



Ένωση Πληροφορικών Ελλάδας

Το Α.Ι. ήταν πάντα εδώ:
Απλά εργαλεία για σύνθετα προβλήματα

Χάρης Γεωργίου (MSc, PhD)

Ένωση Πληροφορικών Ελλάδας

Στόχοι:

- Πρώτος “καθολικός” φορέας εκπροσώπησης πτυχιούχων Πληροφορικής.
- Αρμόδιος φορέας εκπροσώπησης επαγγελματιών Πληροφορικής.
- Αρμόδιος επιστημονικός “συμβουλευτικός” φορέας για το Δημόσιο.
- Αρωγός της Εθνικής Ψηφιακής Στρατηγικής & Παιδείας της χώρας.

<https://www.epe.org.gr>



Τομείς παρέμβασης

Ποιοι είναι οι κύριοι τομείς παρεμβάσεων της ΕΠΕ;

- ① Εθνική Ψηφιακή Στρατηγική & Οικονομία
- ② Εργασιακά (ΤΠΕ), Δημόσιος & ιδιωτικός τομέας
- ③ Παιδεία (Α', Β', Γ')
- ④ Έρευνα & Τεχνολογία
- ⑤ Έργα & υπηρεσίες ΤΠΕ
- ⑥ Ασφάλεια συστημάτων & δεδομένων
- ⑦ Ανοικτά συστήματα & πρότυπα
- ⑧ Χρήση ΕΛ/ΛΑΚ
- ⑨ Πνευματικά δικαιώματα
- ⑩ Κώδικας Δεοντολογίας (ΤΠΕ)
- ⑪ Κοινωνική μέριμνα (ICT4D)





Harris Georgiou (MSc, PhD) – <https://github.com/xgeorgio/info>

- R&D: Associate post-doc researcher and lecturer with the University Athens (NKUA) and University of Piraeus (UniPi)
- Consultant in Medical Imaging, Machine Learning, Data Analytics, Signal Processing, Process Optimization, Dynamic Systems, Complexity & Emergent A.I., Game Theory
- HRTA member since 2009, LEAR / scientific advisor
- HRTA field operator (USAR, scuba diver)
- Wilderness first aid, paediatric (child/infant)
- Humanitarian aid & disaster relief in Ghana, Lesvos, Piraeus
- Support of unaccomp. minors, teacher in community schools
- Streetwork training, psychological first aid & victim support
- 2+4 books, 170+ scientific papers/articles (and 5 marathons)

Επισκόπηση

- Περιεχόμενα:
 - Τι είναι η Μηχανική Μάθηση και η Αναλυτική Δεδομένων (ML/DA).
 - Προπαρασκευή δεδομένων (pre-processing), είδη προβλημάτων ML/DA.
 - Επεξεργασία αρχείων, command-line tools, SQL, αρχεία κειμένου (imports).
 - Προηγμένα εργαλεία στην κονσόλα (command-line), σε IDE, στο cloud.
- Αναφορές:
 - «Εισαγωγή στη Μηχανική Μάθηση και στην Αναλυτική Δεδομένων», Χ. Γεωργίου, Α' κύκλος ανοικτών μαθημάτων ΕΠΕ – <https://youtu.be/mlU4SvyfRqA>
 - «Εισαγωγή στην Αναλυτική Δεδομένων με τη γλώσσα R», Χ. Γεωργίου, Β' κύκλος ανοικτών μαθημάτων ΕΠΕ – <https://youtu.be/7aaNnXxyzgl>
 - Dunham: Data Mining – Introductory and Advanced Topics. Prentice Hall, 2003.
 - Tan, Steinbach, Kumar: Introduction to Data Mining. Addison Wesley, 2006.

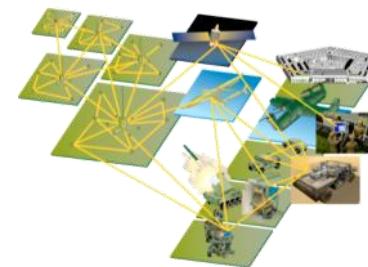
Μέρος I: Οργάνωση της εργασίας (ML/DA)

1. Μηχανική Μάθηση και Αναλυτική Δεδομένων (ML/DA).
2. Προεπεξεργασία / προπαρασκευή δεδομένων.
3. Τι είναι το ML/DA pipeline;
4. Τι προβλήματα έχουμε;



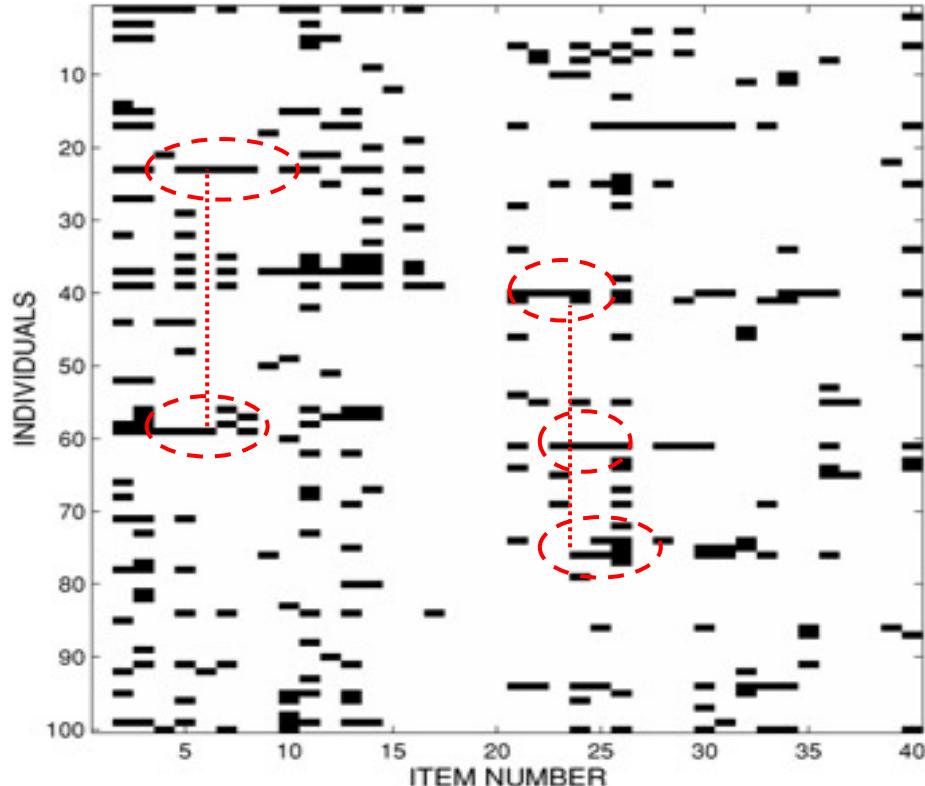
Δεδομένα παντού ...

- Παράγονται όλο και περισσότερα δεδομένα:
 - Τραπεζικά, τηλεπικοινωνιακά, ...
 - Επιστημονικά δεδομένα:
αστρονομικά, βιολογικά κλπ.
 - Κείμενα στο web κ.α.
- Αποθηκεύονται όλο και περισσότερα δεδομένα:
 - Γρήγορη / φθηνή τεχνολογία αποθήκευσης
 - Ικανά ΣΔΒΔ για μεγάλες ΒΔ



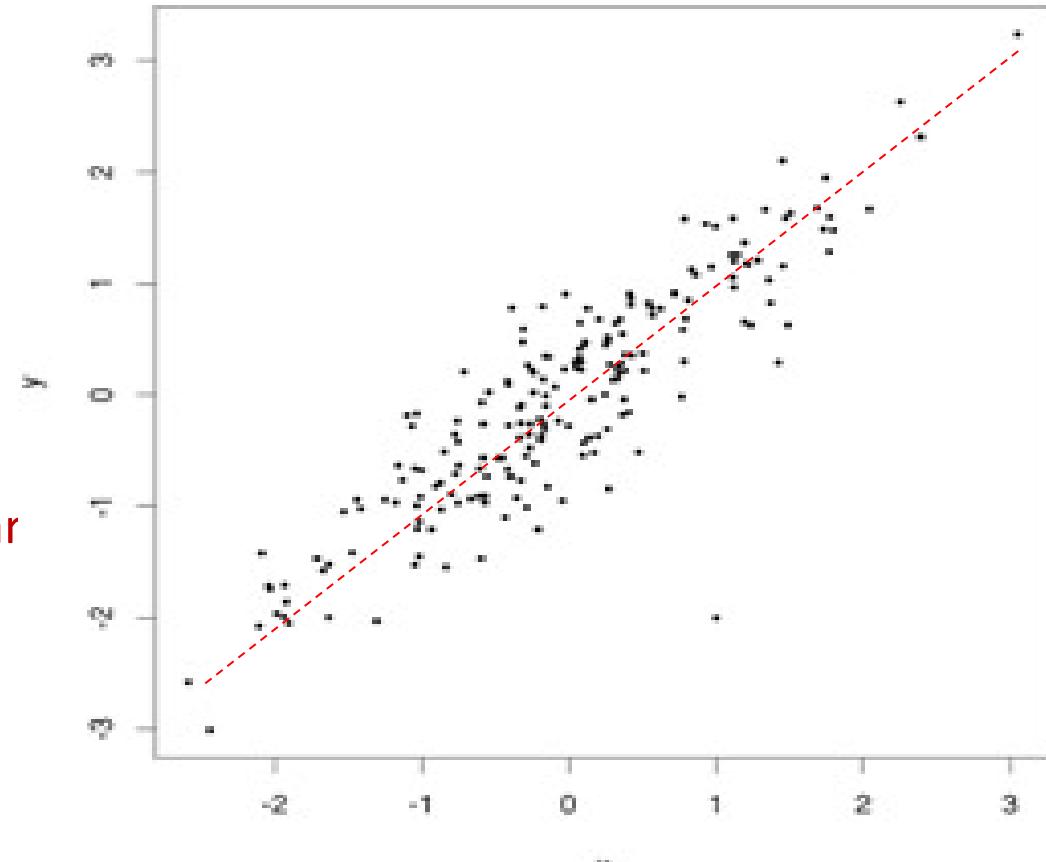
Δεδομένα...

