



Μηχανική Μάθηση Machine Learning ΜΑΘΗΜΑ 1

Δρ. Παπαδάκης Στυλιανός,
Τμήμα Διοικητικής Επιστήμης & Τεχνολογίας
Ελληνικό Μεσογειακό Πανεπιστήμιο
e-mail: spap@hmu.gr

Μηχανική Μάθηση

- Προτεινόμενο σύγγραμμα:
 - Κων/νος Διαμαντάρας Τεχνητά Νευρωνικά δίκτυα, Εκδόσεις Κλειδάριθμος
 - Παπαδάκης Στέλιος και Διαμαντάρας Κώστας, 'Προγραμματισμός και αρχιτεκτονική συστημάτων παράλληλης Επεξεργασίας, Εκδόσεις Κλειδάριθμος
 - Eberhart, R.C., Shi, Y., 2011. Computational intelligence: concepts to implementations. Elsevier.
 - Σημειώσεις
- Τρόπος Εξέτασης
 - Γραπτές εξετάσεις
 - Εργασίες
- Ώρες συνεργασίας με τους φοιτητές
 - 24/7
 - Email: spap@hmu.gr

Νοημοσύνη

- Ακριβής, σαφής και μοναδικός ορισμός δεν υπάρχει
- Αποτελεί ανοικτό φιλοσοφικό ζήτημα, (π.χ. επιστημολογία, γνωσιολογία)
- Γνωρίζουμε κάποιες ιδιότητες που χαρακτηρίζουν τη νοημοσύνη
 - Ικανότητα επεξεργασίας δεδομένων
 - Προσαρμογή σε κάποιο περιβάλλον
 - Ικανότητα εξαγωγής γνώσης από τα δεδομένα
 - Λήψη αποφάσεων
- Δεν είναι η μοναδική έννοια που δεν είναι κατανοητή (δες π.χ. Ενέργεια, χρόνος)

Μάθηση

- Και εδώ ακριβής, σαφής και μοναδικός ορισμός δεν φαίνεται να υπάρχει.
- Αποτελεί επίσης ανοικτό φιλοσοφικό ζήτημα
- Ένας **βολικός ορισμός** για τις ανάγκες του μαθήματος θα μπορούσε να είναι:

Η λήψη δεδομένων από το περιβάλλον η επεξεργασία τους και αξιοποίηση τους, με στόχο την αμφίδρομη προσαρμογή, τον έλεγχο και τη σταδιακή βελτίωση στην εκτέλεση κάποιων λειτουργιών

Μηχανική μάθηση

- Στο πεδίο της μηχανικής μάθησης περιορίζουμε την έννοια της μάθησης στην ικανότητα προσαρμογής τεχνητών **συστημάτων** σε συγκεκριμένα δεδομένα
- **Alan Turing:** “Υπολογιστικές μηχανές και Νοημοσύνη”
 - Μπορούν οι μηχανές να σκεφτούν;
 - Μπορούν οι μηχανές να κάνουν αυτό που κάνουμε εμείς;
- **Tom Mitchel:** Ένα πρόγραμμα υπολογιστή λέμε ότι μαθαίνει από την εμπειρία E ως προς μια ομάδα εργασιών T και ένα μέτρο επίδοσης P , Αν η επίδοση P στις εργασίες βελτιώνεται με την εμπειρία E .
- Αν μια μηχανή παρουσιάζει κάποιες από τις ιδιότητες της μάθησης τότε μιλάμε για μηχανική μάθηση.
- Είναι ένα σύνολο αλγορίθμων, μεθοδολογιών, είτε εμπειρικών είτε μαθηματικών (στατιστική) τα οποία δίνουν τη δυνατότητα σε υπολογιστικές μηχανές να μαθαίνουν από την εμπειρία.

