

Entorn de desenvolupament

Prerequisits:

- Mac OS 10.7.3 + JRE 1.6.0.29¹ + XCode 4.3.x + XCode command line tools + MacPorts²

Instal·lar Python i els paquets necessaris de Python través de MacPorts:

```
# xcode-select -switch /Applications/Xcode.app/Contents/Developer/  
# port selfupdate; port upgrade outdated; port uninstall inactive  
# port install python27  
# port select --set python python27  
# port install py27-htmldocs  
# port install py27-numpy  
# port install py27-scipy  
# port install py27-matplotlib  
# port install py27-scikits-learn
```

Instal·lar l'IDE NetBeans 6.9.1³, a continuació instal·lar des de l'IDE el complement “Python” (Eines > Complements > Complements disponibles > Python).

Afegir una nova plataforma de Python que etiqueti a la de MacPorts. Dins de l'IDE anar a “Eines > Python Platforms” i crear una nova plataforma on el binari es troba en /opt/local/bin/python, establir la nova plataforma com la predeterminada.

NOTA:

En obrir el projecte python de la PAC, en les propietats del projecte establir la plataforma python configurada en el pas previ (Category: Python > Python Platform : [etiqueta plataforma python]).

Exercici 1

#	Nom de l'atribut	Tipus d'atribut	?
1	survival	Numèric	2
2	still-alive	Booleà (utilitzat com classe)	1
3	age-at-heart-attack	Numèric	5
4	pericardial-effusion	Booleà (tractat com numèric, probabilitat que sigui cert)	1
5	fractional-shortening	Numèric	8
6	epss	Numèric	15
7	lvdd	Numèric	11

1 JRE for Lion: http://support.apple.com/es_ES/downloads/#macos

2 MacPorts for Lion: <http://www.macports.org/>

3 IDE NetBeans 6.9.1: http://netbeans.org/downloads/6.9.1/index.html?pagelang=ca_ES

8	wall-motion-score	Ordinal: utilitzar wall-motion-index	4
9	wall-motion-index	Numèric	1
10	mult	(variable derivada)	4
11	name	Nominal	0
12	group	Nominal	22
13	alive-at-1	Booleà (derivat atribut 1 i 2)	58

Les variables de color gris es descarten per no aportar informació o bé perquè aquesta està duplicada.

La informació dels pacients s'ha codificat amb python mitjançant classes. La classe ***Patients*** conté una llista d'instàncies de la classe ***Patient***. La classe ***Patient*** conté la informació de cada pacient. Els atributs de la classe ***Patient*** són instàncies de les següents classes:

- classe ***AgeAtHeartAttack***
- classe ***AliveAtOne***
- classe ***Epss***
- classe ***FractionalShortening***
- classe ***Group***
- classe ***Lvdd***
- classe ***Mult***
- classe ***Name***
- classe ***PericardialEffusion***
- classe ***StillAlive***
- classe ***Survival***
- classe ***WallMotionIndex***
- classe ***WallMotionScore***

L'accés a la informació dels atributs que formen part la classe ***Patient***, es a dir per a cadascuna de les 13 classes anteriors, es realitza mitjançant mètodes ***setter/getter*** i amb control per codi per si conté un valor nul ***hasValue_(atribut)***.

La inicialització de la llista de pacients, es realitza passant-li a la constructora de ***Patients*** el nom del fitxer que conté les dades dels pacients.

Per a les observacions numèriques de cada Pacient que no tenen cap valor ("?"), se'ls assigna la mitjana aritmètica de les observacions que si en tenen.

Resultat de l'execució de l'activitat 1:

Activitat: 1

Pacients:

llegint dades del fitxer...

Descartant els pacients que no han superat el temps mínim d'observació i figuren com a vius...

Nombre de variables booleans inicialment no assignades: 0

Nombre de variables numeriques inicialment no assignades: 38
Tractant variables numeriques sense assignar: assignar mitja aritmetica...
Tractant valor de les variables: ranging...
Nombre de pacients: 131

Exercici 2

Resultat de l'execució de l'activitat 2:
Activitat: 2

Pacients:
Llegint dades del fitxer...
Descartant els pacients que no han superat el temps minim d'observacio i figuren com a vius...
Nombre de variables booleans inicialment no assignades: 0
Nombre de variables numeriques inicialment no assignades: 38
Tractant variables numeriques sense assignar: assignar mitja aritmetica...
Tractant valor de les variables: ranging...
Nombre de pacients: 131

Es requereixen 5 components per obtenir un 95.5112441374 % de variancia
Variables seleccionades: ['Survival', 'AgeAtHeartAttack', 'WallMotionIndex', 'Epss', 'Lvdd']

Utilitzant 6 de les 7 variables numèriques indicades en l'exercici 1, totes menys *pericardial-effusion*, al codi es determina amb *num_variables=6*.