- 2-D πίνακας: πελάτες (γραμμές) προς προϊόντα (στήλες)
- Στόχος ανάλυσης: (εύκολος;) εντοπισμός συλλογικής συμπεριφοράς
- Αποτέλεσμα: Αναγνώριση ισχυρών εξαρτήσεων μεταξύ μεταβλητών
- **Στον πίνακα: «συνεμφάνιση» τιμών ή γεγονότων**

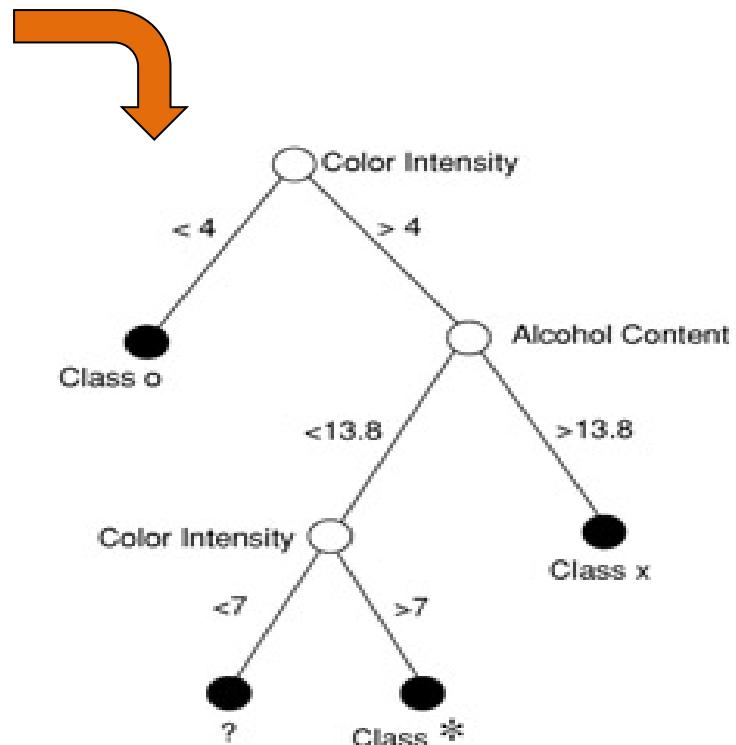
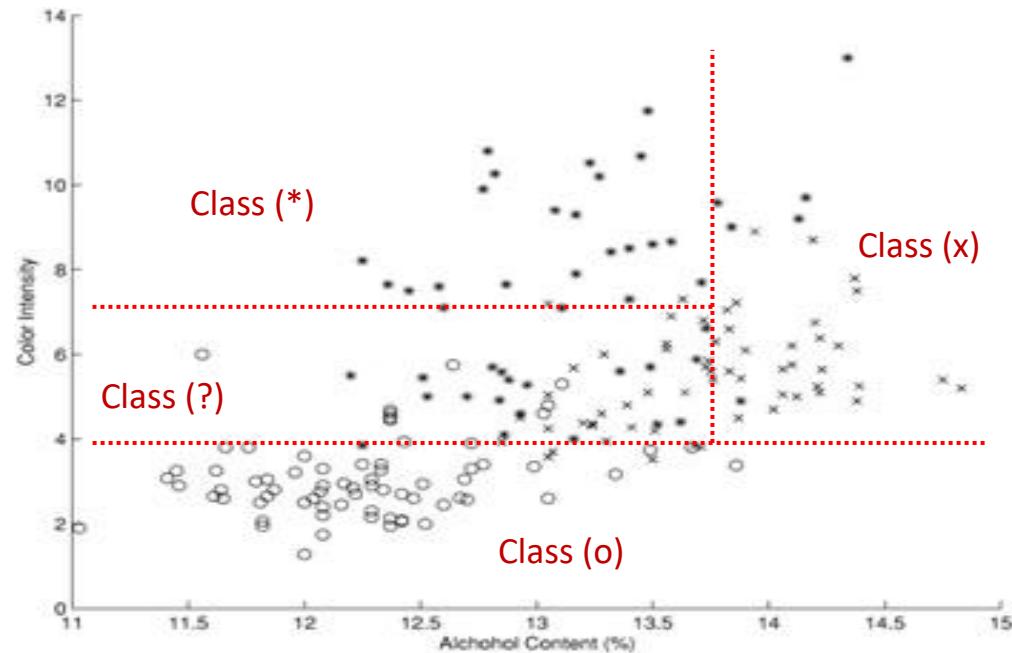


Δεδομένα...

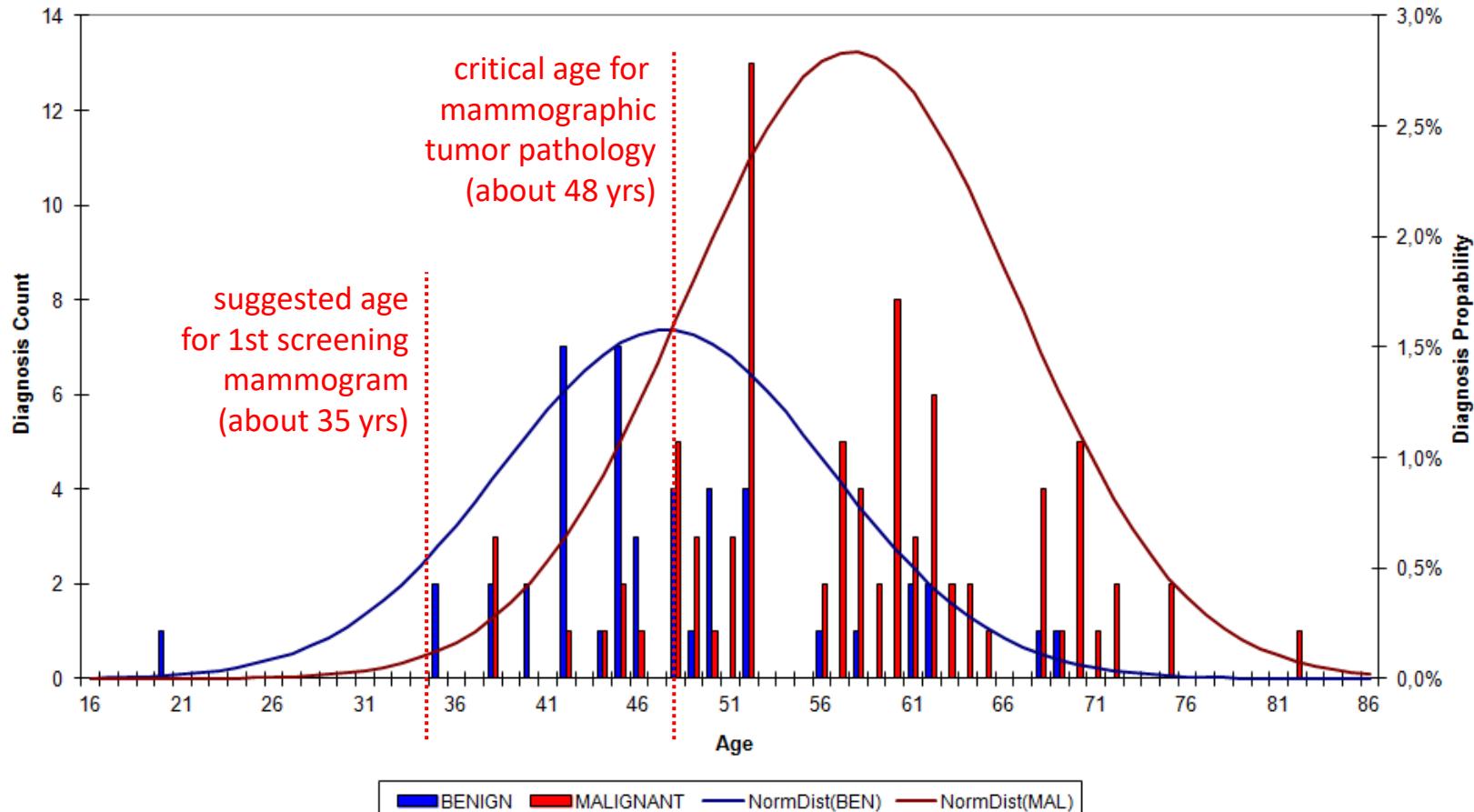
- 2-D διάνυσμα (x, y)
- Στόχος ανάλυσης: πρόβλεψη «αναμενόμενης» εξόδου για οποιαδήποτε είσοδο
- Εναλλακτικά: εντοπισμός «ακραίας συμπεριφοράς»
- Γραμμική προσαρμογή (linear regression): η «καλύτερη» ευθεία που «εξηγεί» τα δεδομένα



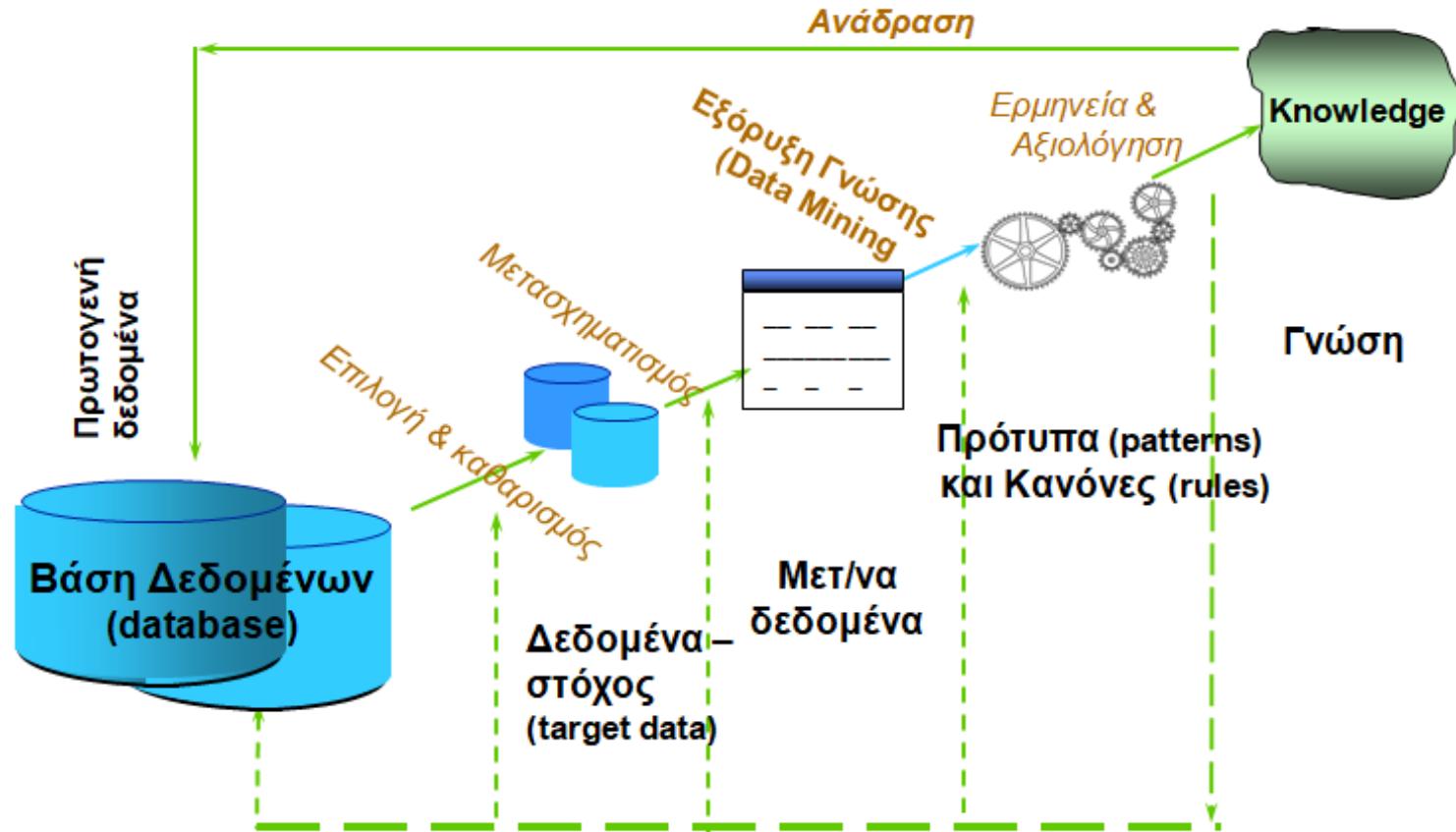
Από τα δεδομένα στη γνώση...