Τεχνητή Νοημοσύνη

- Χωρίς να ξέρουμε τι ακριβώς είναι η μάθηση και τι ακριβώς είναι η νοημοσύνη, θεωρούμε ότι η μάθηση είναι βασική συνιστώσα της νοημοσύνης
- Άρα μπορούμε να θεωρήσουμε ότι και η μηχανική μάθηση είναι βασική συνιστώσα της τεχνητής νοημοσύνης
- Αν μια μηχανή έχει την ικανότητα μάθησης τότε έχει ιδιότητες νοημοσύνης.
- Η τεχνητή νοημοσύνη είναι η φιλοδοξία κατασκευής νοημόνων μηχανών
- Υλοποιείται κατά βάση με
 - μαθηματική λογική (π.χ. prolog)
 - Δένδρα αποφάσεων

Παράδειγμα επίλυσης λογικών προβλημάτων Puzzle

- Βρισκόμαστε σε ένα νησί το οποίο έχει δυο τύπους κατοίκων
 - **Ιππότες** οι οποίοι λένε πάντα αλήθεια
 - **Σκλάβους** οι οποίοι λένε πάντα ψέματα
- Συναντάμε τυχαία δυο κατοίκους A,B και θέλουμε να δούμε αν είναι ιππότες η σκλάβοι με βάση τα λεγόμενα τους.
- Κάνουμε την εξής ερώτηση στον A,B: Είστε ίδιοι ή διαφορετικοί?
 - A: Είμαστε Διαφορετικοί
 - B: Είμαστε Ίδιοι

Επίλυση

- Παίρνουμε όλες τις δυνατές περιπτώσεις στον παρακάτω πίνακα αληθείας:

A	B	A: Είμαστε Διαφορετικοί	B: Είμαστε ίδιοι
Ιππότης	Ιππότης	0	1
Ιππότης	Σκλάβος	1	1
Σκλάβος	Ιππότης	0	0
Σκλάβος	Σκλάβος	1	0

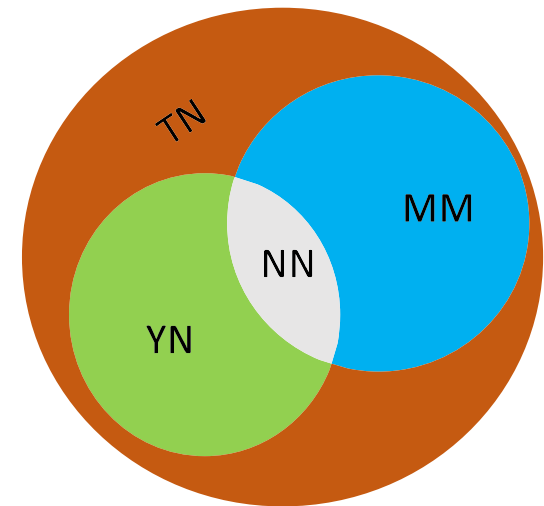
- **Prolog:** Γλώσσα λογικού προγραμματισμού
- Προσπάθεια επίτευξης νοημοσύνης μέσω της λογικής
- Περιορισμοί της λογικής, αυστηρή μαθηματική γλώσσα, χαλαρή φυσική γλώσσα
- Για κάθε πρόβλημα πρέπει να υπάρξει διαφορετικός προγραμματισμός της μηχανής επικεντρωμένος στο πρόβλημα ώστε να μπορέσει να το επιλύσει.

Υπολογιστική νοημοσύνη

- Είναι ένα σύνολο
 - Αλγορίθμων,
 - Μεθοδολογιών
 - Στατιστικής
 - Εμπειρικών κανόνωνπου προσπαθεί να επιτύχει νοημοσύνη μέσω επεξεργασίας δεδομένων
- Περιλαμβάνει κατά βάση τις θεματικές περιοχές:
 - Εξελικτικοί αλγόριθμοι (προσομοίωση της φυσικής επιλογής)
 - Νευρωνικά δίκτυα (προσομοίωση του εγκεφάλου)
 - Ασαφή συστήματα (προσομοίωση του ανθρώπινου τρόπου σκέψης)