Si s'utilitzen les 7 variables numèriques indicades en l'exercici 1 (*num_variables=7*) es requereixen 6 components, sent la variable *pericardial-effusion* la de menor rellevància de les 6 components:
Activitat: 2

Pacients:
Llegint dades del fitxer...
Descartant els pacients que no han superat el temps minim d'observacio i figuren com a vius...
Nombre de variables booleans inicialment no assignades: 0
Nombre de variables numeriques inicialment no assignades: 38
Tractant variables numeriques sense assignar: assignar mitja aritmetica...
Tractant valor de les variables: ranging...
Nombre de pacients: 131

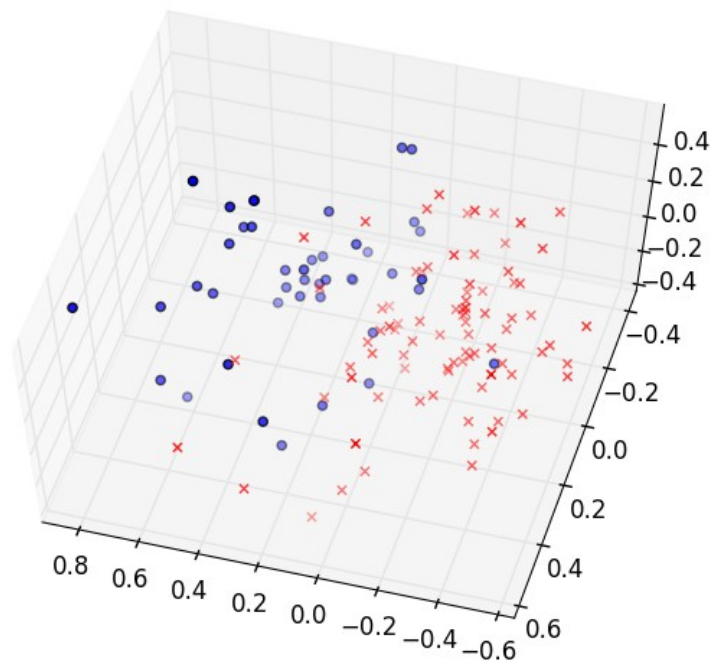
Es requereixen 6 components per obtenir un 97.116793007 % de variancia
Variables seleccionades: ['Survival', 'AgeAtHeartAttack', 'WallMotionIndex', 'FractionalShortening', 'Lvdd', 'PericardialEffusion']

Exercici 3

Resultat de l'execució de l'activitat 3 (la variable *pericardial-effusion* no és de les 3 amb més variança (VAPS principals), dóna el mateix incloure-la o no en la selecció dels VEPS/VAPS):
Activitat: 3

Pacients:
llegint dades del fitxer...
Descartant els pacients que no han superat el temps mínim d'observació i figuren com a vius...
Nombre de variables booleans inicialment no assignades: 0
Nombre de variables numèriques inicialment no assignades: 38
Tractant variables numèriques sense assignar: assignar mitja aritmètica...
Tractant valor de les variables: ranging...
Nombre de pacients: 131

Percentatge de variances amb 3 components 76.7103476529 %
PCA amb 3 components: amb els 3 VEPS amb més VAP...
Separant valors PCA en classes, utilitzant StillAlive com a classe...
Plotejant valors PCA segons la classe...



Conclusió: es pot observar en la gràfica l'aparició de dos clusters més o menys aïllats segons la classe *still-alive*.

Exercici 4

Resultat de l'execució de l'activitat 4 (6 variables):

Activitat: 4

Pacients:

llegint dades del fitxer...

Descartant els pacients que no han superat el temps mínim d'observació i figuren com a vius...

Nombre de variables booleans inicialment no assignades: 0

Nombre de variables numèriques inicialment no assignades: 38

Tractant variables numèriques sense assignar: assignar mitja aritmètica...

Tractant valor de les variables: ranging...

Nombre de pacients: 131

Separant les dades en dos sectors segons la classe...

MDS amb Pearson Score...

20397.8924596

70700.2449554

Plotejant MDS segons la classe...

MDS amb Euclidean Distance...