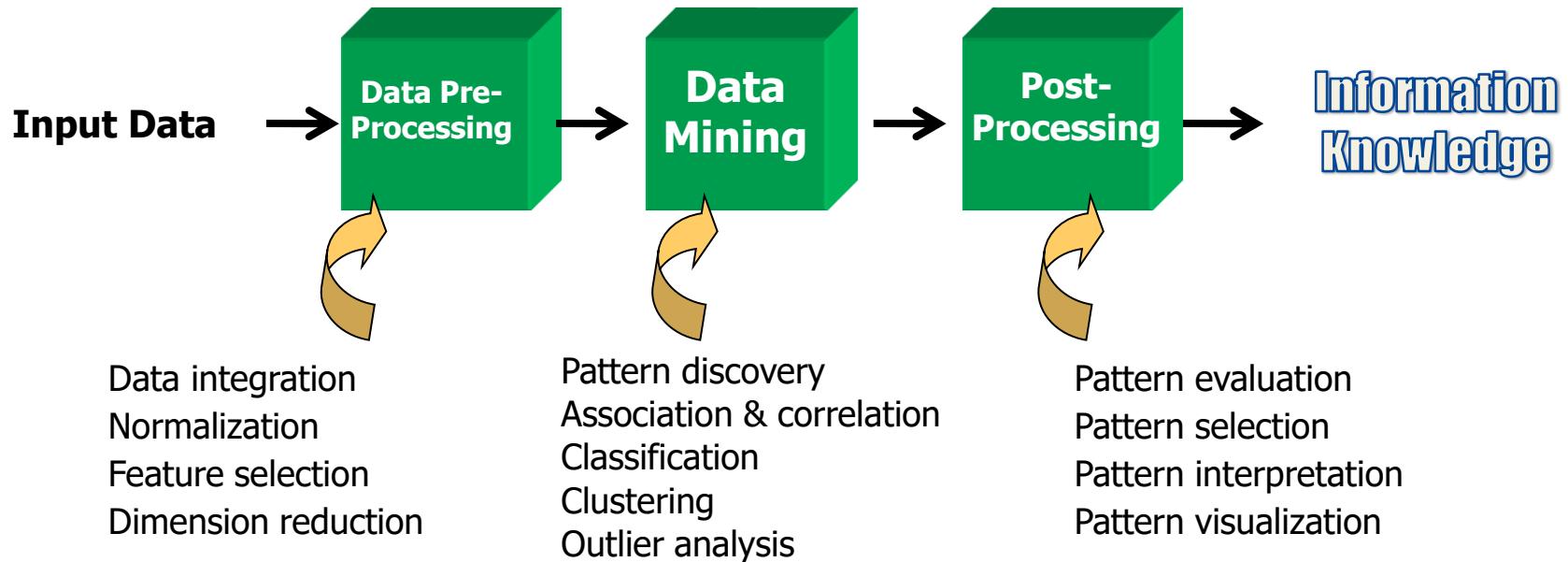
Age Distributions vs Benign/Malignant



Η “σκάλα” της διαδικασίας KDD



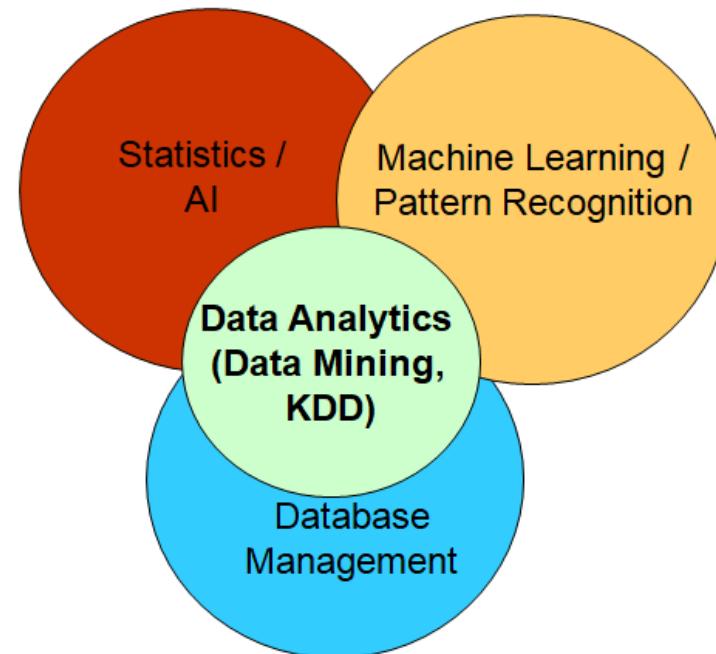
Εναλλακτικά...



- Τυπική θεώρηση από τη σκοπιά της Στατιστικής και της Μηχανικής Μάθησης

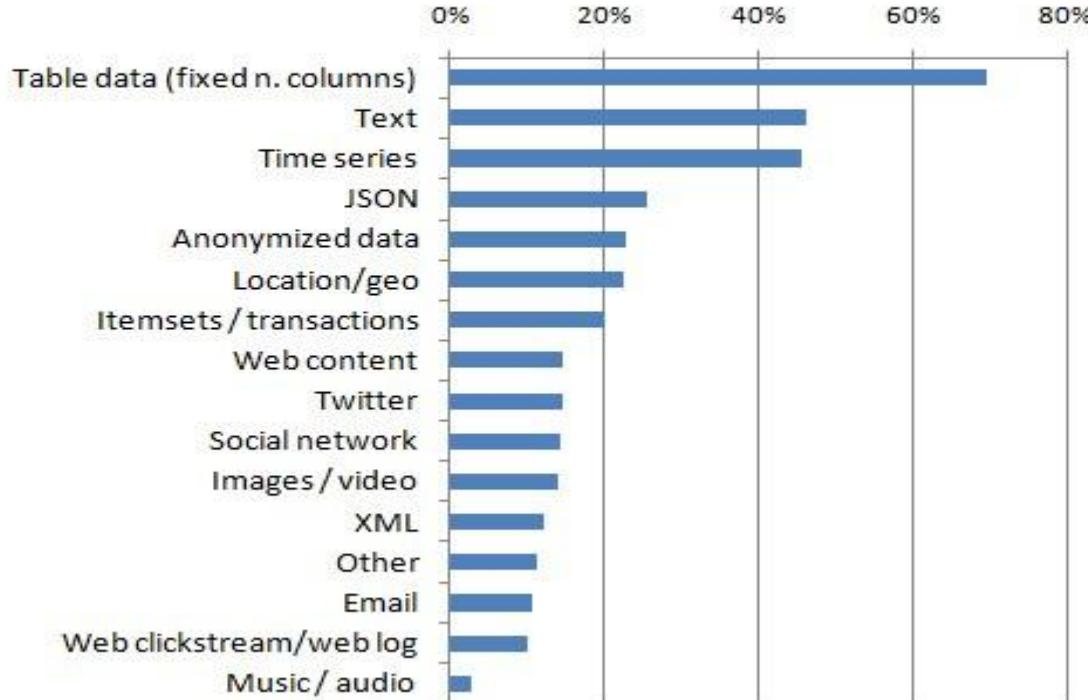
Σχετικά επιστημονικά πεδία

- Στατιστική / «Τεχνητή Νοημοσύνη», Μηχανική Μάθηση / Αναγνώριση Προτύπων, Διαχείριση Βάσεων Δεδομένων
- Οι παραδοσιακές τεχνικές επεξεργασίας δεδομένων που μας προσφέρουν αυτές οι επιστημονικές περιοχές μπορεί να είναι ανεφάρμοστες λόγω:
 - του μεγάλου όγκου,
 - των πολλών διαστάσεων,
 - της ετερογένειας των δεδομένων,
 - των απαιτήσεων επεξεργασίας,
 - ...