Μηχανική μάθηση

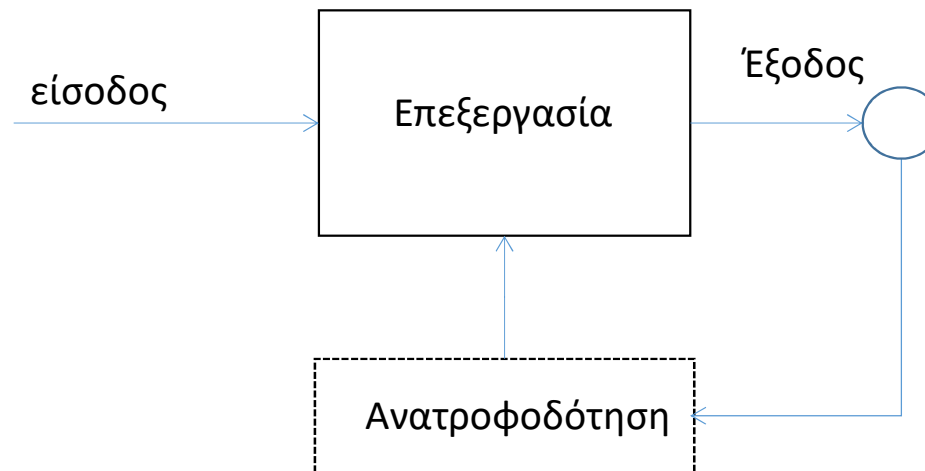
- Είναι επίσης ένα σύνολο αλγορίθμων, μεθοδολογιών που προσπαθεί να επιτύχει νοημοσύνη μέσω επεξεργασίας δεδομένων.
- Περιλαμβάνει οτιδήποτε δεν είναι υπολογιστική νοημοσύνη (ΕΑ, ΝΔ, ΑΣ)
- Συμπλήρωμα/επέκταση της υπολογιστικής νοημοσύνης



Κάθε πρόβλημα αντιμετωπίζεται ως «Μαύρο Κουτί»

- Δεν είναι γνωστή η αναλυτική συμπεριφορά του συστήματος
- Παραδείγματα:
 - «Ένα νέο κρασί θα αρέσει στους πελάτες» ? Μπορούμε να το ξέρουμε αυτό πριν το κρασί να βγει στην αγορά?
 - Ένας νέος δανειολήπτης θα είναι συνεπής? Μπορούμε να το ξέρουμε αυτό πριν δώσουμε το δάνειο?
 - Ένας νέος φοιτητής θα αποφοιτήσει εγκαίρως?
 - Μπορούμε να αναγνωρίσουμε την ύπαρξη μιας ασθένειας από μια μαγνητική/αξονική
 - Μπορούμε να φτιάξουμε ένα νάνο- ρομπότ, το οποίο να αναγνωρίζει ιούς η καρκινικά κύτταρα?

Σύστημα



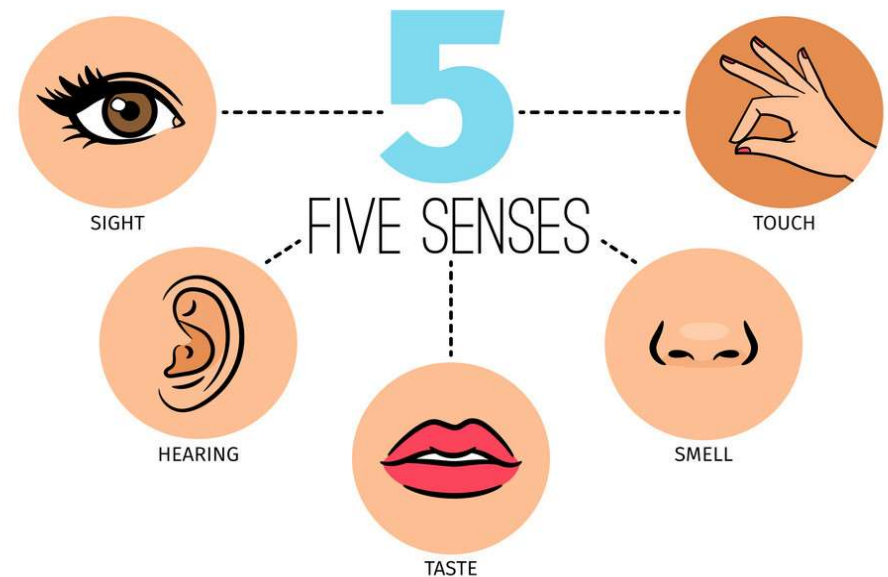
- Είσοδος, Έξοδος → **Δεδομένα**
- Επεξεργασία → **Μοντέλα**
- Ανατροφοδότηση $\leftrightarrow$ Επιβλεπόμενη, μη επιβλεπόμενη μάθηση
- Λειτουργίες: Πρόβλεψη, αναγνώριση, κατάταξη, κατανόηση γλώσσας, παραγωγή ομιλίας, ανάπτυξη στρατηγικής, κτλ.

