier de sortie TXT

TANAGRA



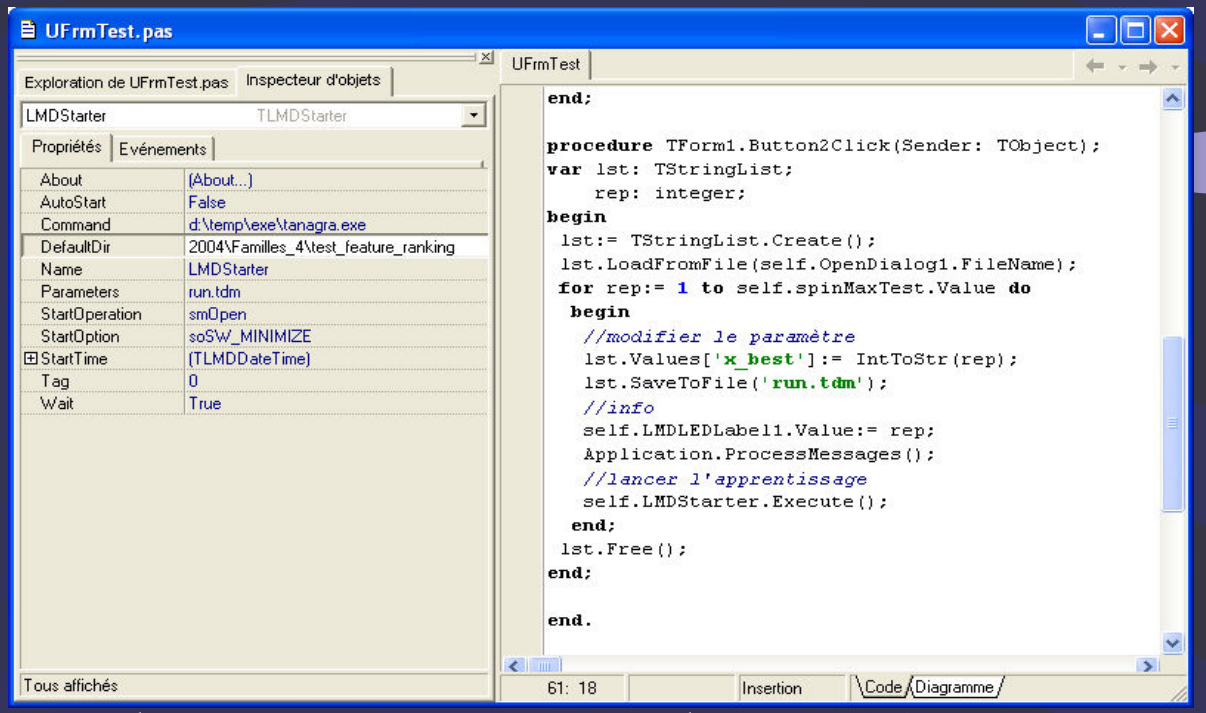
Évaluation et comparaisons (2/2)

Utiliser les mêmes individus en apprentissage, puis comparer sur les mêmes individus en test

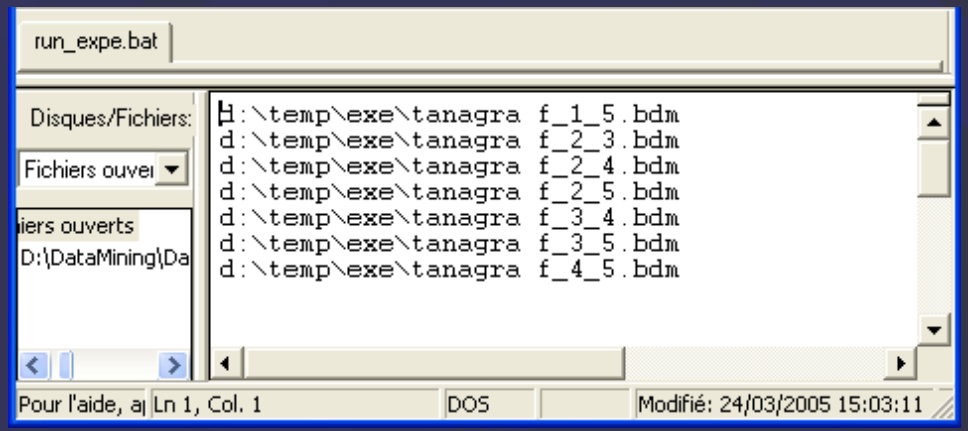
TANAGRA

- Origine, architecture
- Interface
- Accès aux données
- Définir les traitements
- Bibliothèque
- Évaluation et comparaisons
- Expérimentations
- Performances
- Exploration graphique

Exploiter les possibilités de la ligne de commande !