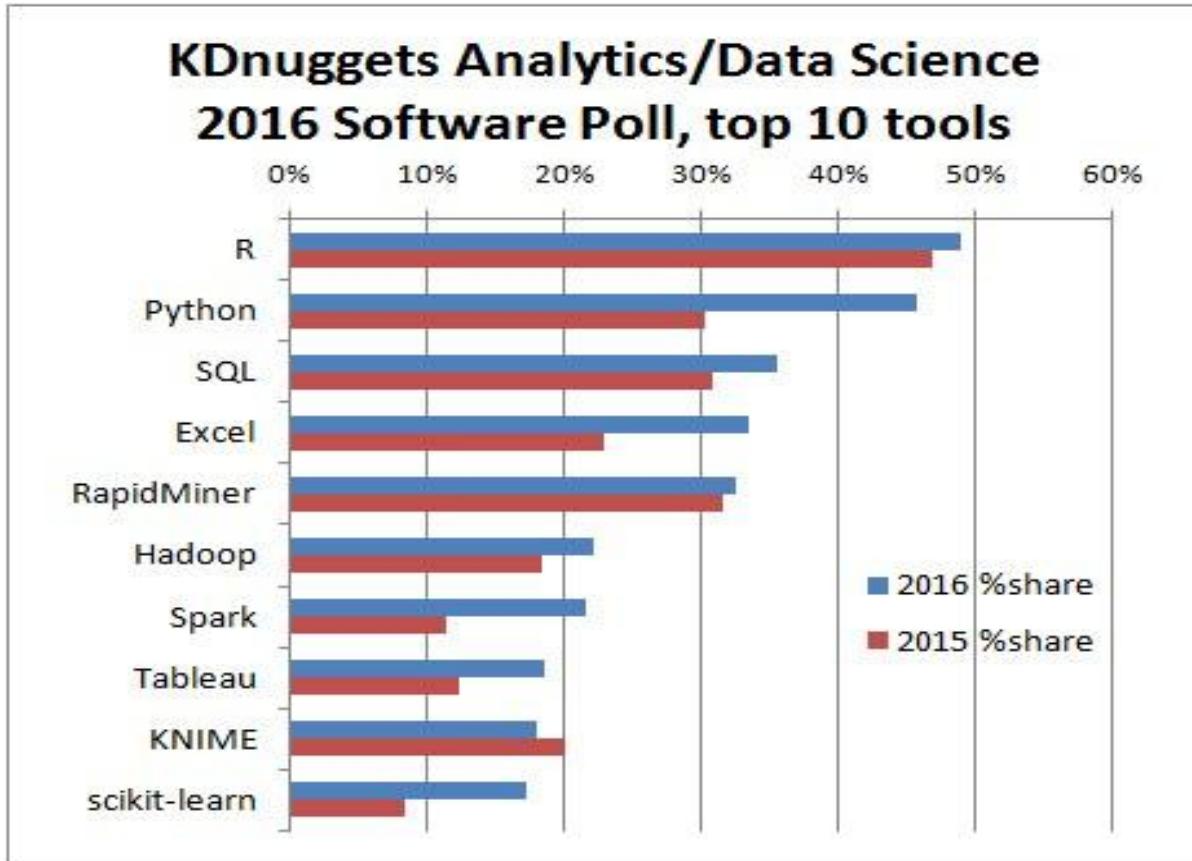


Τι δεδομένα αναλύουμε συνήθως ...

KDnuggets Poll: Data Types Analyzed, 2017



Με τι λογισμικό ...



Βασικές έννοιες & γνώσεις

- **Γνωριμία με τα δεδομένα**
 - Προπαρασκευή δεδομένων
- **Τεχνικές και αλγόριθμοι**
 - Κατηγοριοποίηση / ταξινόμηση (classification)
 - Ανάλυση συστάδων (cluster analysis)
 - Μοντέλα προσαρμογής (regression)
 - Εξόρυξη συχνών προτύπων (frequent pattern mining)
- **Παραγόμενα αποτελέσματα**
 - Προτυποποίηση κώδικα
 - Βελτιστοποίηση διαδικασιών (process optimization)
 - Ενσωματωμένα συστήματα (IoT, edge, ...)



Μέρος II: Επιλογή απλών εργαλείων

1. Διαχείριση & βασική επεξεργασία αρχείων.
2. AWK ως εργαλείο προγραμματισμού.
3. SQL για in-server επεξεργασία δεδομένων.
4. Εισαγωγή αρχείων κειμένου (imports).



CMD: tar, gzip, split/merge

```
tar -czvf name-of-archive.tar.gz /path/to/directory-or-file
```

Here's what those switches actually mean:

- **-c:** Create an archive.
- **-z:** Compress the archive.
- **-v:** Display progress in verbose mode. The v is always followed by a colon.
- **-f:** Allows you to specify the archive file name.

Gzip

The most basic way to use gzip to compress a file is:

```
% gzip filename
```

This will create a compressed version of *filename* with a .gz extension. The process type:

```
% gzip -d filename.gz or % gunzip filename.gz
```

Which removes *filename.gz* and replaces it with *filename*.

To prevent gzip from overwriting a file that you don't want to lose, you can use the -c switch where you can manipulate it as you like.

(i.e. `% gzip -c filename > compressedfile.gz`)

```
split -b 1M bigfile.txt bigfile-part-
```

In the above command, `-b` is used to specify the maximum size of each part, which is 1 MB in this case.

`bigfile.txt` is the input file, and `bigfile-part-` is the prefix that will be used for the output files.

The split command will create multiple output files, each with the prefix `bigfile-part-` followed by a two-letter suffix. For example, the first file will be named `bigfile-part-aa`, the second file will be named `bigfile-part-ab`, and so on.

File Concatenation

To concatenate multiple files into a single file, you can use the `cat` command along with the file names in the desired order. For example, let's say we have three files named 'file1.txt', 'file2.txt', and 'file3.txt' that we want to concatenate into a single file named 'allfiles.txt'.

Here's how we can use the `cat` command to achieve that:

```
...
```

```
cat file1.txt file2.txt file3.txt > allfiles.txt
```

In the above command, 'file1.txt', 'file2.txt', and 'file3.txt' are the input files that will be concatenated into a single file. The '`>`' operator is used to redirect the output to a new file named 'allfiles.txt'.

CMD: sed, find, grep, awk

NAME
awk - pattern scanning and processing language

SYNOPSIS
awk [-F *sepstring*] [-v *assignment*]... *program* [*argument...*]
awk [-F *sepstring*] -f *progfile* [-f *progfile*]... [-v
assignment]... [*argument...*]

DESCRIPTION

OPTIONS

OPERANDS

STDIN

INPUT FILES

ENVIRONMENT VARIABLES

<i>LANG</i>	<i>LC_MESSAGES</i>
<i>LC_ALL</i>	<i>LC_NUMERIC</i>
<i>LC_COLLATE</i>	<i>NLSPATH</i>
<i>LC_CTYPE</i>	<i>PATH</i>

ASYNCHRONOUS EVENTS

STDOUT

STDERR

OUTPUT FILES

EXTENDED DESCRIPTION

Overall Program Structure
pattern { action }

SED(1)

User Commands

SED(1)

NAME

sed - stream editor for filtering and transforming text

SYNOPSIS

sed [OPTION]... [script-only-if-no-other-]

just one task. In fact, one of the options in command on whatever results **find** returns. d to search for a file by *content* rather than

-exec grep -Hi penguin {} \;
like penguins.
guins are fun.

--debug

annotate program execution

-e *script*, --expression=*script*

add the script to the commands to be executed

stream editor is transformations input from a s similar to an edits (such as one pass over ntly more bility to ch particularly types of

g of pattern

AWK: data filtering

```
flights.csv
1 4744622522;7569618;GES011;LEPA;LPPT;I
;M;BE40;30/4/2016;30/4/2016;06:38:04;39.553612;2.758046;5;153.26;60.298;133.125;75.937;297
5;1461998284
2 4744625850;7569618;GES011;LEPA;LPPT;I
;M;BE40;30/4/2016;30/4/2016;06:38:09;39.55558;2.76227;8;156.289;59.743;135;78.75;3231;1461
998289
3 4744627330;7569618;GES011;LEPA;LPPT;I
;M;BE40;30/4/2016;30/4/2016;06:38:14;39.557549;2.766688;10;160.474;59.839;138.75;80.625;30
75;1461998294

LPPT LEPA 7/4/2016

xgeorgio@CircaQ8 ~
$ awk --field-separator=';' '$9 ~ /??4.2016/ {print $4 " " $5 " " $9}' flights
.csv | uniq
LEPA LPPT 30/4/2016
LPPT LEPA 7/4/2016

xgeorgio@CircaQ8 ~
$ awk --field-separator=';' '$9 ~ /??4.2016/ {print $4 " " $5 " " $9}' flights
.csv | uniq -c
 1082 LEPA LPPT 30/4/2016
  997 LPPT LEPA 7/4/2016

xgeorgio@CircaQ8 ~
$ wc flights.csv
 2078   4159  314407 flights.csv
```

AWK: linear interpolation

ts.csv interp1.awk

```
awk 'BEGIN{dt=0.5;L=0;N=0;t0=0;y0=0;t=0;y=0;sl=0;b=0}; N>0{L++;N++;sl=($2-y0)/($1-t0);b=$2-sl*$1; t=t0;while(t<$1){y=sl*t+b;print L ":" N ":"[" t0 "," $1 "]:\t(" sl ")t+" b "); \t(" t "," y ");t+=dt;N++};t0=$1;y0=$2; L==0{t0=$1;y0=$2;L++;N++}; END{print L " -> " N " samples"}' TSdata.txt >res.txt
```

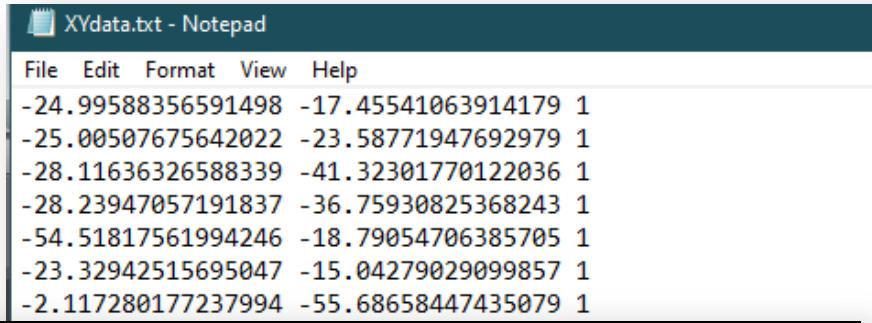
```
2:2:[1,2]: (-4)t+(16): (1,12)
2:3:[1,2]: (-4)t+(16): (1.5,10)
3:5:[2,3]: (-6)t+(20): (2,8)
3:6:[2,3]: (-6)t+(20): (2.5,5)
4:8:[3,4]: (-5)t+(17): (3,2)
4:9:[3,4]: (-5)t+(17): (3.5,-0.5)
5:11:[4,5]: (-4)t+(13): (4,-3)
5:12:[4,5]: (-4)t+(13): (4.5,-5)
6:14:[5,7]: (-2)t+(3): (5,-7)
6:15:[5,7]: (-2)t+(3): (5.5,-8)
6:16:[5,7]: (-2)t+(3): (6,-9)
6:17:[5,7]: (-2)t+(3): (6.5,-10)
```