Δεδομένα

- Ένα μοντέλο που φιλοδοξεί να είναι ευφυές θα πρέπει να έχει τη δυνατότητα να επεξεργάζεται τα δεδομένα που ένα έμβιο όν μπορεί να επεξεργαστεί δηλαδή:

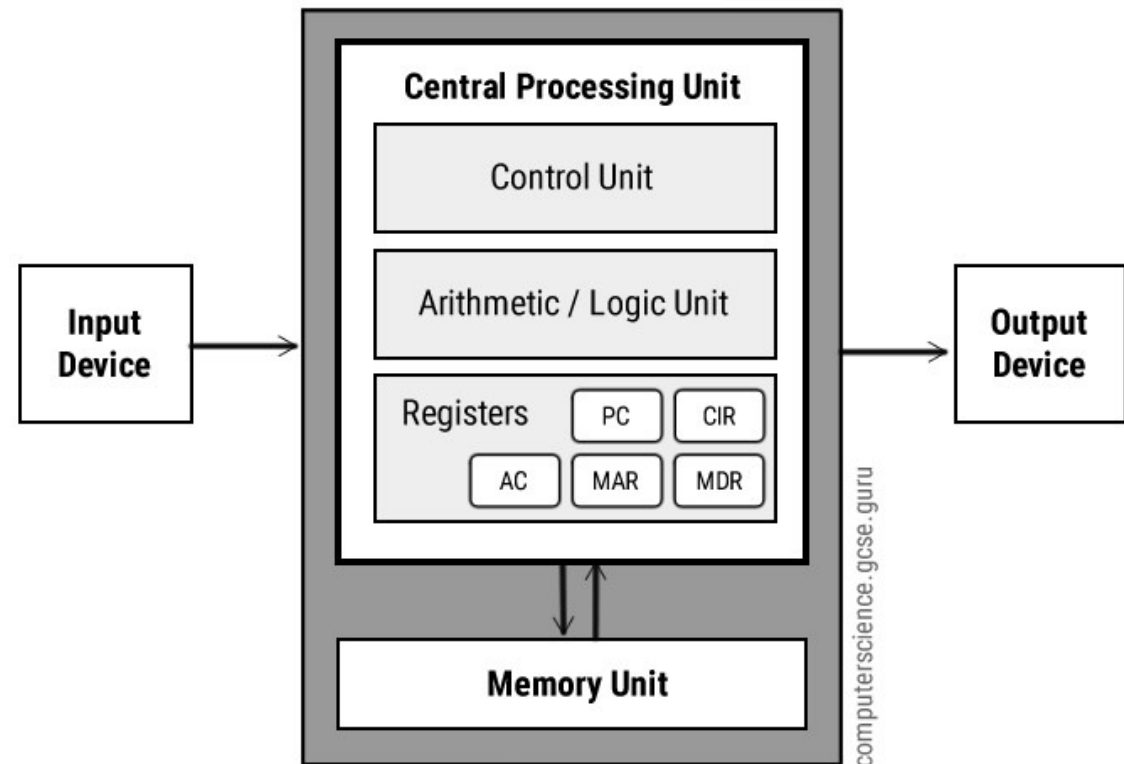
- Εικόνα
- Ήχο
- υφή
- Γεύση
- Οσμή

- Δεν υπάρχει κάποια αίσθηση για αριθμητικά δεδομένα!



Δεδομένα

- Von Neumann υπολογιστικές αρχιτεκτονικές
- Χειρίζονται μόνο Αριθμητικά Δεδομένα
- Ο Άνθρωπος δεν χειρίζεται αριθμούς!
- Ο Άνθρωπος τους Αριθμούς τους προσλαμβάνει ως εικόνες!