7778.7629096

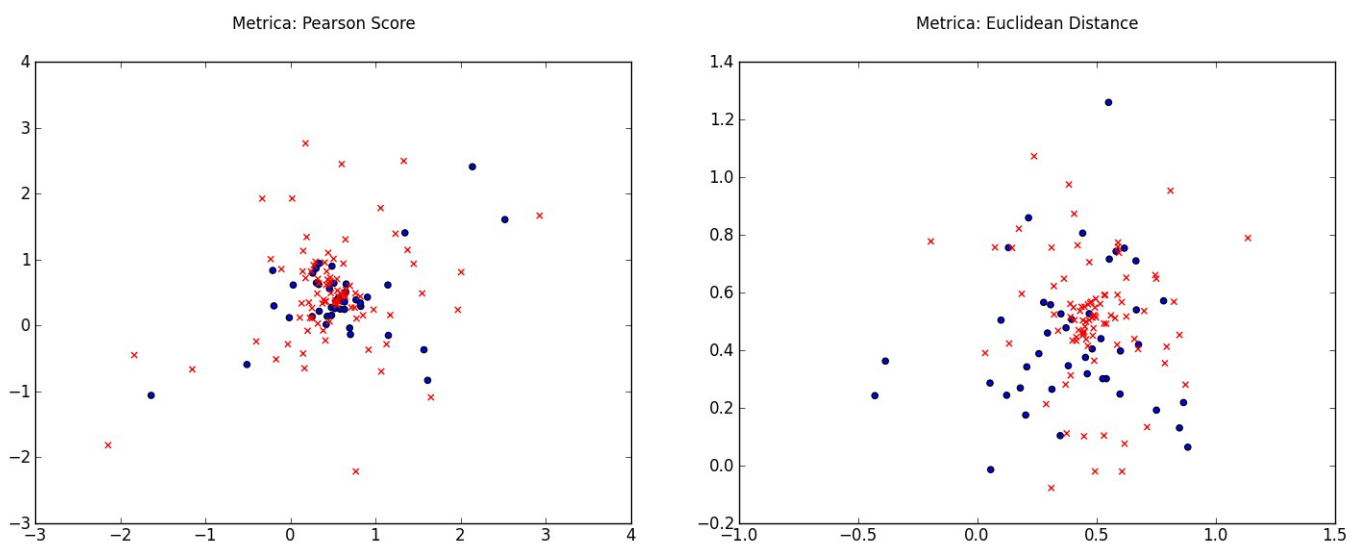
6545.00582185

6179.59888946

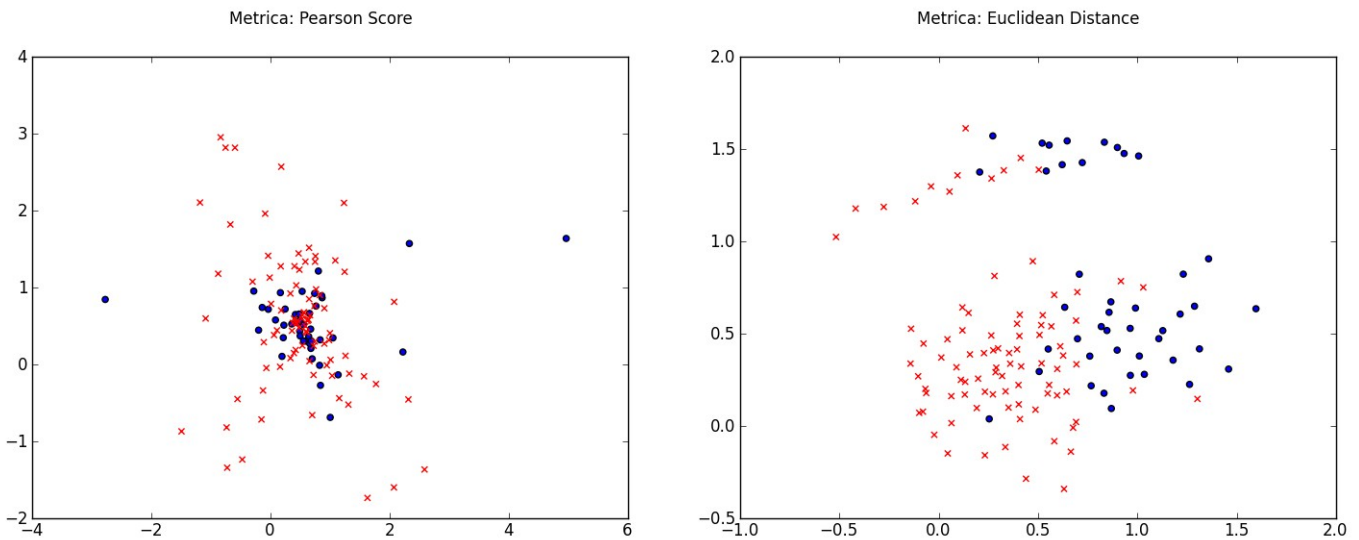
6547.81837913

Plotejant MDS segons la classe...

Utilitzant 6 de les 7 variables indicades en l'exercici 1, totes menys *pericardial-effusion*, al codi es determina amb *num_variables=6*:



Si s'utilitzen les 7 variables numèriques indicades en l'exercici 1 (*num_variables=7*):



Conclusió: Si s'utilitzen 6 variables numèriques (no incloent *pericardial-effusion*) les observacions podrien tenir alguna correlació interna, i si s'utilitzen les 7 variables numèriques (incloent *pericardial-effusion*) les observacions podrien seguir una distribució amb major dispersió.

Exercici 5

Resultat de l'execució de l'activitat 5:

Activitat: 5

Pacients:

llegint dades del fitxer...

Descartant els pacients que no han superat el temps mínim d'observació i figuren com a vius...

Nombre de variables booleans inicialment no assignades: 0

Nombre de variables numèriques inicialment no assignades: 38

Tractant variables numèriques sense assignar: assignar mitja aritmètica...

Tractant valor de les variables: ranging...

Nombre de pacients: 131

Exercici 1...

Entrenament LDA ex 1...

Puntuació: 0.561622283084

Exercici 2...

Es requereixen 6 components per obtenir un 97.116793007 % de variància

Entrenament LDA ex 2...

Puntuació: 0.559000058272

En eliminar una component en l'exercici 2, s'elimina aquella que té menys variància, no obstant es rep menys puntuació al fer una predicció sobre les dades d'entrenament.