TSdata.txt - Notepad

File Edit Format View Help

```
1 12
2 8
3 2
4 -3
5 -7
```

AWK: outliers detection



XYdata.txt - Notepad

File	Edit	Format	View	Help
-24.99588356591498	-17.45541063914179	1		
-25.00507675642022	-23.58771947692979	1		
-28.11636326588339	-41.32301770122036	1		
-28.23947057191837	-36.75930825368243	1		
-54.51817561994246	-18.79054706385705	1		
-23.32942515695047	-15.04279029099857	1		
-2.117280177237994	-55.68658447435079	1		

```
xgeorgio@CircaQ8 ~
$ awk 'BEGIN{L=2;n1=0;m1x=-30;m1y=-30;s1x=15;s1y=12;n2=0;m2x=70;m2y=70;s2x=10;s2y=7;dx=0;dy=0}; $3==1{dx=($1-m1x)/s1x;dx=(dx<0)?-dx:dx;dy=($2-m1y)/s1y;dy=(dy<0)?-dy:dy;if((dx>L) || (dy>L)){n1++;print "1: " dx " " dy}}; $3==2{dx=($1-m2x)/s2x;dx=(dx<0)?-dx:dx;dy=($2-m2y)/s2y;dy=(dy<0)?-dy:dy;if((dx>L) || (dy>L)){n2++;print "2: " dx " " dy}}; END{print "Outliers: N1=" n1 ", N2=" n2}' XYdata.txt
1: 1.85885 2.14055
1: 1.12024 2.3831
1: 0.873892 2.10238
1: 2.29492 0.909004
2: 0.486175 2.02169
2: 2.46721 1.26579
2: 2.18952 0.355079
2: 2.7715 1.48925
2: 0.145815 2.06641
Outliers: N1=4, N2=5
```

AWK: MDC classifier

```
57.83235125529826 74.52355100626579 2
70.29677449823477 64.48368848816855 2
74.17214649392676 78.31174156292951 2
66.27717399315232 75.91534413755186 2
97.71502733449357 80.42478078169438 2
62.75443458191603 71.00979441911306 2
68.581296878849 63.11891481559147 2
83.061131722715 73.89127488307699 2
85.62846221271349 64.02996796593014 2
73.57920512645296 76.11874245572909 2
69.13089429947014 75.8186395018914 2
54.65691319463033 67.91187962341608 2
82.67747215691931 73.77860757022388 2
72.16555384914786 63.5658671557541 2
62.26641707266825 77.04233803851464 2
68.54184781282795 55.53516295633679 2
```

xgeorgio@CircaQ8 ~

```
$ awk 'BEGIN{n1=0;m1x=0;m1y=0;n2=0;m2x=0;m2y=0}; $3==1{n1++;m1x+=$1;m1y+=$2}; $3==2{n2++;m2x+=$1;m2y+=$2}; END{print "Classes: N1=" n1 ", N2=" n2 "\n1: (" m1x/n1 "," m1y/n1 ") \n2: (" m2x/n2 "," m2y/n2 ")" }' XYdata.txt
```

XYdata.txt - Notepad

File Edit Format View Help

```
-24.99588356591498 -17.45541063914179 1
-25.00507675642022 -23.58771947692979 1
```

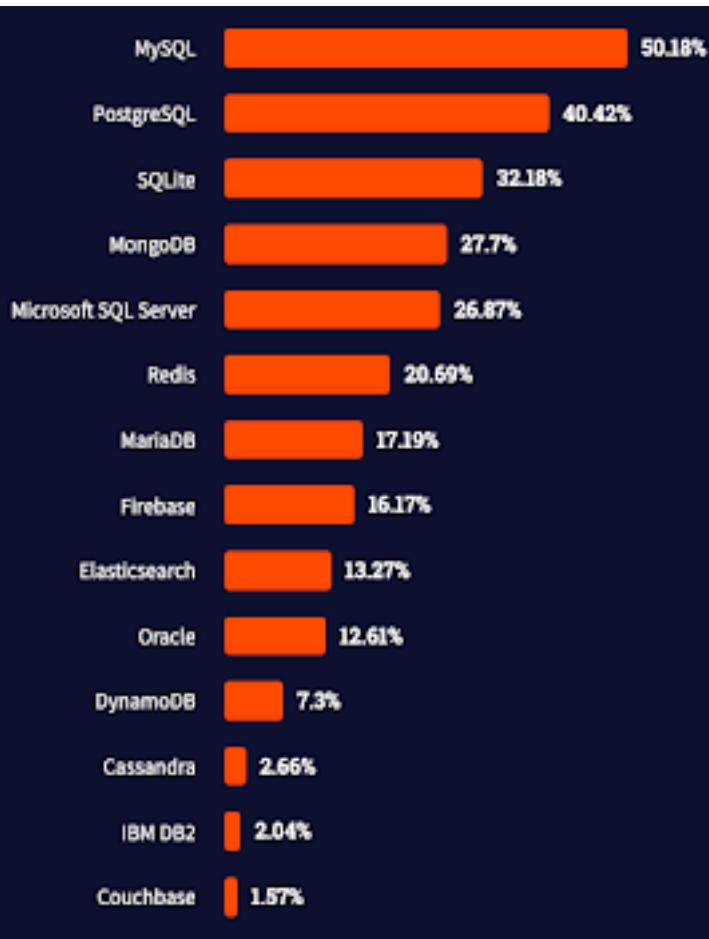
Classes: N1=60, N2=40
1: (-28.8887,-30.9343)
2: (70.4564,68.8968)

AWK: k-nn classifier

XYdata.txt - Notepad			
File	Edit	Format	View
-24.99588356591498	-17.45541063914179	1	
-25.00507675642022	-23.58771947692979	1	
-28.11636326588339	-41.32301770122036	1	
-28.23947057191837	-36.75930825368243	1	
-54.51817561994246	-18.79054706385705	1	
-23.32942515695047	-15.04279029099857	1	
-2.117280177237994	-55.68658447435079	1	

```
91: 58.2384 -> 2
92: 55.6203 -> 2
93: 52.1697 -> 2
94: 48.7571 -> 2
95: 33.3263 -> 2
96: 57.8583 -> 2
97: 43.4798 -> 2
98: 45.1941 -> 2
99: 36.5558 -> 2
min.dist[80]: 27.1149 -> class.id: 2
```

```
xgeorgio@CircaQ8 ~
$ awk 'BEGIN{x=35;y=41;i=0;mx=0} {ds[i]=sqrt((x-$1)*(x-$1)+(y-$2)*(y-$2));cl[i]
=$3;print i ":" ds[i] " -> " cl[i];i++}; END{for(k=1;k<length(ds);k++){if(ds[k]
<ds[mx]){mx=k}};print "min.dist[" mx "]": " ds[mx] " -> class.id: " cl[mx]}' XY
data.txt
```



SQL: linear regression

```
select avg(x) as x_bar,  
       avg(y) as y_bar  
  from ols;
```

```
select x, avg(x) over () as x_bar,  
       y, avg(y) over () as y_bar  
  from ols;
```

```
select sum((x - x_bar) * (y - y_bar)) / sum((x - x_bar) * (x - x_bar)) as slope  
from (  
    select x, avg(x) over () as x_bar,  
          y, avg(y) over () as y_bar  
     from ols) s;
```

```
select slope,  
      y_bar_max - x_bar_max * slope as intercept  
from (  
    select sum((x - x_bar) * (y - y_bar)) / sum((x - x_bar) * (x - x_bar)) as slope,  
          max(x_bar) as x_bar_max,  
          max(y_bar) as y_bar_max  
     from (  
        select x, avg(x) over () as x_bar,  
              y, avg(y) over () as y_bar  
         from ols) s;  
)
```

```
-- insert data samples --  
INSERT INTO DATAXY VALUES ( 1, 1.0, -1.0, -1); -- classID: '-1' --  
-1.0, -1);  
1.0, -1);  
0.0, -1);  
2.0, -1);  
  
-1.0, 1); -- classID: '+1' --  
-1.0, 1);  
1.0, 1);  
0.0, 1);  
2.0, 1);
```

$$y = mx + b$$

$$m = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b = \bar{y} - m\bar{x}$$

SQL: k-nn classifier

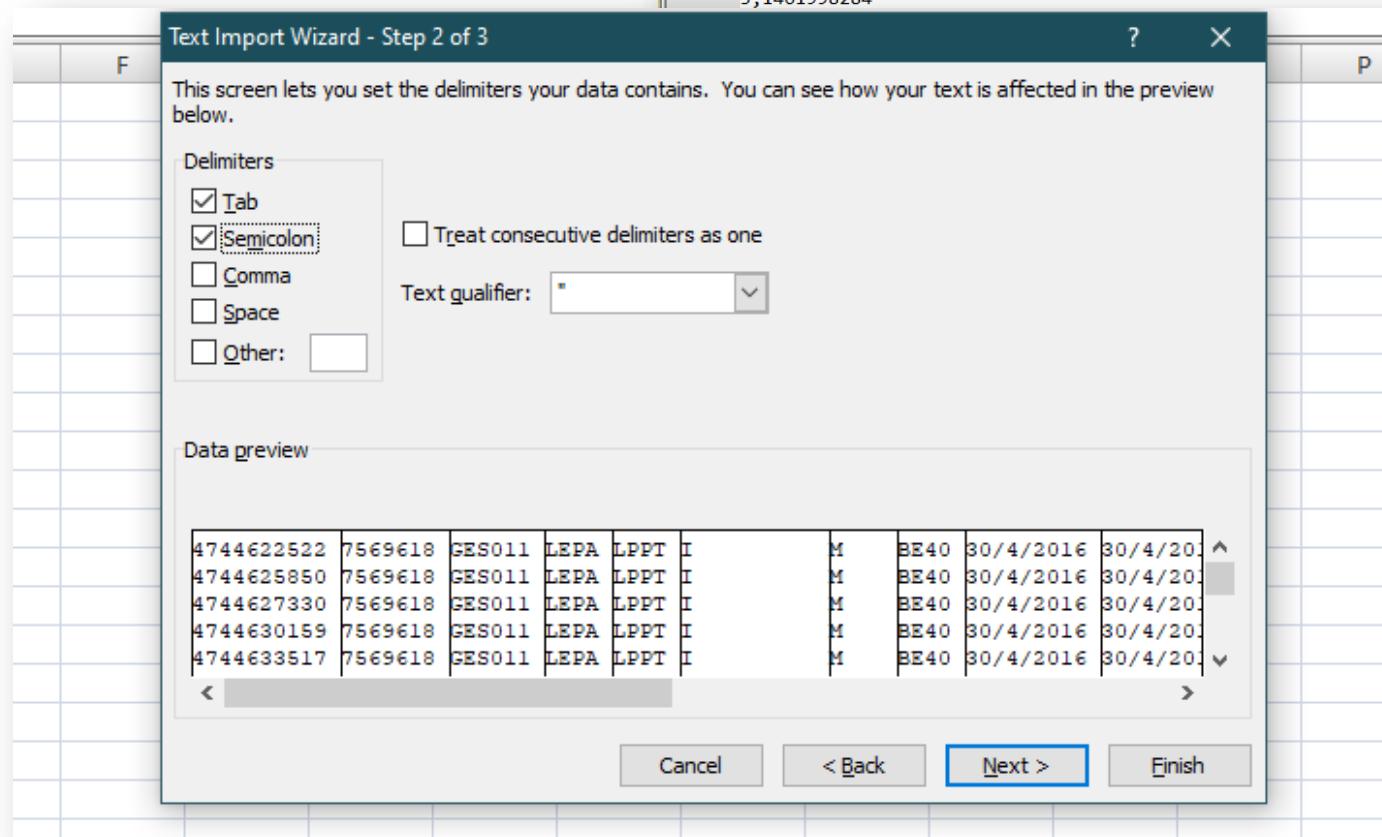
```
-- k-nn fetch, simple version (works for N=1 test sample) --
-- Note: As it is, for N>1 the results of all testing
-- samples are merged into one table without grouping
-- i.e., some FOR-EACH structure is necessary for this.

SELECT
    --id1, id2, cidk -- display the top K neighbours
    id2, SUM(cidk)/ABS(SUM(cidk)) -- display only the result
FROM TESTXY, (
    SELECT
        DATAXY.id AS id1, TESTXY.id AS id2,
        -- calculate the Euclidean distance (squared) --
        ((DATAXY.x-TESTXY.x) * (DATAXY.x-TESTXY.x)
        + (DATAXY.y-TESTXY.y) * (DATAXY.y-TESTXY.y)) AS dist,
        -- keep the corresponding class ID --
        DATAXY.cid AS cidk
    FROM DATAXY, TESTXY
    ORDER BY id2, dist ASC -- sort against distances
    LIMIT 5 -- this is the K parameter
);
```

```
-- insert data samples --
INSERT INTO DATAXY VALUES ( 1, 1.0, -1.0, -1); -- classID: '-1' --
2.0, -1.0, -1);
2.5, 1.0, -1);
3.0, 0.0, -1);
3.5, 2.0, -1);

11.0, -1.0, 1); -- classID: '+1' --
12.0, -1.0, 1);
12.5, 1.0, 1);
13.0, 0.0, 1);
13.5, 2.0, 1);
```