Δεδομένα

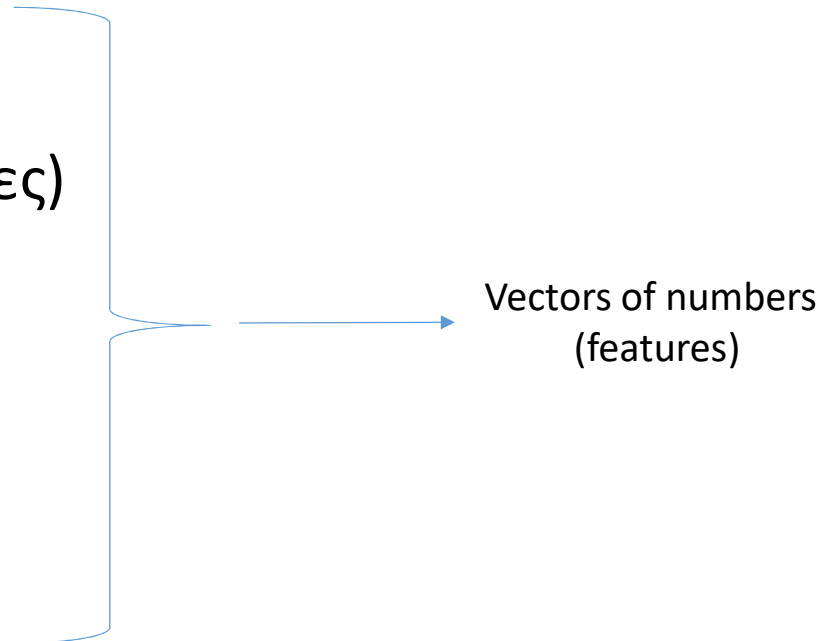
- Οι τύποι δεδομένων εισόδου εξόδου είναι συνήθως τύποι δεδομένων που επεξεργάζονται τα έμβια όντα, μέσω των αισθήσεων π.χ.
 - Αριθμοί
 - Εικόνες
 - Ήχοι
 - Αντικείμενα (σχήμα υφή)
- Επειδή, οι μηχανές χειρίζονται αριθμητικά δεδομένα θα πρέπει να υπάρξει μια αριθμητική αναπαράσταση των αντικειμένων που πρόκειται να χειριστεί το σύστημα.
- Η συνήθης αναπαράσταση είναι η **διανυσματική**. Δηλαδή μια οντότητα αναπαρίσταται με ένα διάνυσμα.
- Την αναπαράσταση αυτή την έχουμε ξαναδεί στις σχεσιακές βάσεις δεδομένων όπου μια οντότητα π.χ. φοιτητής αναπαρίσταται με ένα σύνολο από αριθμητικά χαρακτηριστικά π.χ.

Φοιτητής=[ηλικία, φύλο, εξάμηνο,]

Αναπαράσταση Δεδομένων

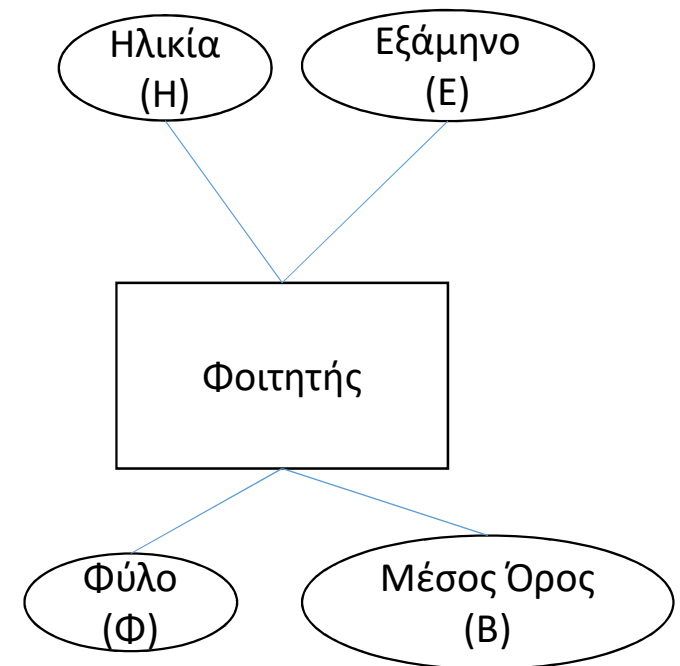
Data Representation

- Εικόνες
- Video (κινούμενες εικόνες)
- Ήχους
- Κείμενα
- Αριθμούς
- Οντότητες



Χαρακτηριστικά (Features, attributes, inputs) Διανυσματική αναπαράσταση

- Διάνυσμα χαρακτηριστικών:
 - $\mathbf{X} = [\Phi, \mathbf{H}, \mathbf{E}, \mathbf{B}] = [x_1, x_2, x_3, x_4]$
- Κάθε χαρακτηριστικό έχει ένα εύρος τιμών:
 - $x_i \in [x_{i,min}, x_{i,max}]$
- Όπως στις ΣΒΔ επιλέγουμε τα χαρακτηριστικά μιας οντότητας που θέλουμε να αποθηκεύσουμε έτσι και εδώ επιλέγουμε εκείνα τα χαρακτηριστικά που θέλουμε να αποτελέσουν τις **εισόδους** και τις **εξόδους** του συστήματος που μελετάμε.
- Τα χαρακτηριστικά που θα επιλέξουμε για την αναπαράσταση της οντότητας είναι **δική μας επιλογή**