(1) En écrivant un programme qui exécute TANAGRA



(2) En écrivant manuellement le fichier script

TANAGRA

- Origine, architecture
- Interface
- Accès aux données
- Définir les traitements
- Bibliothèque
- Évaluation et comparaisons
- Expérimentations
- Performances
- Exploration graphique

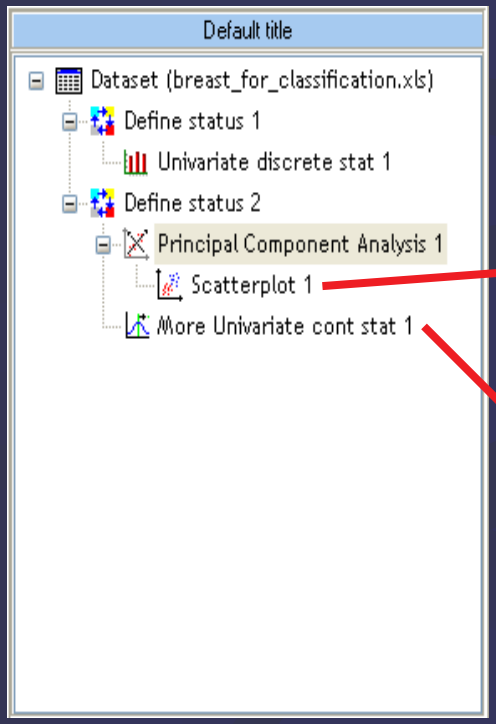
- Pas de limitation théorique (espace adressable) ☺
- Toutes les données en mémoire ☹
- Exécutable binaire sans moteur intermédiaire ☺☺
- Organisation du code → Aude à la simplicité et au minimalisme
  - Rapport code interface / code calcul ☺☺☺
  - Choix optimisés pour le temps de calcul ☺☺

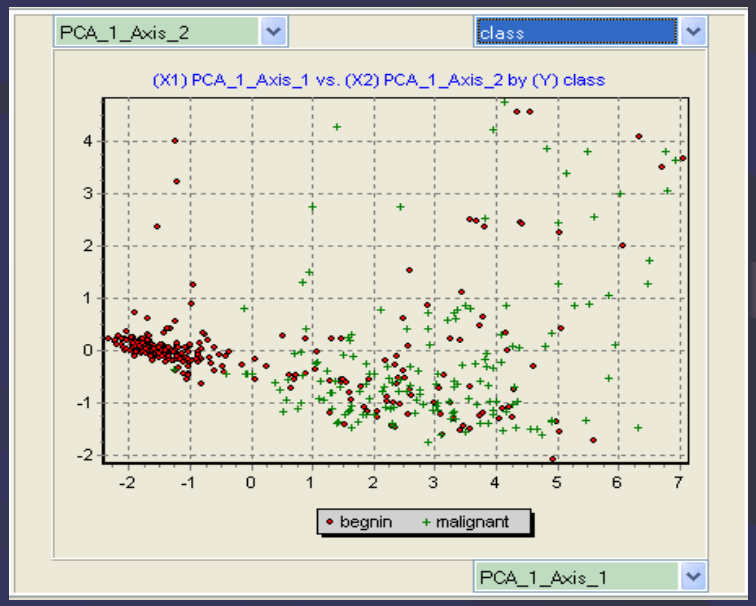
  (importation, arbres de décision, …)
  - Gestion mémoire moins performante ☹

  (règles d'association, calcul matriciel, …)

TANAGRA

- Origine, architecture
- Interface
- Accès aux données
- Définir les traitements
- Bibliothèque
- Évaluation et comparaisons
- Expérimentations
- Performances
- Exploration graphique





Values	Count	Percent	Histogram
x_<_1.9000	353	50.50%	
1.9000_=<_x_<_2.8000	59	8.44%	
2.8000_=<_x_<_3.7000	56	8.01%	
3.7000_=<_x_<_4.6000	44	6.29%	
4.6000_=<_x_<_5.5000	34	4.86%	
5.5000_=<_x_<_6.4000	30	4.29%	
6.4000_=<_x_<_7.3000	30	4.29%	
7.3000_=<_x_<_8.2000	28	4.01%	
8.2000_=<_x_<_9.1000	7	1.00%	
x>=_9.1000	58	8.30%	

Minimaliste et sans interactivité

TANAGRA

Bilan des fréquentations
Le virage statistique en août 2005

Découverte d'une autre communauté avec ses logiciels gratuits « fétiches » :

- R
- DATAPLOT
- OPEN STAT
- VISTA
- WINIDAMS
- etc.

http://members.aol.com/johnp71/javasta2.html
http://freestatistics.altervista.org/stat.php

Bilan de fréquentation de votre site		Le 4/12/2005
Visiteurs uniques		43
Visites		44
Pages vues		51
Nbr de pages vues par visiteurs		1.19
Nbr de pages vues par visites		1.16
Nbr de visites par visiteurs		1.02
Durée moyenne d'une visite		6 ' 00 "
Nbr de visites à une seule page		36
Nbr de pages visitées différentes		4

Estimation pour la journée en cours		J-1	
Visiteurs uniques	51	61	-16.4 %
Visites	52	64	-18.8 %
Pages vues	61	79	-22.8 %

Progression & Bilan annuel

Mois	Visiteurs	Visites	Pages vues
Juillet 2005	906	989	1348
Août 2005	1596 +76.16 %	1781 +80.08 %	2510 +86.20 %
Septembre 2005	1806 +13.16 %	1972 +10.72 %	2621 +4.42 %
Octobre 2005	1988 +10.08 %	2163 +9.69 %	2950 +12.55 %
Novembre 2005	2300 +15.69 %	2495 +15.35 %	3289 +11.49 %