Data import (text)



F	P
	289;59.743;135;78.75;3231;1461
	50.474;59.839;138.75;80.625;30
	51.285;60.007;139.687;80.625;3
	53.038;59.601;140.625;82.5;325
	.688;59.676;149.062;87.187;323
	74.448;59.3;150;89.062;3113;14
	77.823;59.243;152.812;90.937;3
	81.681;58.934;155.625;93.75;29
	Col:1 Pos:1 Windows (C)

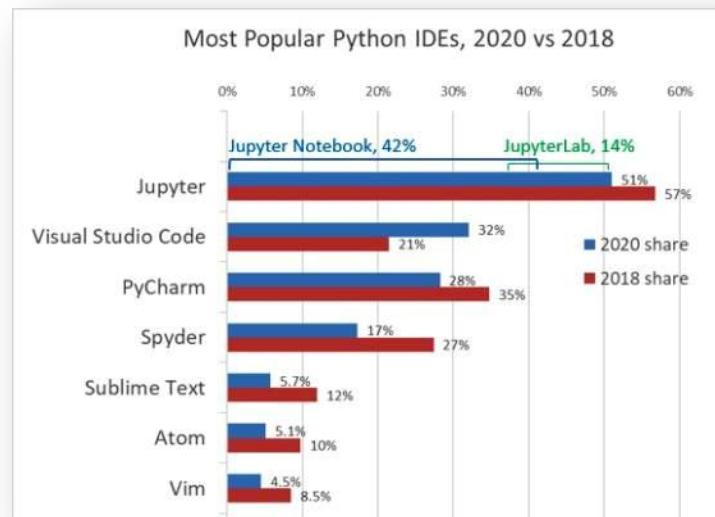
Μέρος III: Προηγμένα εργαλεία

1. Χρήση στην κονσόλα (command-line).
2. Χρήση σε περιβάλλον IDE.
3. Χρήση online στο cloud.
4. Και όχι μόνο...

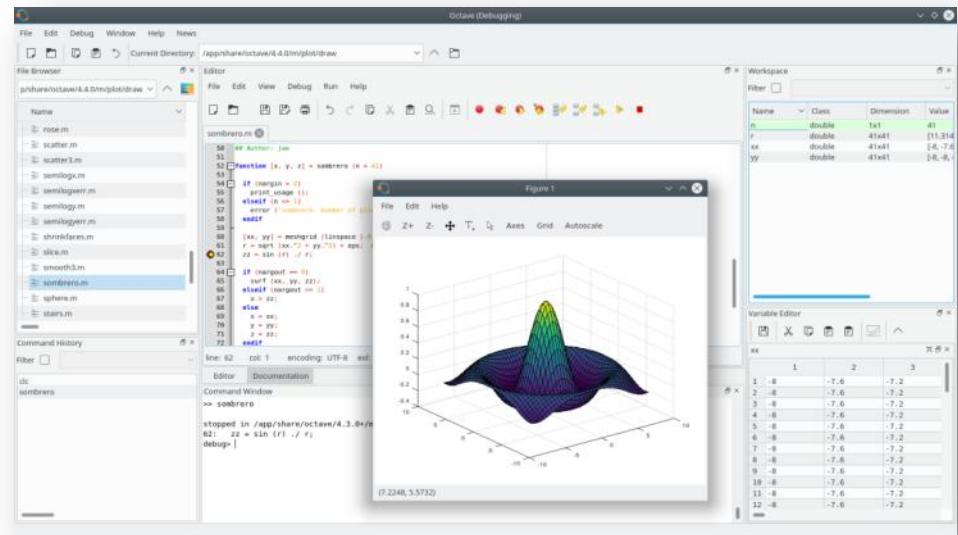


Specialized scripting: Python, Octave/Matlab, ...

```
import numpy as np  
x = np.arange(data)  
y = np.arange(data)  
m, b = np.polyfit(x, y, 1)
```



```
A = [X + ones(n,1)] \ Y
```



Linear Regression



```
## correlation between CPI and year / quarter
cor(year, cpi)

## [1] 0.9096316

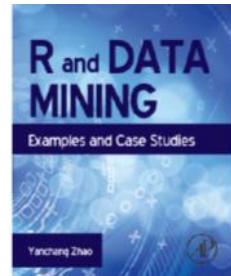
cor(quarter, cpi)

## [1] 0.3738028

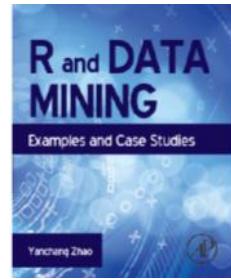
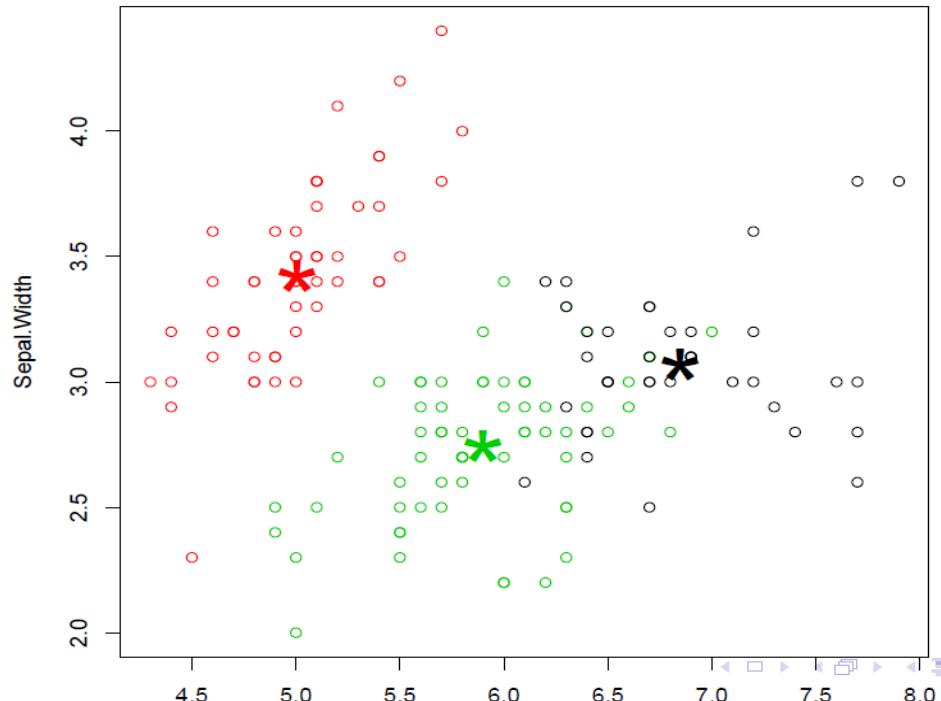
## build a linear regression model with function lm()
fit <- lm(cpi ~ year + quarter)
fit

## 
## Call:
## lm(formula = cpi ~ year + quarter)
## 
## Coefficients:
## (Intercept)          year        quarter
## -7644.488           3.888       1.167
```

$$cpi = c_0 + c_1 * \text{year} + c_2 * \text{quarter},$$



```
# plot clusters and their centers  
plot(iris2[c("Sepal.Length", "Sepal.Width")], col=iris.kmeans$cluster)  
points(iris.kmeans$centers[, c("Sepal.Length", "Sepal.Width")],  
       col=1:3, pch="*", cex=5)
```



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter

Choose None

Current relation

Relation: breast-cancer Attributes: 10 Instances: 286 Sum of weights: 286

Attributes

All None Invert Pattern

No.	Name
1	age
2	menopause
3	tumor-size
4	inv-nodes
5	node-caps
6	deg-malig
7	breast
8	breast-quad
9	irradiat

Remove

Selected attributes

Name: age
Missing: 0

No. 1 2 3 4 5

Class: Class (Nom)

Start Stop

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

Classifier output

Correctly Classified Instances 601 78.2552 %
Incorrectly Classified Instances 167 21.7448 %
Kappa statistic 0.4966
Mean absolute error 0.3063
Root mean squared error 0.3908
Relative absolute error 67.3928 %
Root relative squared error 81.9907 %
Total Number of Instances 768

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
0	0.890	0.418	0.799	0.890	0.842	0.504
1	0.582	0.110	0.739	0.582	0.651	0.504
Weighted Avg.	0.783	0.310	0.778	0.783	0.775	0.504

==== Confusion Matrix ===

a	b	--- classified as
445	55	a = tested_negative
112	156	b = tested_positive

Status

OK

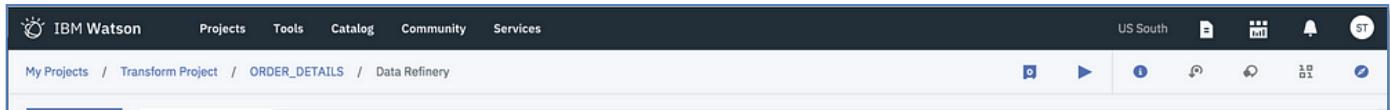
Log x 0

WatsonX (IBM)

IBM Watson Projects Tools Catalog Community Services

My Projects / Transform Project / ORDER_DETAILS / Data Refinery

US South



x Operation Code an operation to clean up

Join

Left join

Returns all rows in the original data set and retains matching rows in the joining data set. Returns the original data set for each matching row in data set.