Τα χαρακτηριστικά αποτελούν την είσοδο στο σύστημα

- Έστω ότι έχουμε $N=5$ φοιτητές. Διαμορφώνονται $N=5$ διανύσματα $m=4$, διαστάσεων έκαστο:
- $\Phi 1=[A,20,A,6.7]$
- $\Phi 2=[K,21,B,6.2]$
- $\Phi 3=[A,26,E,7.1]$
- $\Phi 4=[K,21,B,8.2]$
- Τα διανύσματα αυτά αποτελούν το σύνολο των δεδομένων:
- Οι τιμές $A=Αγόρι$, $K=Κορίτσι$ είναι σύμβολα(**nominal attributes**). Συνήθως τα αντικαθιστούμε με ακέραιους αριθμούς π.χ $A=0, K=1$. Ομοίως για το εξάμηνο(E): $A=1, B=2, Γ=3, Δ=4, E=5$
- Τελικά διαμορφώνεται ένας δισδιάστατος πίνακας $N \times m$

ID	Φ	H	E	B
Φ1	A	20	A	6.7
Φ2	K	21	B	6.2
Φ3	A	26	E	7.1
Φ4	K	21	B	8.2
Φ5	K	23	A	8.1



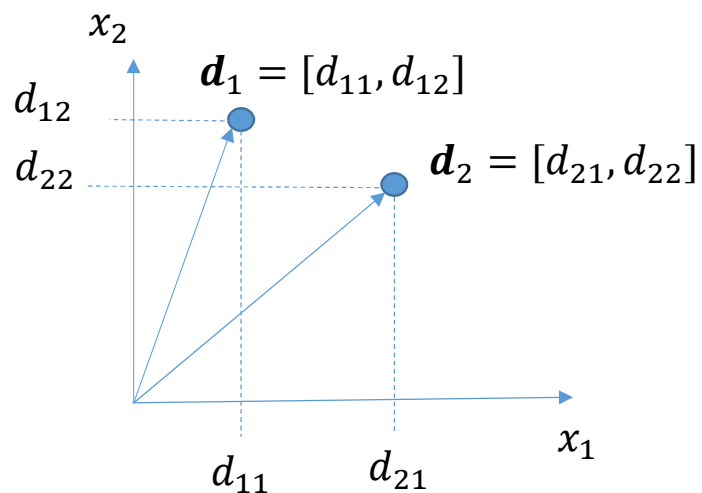
ID	Φ	H	E	B
Φ1	0	20	1	6.7
Φ2	1	21	2	6.2
Φ3	0	26	5	7.1
Φ4	1	21	2	8.2
Φ5	1	23	1	8.1

Διανυσματική αναπαράσταση

ID	Φ	H	E	B
Φ1	0	20	1	6.7
Φ2	1	21	2	6.2
Φ3	0	26	5	7.1
Φ4	1	21	2	8.2
Φ5	1	23	1	8.1

- Κάθε γραμμή του πίνακα εκφράζει ένα **διάνυσμα** ή **στιγμιότυπο** ή **δεδομένο**.
- Υπάρχουν τόσες γραμμές όσα είναι και τα δεδομένα.
- Κάθε στήλη εκφράζει ένα **χαρακτηριστικό** ή **ιδιότητα** ή **είσοδο** σε ενδεχόμενο σύστημα.
- π.χ. $H \rightarrow \mathbf{X}_2 = [20, 21, 26, 21, 23]^T$

Γεωμετρική Ερμηνεία της Διανυσματικής Αναπαράστασης



- Πράξεις μεταξύ των διανυσμάτων
- Απόσταση (metric)
- Μέτρο διανύσματος
- Γενίκευση σε πολλές διαστάσεις

	x_1	x_2
$\mathbf{d}_1 = [d_{11}, d_{12}]$		
$\mathbf{d}_2 = [d_{21}, d_{22}]$		

Αναπαράσταση Εξαγωγή χαρακτηριστικών (Feature Extraction)