The default suffix for each data set will be used to differentiate any duplicate column names in the data set.

Source

ORDER_DETAILS	Data set to join
*Suffix .x	ORDER_HEAD

JOIN KEYS	JOIN KEYS
ORDER_DETAILS	ORDER_HEAD
ORDER_DETAIL_CO...	ORDER_NUM

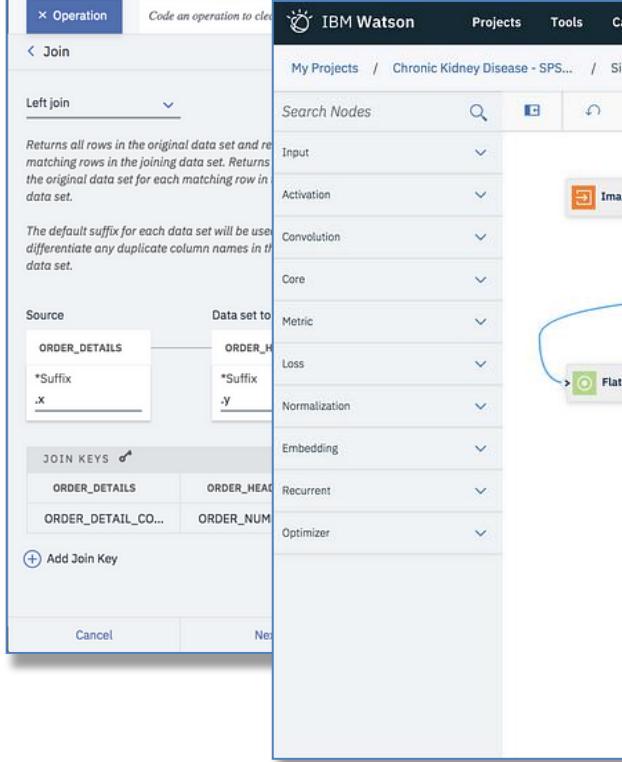
+ Add Join Key

Cancel Next

IBM Watson Projects Tools Catalog Community Services

My Projects / Chronic Kidney Disease - SPS... / Single Convolution layer on M

US South



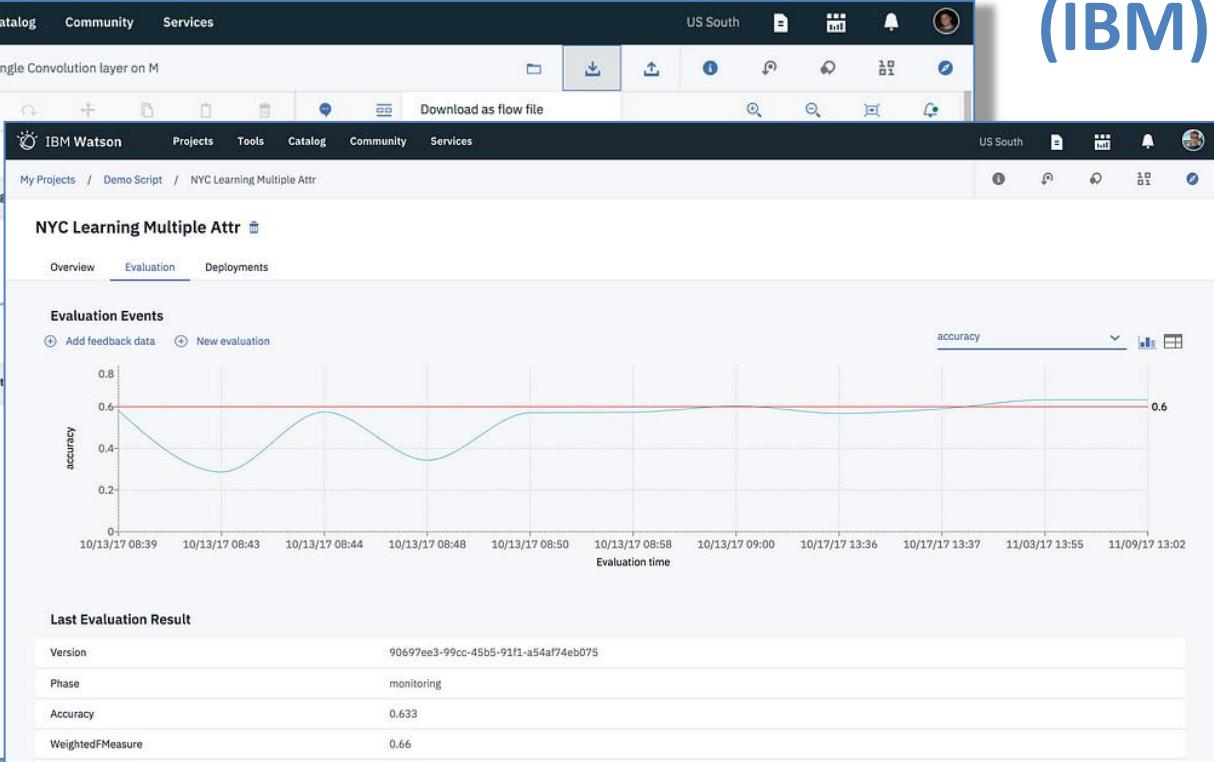
IBM Watson Projects Tools Catalog Community Services

My Projects / Demo Script / NYC Learning Multiple Attr

US South

Evaluation Events

Add feedback data New evaluation



Evaluation Events

accuracy

accuracy

10/13/17 08:39 10/13/17 08:43 10/13/17 08:44 10/13/17 08:48 10/13/17 08:50 10/13/17 08:58 10/13/17 09:00 10/17/17 13:36 10/17/17 13:37 11/03/17 13:55 11/09/17 13:02

Evaluation time

Last Evaluation Result

Version	90697ee3-99cc-45b5-91f1-a54af74eb075
Phase	monitoring
Accuracy	0.633
WeightedFMeasure	0.66
WeightedRecall	0.633
WeightedPrecision	0.693

Performance Monitoring Edit configuration





File Edit View Insert Cell Kernel Help

Python 3



Markdown Cell Toolbar

```
In [1]: # setup the matplotlib graphics library and configure it to show
# figures inline in the notebook
%matplotlib inline
import matplotlib.pyplot as plt
import numpy as np
```

```
In [2]: # make qutip available in the rest of the notebook
from qutip import *

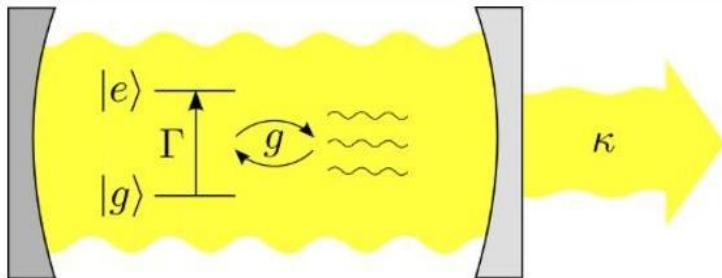
from IPython.display import Image
```

Introduction and model

Consider a single atom coupled to a single cavity mode, as illustrated in the figure below. If there atom excitation rate Γ exceeds the relaxation rate, a population inversion can occur in the atom, and if coupled to the cavity the atom can then act as a photon pump on the cavity.

```
In [3]: Image(filename='images/schematic-lasing-model.png')
```

Out[3]:



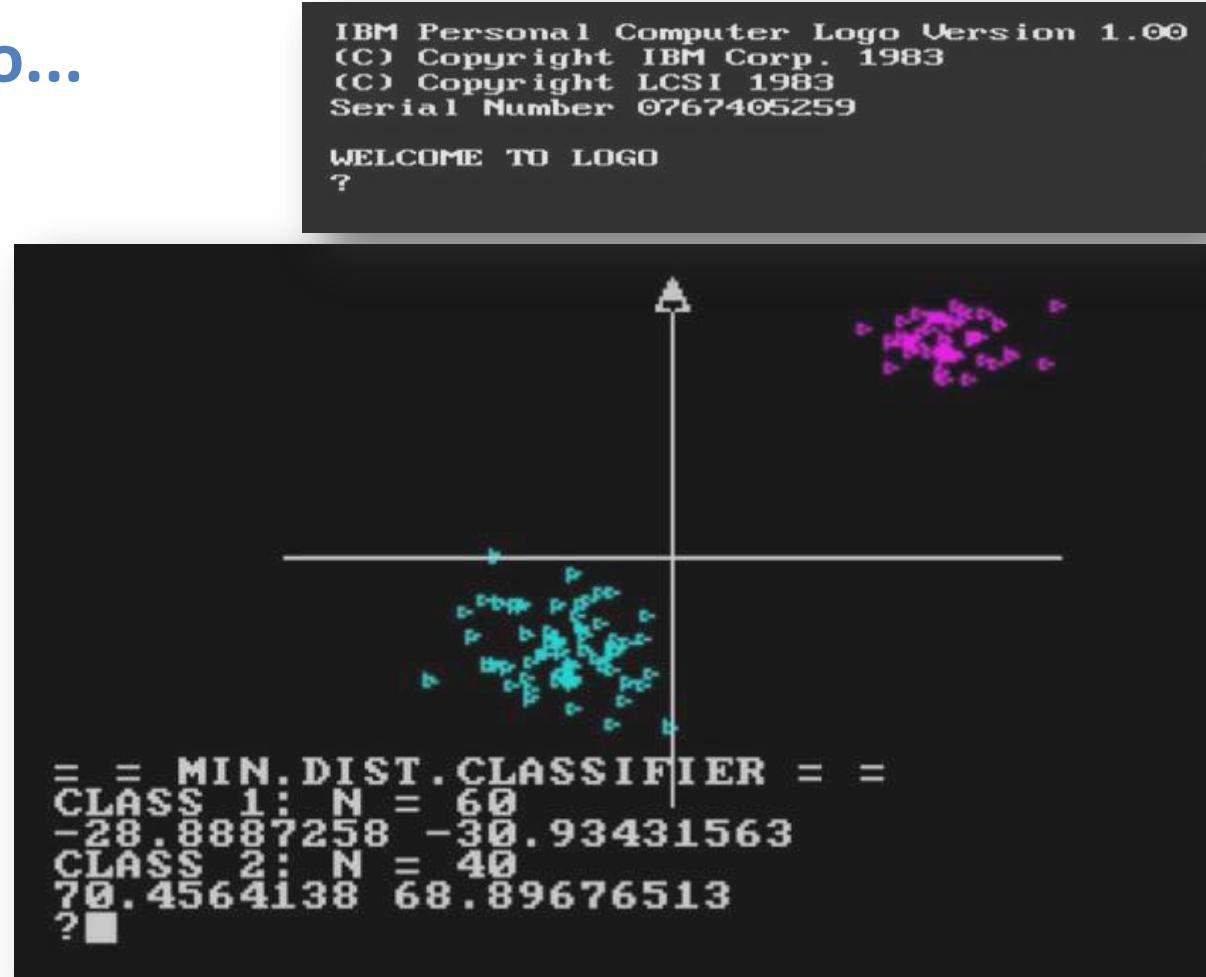
The image displays a composite screenshot of several software interfaces used for data analysis and visualization:

- Top Left:** A file browser window showing a directory structure with notebooks like "notebooks.ipynb", "audio.ipynb", "images.ipynb", "Altair.ipynb", "Copy.ipynb", "Data.ipynb", "Fastai.ipynb", and "Julia.ipynb".
- Top Center:** A Jupyter Notebook interface titled "In Depth: Linear Regression". It contains a section titled "Just as naive Bayes (discussed earlier in In Depth: Naive Bayes Classification) is a good starting point for classification tasks, linear regression models are a good starting point for regression tasks. Such models are popular because they can be fit very quickly, and are very interpretable. You are probably familiar with the simplest form of a linear regression model (i.e., fitting a straight line to data), but such models can be extended to model more complicated data behaviors." Below this is a code cell with a plot of Seattle weather data.
- Middle Left:** A Jupyter Notebook interface titled "Simple". It shows a code cell with a scatter plot and a corresponding notebook metadata section.
- Middle Right:** A Jupyter Notebook interface titled "Seattle Weather: 2012-2015". It displays a scatter plot of "Maximum Daily Temperature (F)" versus time.
- Bottom Left:** A Python 3 Idle interface showing a code editor with a script named "sklearn.ipynb". The code imports numpy, matplotlib, and scikit-learn, generates data, performs calculations using a spline, and plots the results.
- Bottom Center:** A Jupyter Notebook interface titled "Julia". It shows a code cell with Julia code for generating data, performing calculations using a spline, and plotting the results.
- Bottom Right:** A Jupyter Notebook interface titled "Eigen". It shows a code cell with Eigen C++ code for matrix operations and printing eigenvalues.

MDC even in Logo...

```
TO RDREC1
MAKE "XYC READLIST
IF :XYC = [] [STOP]
PRINT :XYC
MAKE "X (FIRST :XYC)
MAKE "Y (FIRST BUTFIRST :XYC)
MAKE "C INT (LAST :XYC)
IF :C = 1 [UPDC1]
IF :C = 2 [UPDC2]
MAKE "X ROUND :X
MAKE "Y ROUND :Y
MARKXY :X :Y :C
RDREC1
END

TO MDC
CS
CT
OPEN "XYDATA.TXT
SETREAD "XYDATA.TXT
MAKE "SX1 0
MAKE "SY1 0
MAKE "N1 0
MAKE "SX2 0
MAKE "SY2 0
```



Σύνοψη

- Περιεχόμενα:
 - Τι είναι η Μηχανική Μάθηση και η Αναλυτική Δεδομένων (ML/DA).
 - Προπαρασκευή δεδομένων (pre-processing), είδη προβλημάτων ML/DA.
 - Επεξεργασία αρχείων, command-line tools, SQL, αρχεία κειμένου (imports).
 - Προηγμένα εργαλεία στην κονσόλα (command-line), σε IDE, στο cloud.
- Αναφορές:
 - «Εισαγωγή στη Μηχανική Μάθηση και στην Αναλυτική Δεδομένων», Χ. Γεωργίου, Α' κύκλος ανοικτών μαθημάτων ΕΠΕ – <https://youtu.be/mlU4SvyfRqA>
 - «Εισαγωγή στην Αναλυτική Δεδομένων με τη γλώσσα R», Χ. Γεωργίου, Β' κύκλος ανοικτών μαθημάτων ΕΠΕ – <https://youtu.be/7aaNnXxyzgl>
 - Dunham: Data Mining – Introductory and Advanced Topics. Prentice Hall, 2003.
 - Tan, Steinbach, Kumar: Introduction to Data Mining. Addison Wesley, 2006.

```

MOVE 1 TO DATA-C(N-T).
ADD 1 TO N-CHANGED.
GO TO LOOP-SCAN.

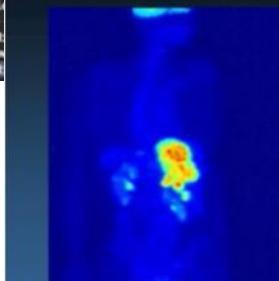
SELECT-CLZ.
ADD DATA-X(N-T) TO SUM2-X.
ADD DATA-Y(N-T) TO SUM2-Y.
ADD 1 TO N-CLZ.
IF DATA-C(N-T) EQUAL 2 GO TO LOOP-SCAN.
MOVE 2 TO DATA-C(N-T).
ADD 1 TO N-CHANGED.

```

```

LOOP-
  id : Integer := 0; -- target ID (counter)
  det : Integer := 0; -- detection slots in sequence
  pwr : Integer := 0; -- rel. power of detection
  pwr0 : Integer := detLimit; -- rel. power baseline (adapt
  disp : Boolean := False; -- target reporting (flag)
  begin
    -- process the FOV slots --
    for p in 1..(seekerData'Length)-1 loop
      -- rel. power is current detection 'step'
      pwr := abs(seekerData(p+1)-seekerData(p));
      if pwr >= detLimit then
        -- detection valid, continue analysis
        if pwr > pwr0 then
          -- strong new 'step' from baseline (new target)
          pwr0 := pwr; -- update the baseline
          det := 0; -- reset the run-length
          disp := False; -- enable target reporting
        end if;
      end if;
    end loop;
    det := 0;
    if
      if
        if
          if
            if
              if
                if
                  if
                    if
                      if
                        if
                          if
                            if
                              if
                                if
                                  if
                                    if
                                      if
                                        if
                                          if
                                            if
                                              if
                                                if
                                                  if
                                                    if
                                                      if
                                                        if
                                                          if
                                                            if
                                                              if
                                                                if
                                                                  if
                                                                    if
                                                                      if
                                                                        if
              end if;
            end if;
          end if;
        end if;
      end if;
    end if;
  end loop;
end;

```



Παράδειγμα ιρισθείσης αντικειμένων μεντηίκης πλούτου (γύναικα) – Wikipedia.org

▶ ▶ ⏪ 12:40 / 2:00:20

- Εικόνα (2-D): Επικαυπτόμενες δομές ιστών
 - Τομογραφία (3-D): Όγκος πληροφοριών
 - Διαφορετικές τεχνολογίες απεικόνισης
 - Διαφορετικά διαγνωστικά χαρακτηριστικά
 - Η διαγνωστική πληροφορία συνήθως δεν είναι καλώς ορισμένη (θόρυβος, ασάφειες δομών)
 - Η διαγνωστική διαδικασία είναι συνήθως ασαφής, πολύπλοκη και βασίζεται στην εμπειρία (ιατρός)
- ⇒ Η χρήση Η/Υ επιτρέπει την αυτόματη επεξεργασία και ενσυνοίση (τομογραφία) μεγάλου όγκου δεδομένων απεικόνισης
- ⇒ ...αλλά εξακολουθεί να έχει σημαντικούς περιορισμούς ως προς τη σημασιολογική ερμηνεία τους (διαγνωστική πληροφορία)

- Hamming (7,4) error correction codes in **R**
- Kmeans clustering in **COBOL**
- Bi-directional Associative Memory (BAM) in **Arduino/C**
- Linear Regression in **SQL, Matlab**
- ...

YouTube:

@ApneaCoding



<https://www.youtube.com/@apneacoding>

<https://www.facebook.com/apneacoding>

Github:

@xgeorgio



<https://github.com/xgeorgio>

<http://apneacoding.eu>

Ένας ψηφιακός κόσμος γεμάτος γνώση για όλους

Σύμφωνα με το Καταστατικό της Ένωσης Πληροφορικών Ελλάδας, ένας από τους βασικούς σκοπούς της λειτουργίας της είναι η προώθηση της γνώσης και χρήσης των πληροφορικών αγαθών από το κοινωνικό σύνολο και η εξάλειψη της τεχνοφοβίας και του "αναλφαβητισμού" στην Πληροφορική.



<https://courses.epe.org.gr>

Σχετικά με τα ανοικτά μαθήματα της Ένωσης Πληροφορικών Ελλάδας:

- ✓ Τα μαθήματα πραγματοποιούνται εξ ολοκλήρου διαδικτυακά, ζωντανά μέσω της πλατφόρμας Zoom.
- ✓ Η συμμετοχή σε όλα τα μαθήματα είναι ελεύθερη για οποιονδήποτε από οποιδήποτε στην Ελλάδα ή στο εξωτερικό.
- ✓ Δεν υπάρχει οικονομικό κόστος ή άλλες προϋποθέσεις συμμετοχής.
- ✓ Οι Εισιγητές είναι μέλη της Ένωσης Πληροφορικών Ελλάδας και πραγματοποιούν τα μαθήματα εθελοντικά.
- ✓ Τα μαθήματα μαγνητοσκοπούνται και παραμένουν διαθέσιμα για σύγχρονη παρακολούθηση στο Αρχείο Μαθημάτων.
- ✓ Η εκπαίδευση που παρέχεται μέσω των ανοικτών διαδικτυακών μαθημάτων είναι άτυπη και δεν παρέχονται βεβαιώσεις παρακολούθησης στους συμμετέχοντες.



Ερωτήσεις



Χάρης Γεωργίου (MSc,PhD)

<https://www.linkedin.com/in/xgeorgio/>

https://twitter.com/xgeorgio_gr