- Για κάθε πρόβλημα πρέπει να γίνει εξαγωγή χαρακτηριστικών (**Feature Extraction**) προκειμένου να πραγματοποιηθεί η διανυσματική αναπαράσταση των αντικειμένων που πρόκειται να επεξεργαστούν από το σύστημα
- Για κάθε **τύπο** αντικειμένου ο τρόπος εξαγωγής των χαρακτηριστικών είναι διαφορετικός.
- Εικόνα, ήχος, χρονοσειρές

Παράδειγμα - Άσκηση

- Μια ασφαλιστική εταιρία θέλει να εκτιμήσει αν ένας νέος ασφαλισμένος θα εκδηλώσει αιφνίδιο θάνατο τα επόμενα πέντε χρόνια

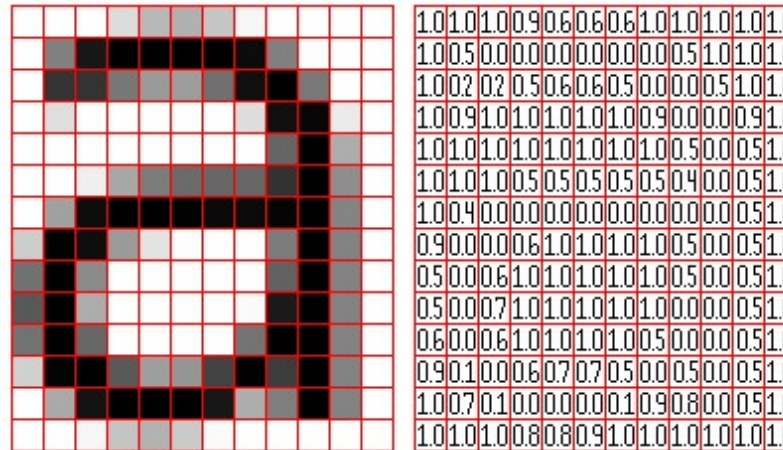
- Εκτιμά ότι οι παράγοντες κινδύνου είναι:

- Αρτηριακή πίεση
- Καρδιακός ρυθμός (παλμοί)
- Ύψος
- Βάρος
- Χοληστερίνη
- Κάπνισμα
- Φύλο
- {ΝΑΙ ΟΧΙ}

AP	KR	H	B	CH	KAP	ΦΥΛλο	y	
----	----	---	---	----	-----	-------	---	--

- Τι δεδομένα θα πρέπει να συλλέξει?
- Πως θα δημιουργήσει ένα ευφυές σύστημα?
- Πως θα μετρήσει την επίδοση/αξιοπιστία του συστήματος

Παράδειγμα - Εικόνα

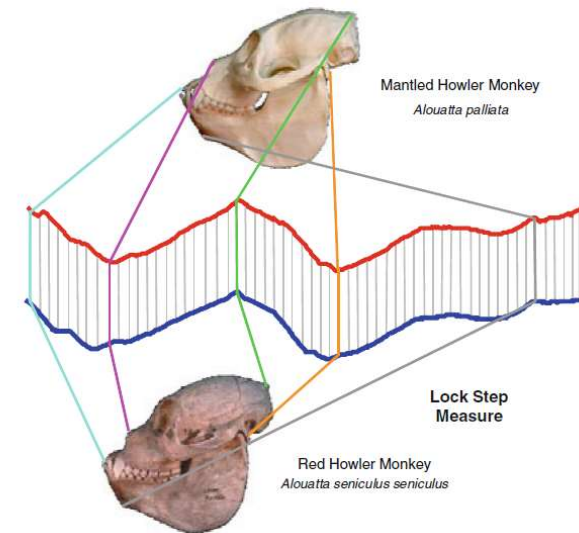


Πηγή: Vats, K. I. (2017). Automatic License Plate Detection in Foggy Condition using Enhanced OTSU Technique.

- Υπάρχουν πολλοί τρόποι αναπαράστασης
- Η αναπαράσταση των εικόνων θα πρέπει να ικανοποιεί συγκεκριμένα χαρακτηριστικά όπως:
- Αν η εικόνα περιστραφεί το διάνυσμα θα πρέπει να παραμένει το ίδιο (rotation invariant)
- Αν η εικόνα μεγεθυνθεί/σμικρυνθεί το διάνυσμα αναπαράστασης θα πρέπει να είναι το ίδιο (scale invariant) αφού πρόκειται για την ίδια εικόνα κτλ.

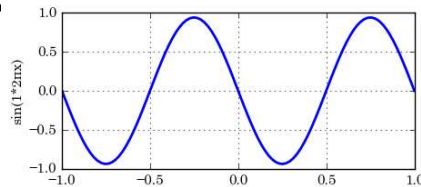
Παράδειγμα- Εικόνα

- Άλλοι τρόποι αναπαράστασης
 - Αναπαράσταση με στατιστικές ροπές
 - Μετασχηματισμός Fourier
 - Συχνότητα εμφάνισης αποχρώσεων
 - Μετασχηματισμοί περιγράμματος
 - Γεωμετρικά χαρακτηριστικά (σχετικές αποστάσεις π.χ απόσταση μάτια-στόμα, απόσταση ματιών

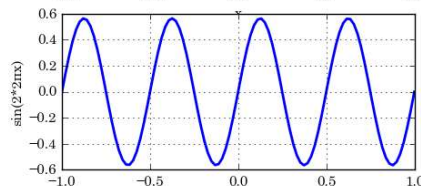


Πηγή: Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., & Keogh, E. (2013). Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26(2), 275-309.

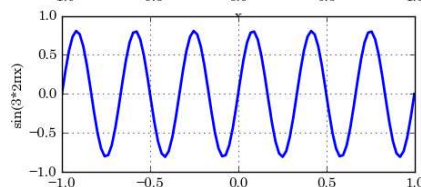
Αναπαράσταση με συναρτήσεις βάσης π.χ. Ανάπτυγμα σε σειρά



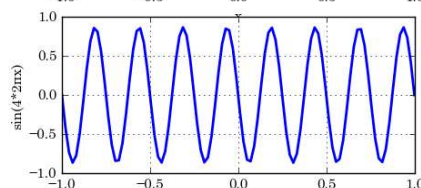
$$y_1(x) = a_1 * \sin(1 * 2\pi x)$$



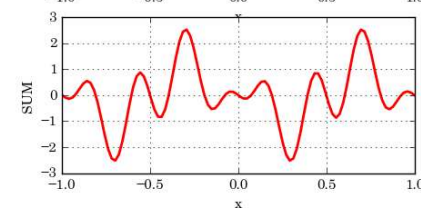
$$y_2(x) = a_2 * \sin(2 * 2\pi x)$$



$$y_3(x) = a_3 * \sin(3 * 2\pi x)$$



$$y_4(x) = a_4 * \sin(4 * 2\pi x)$$



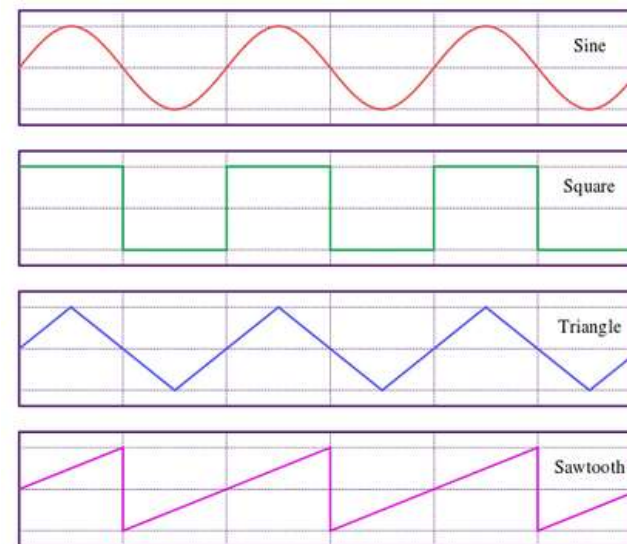
$$y(x) = y_1(x) + y_2(x) + \dots + y_4(x)$$

Μηχανική Μάθηση, Δρ. Παπαδάκης Στέλιος, Τμήμα
Διοικητικής Επιστήμης & Τεχνολογίας, Ελληνικό Μεσογειακό
Πανεπιστήμιο, e-mail: spar@hmu.gr

- Το ζητούμενο είναι η αναπαράσταση της κόκκινης καμπύλης
- Θεωρούμε ότι συνίσταται από ένα σύνολο ημιτόνων διαφορετικών συχνοτήτων
- Υπολογίζουμε τους συντελεστές $a_1, \dots, a_4$ έτσι ώστε να προσεγγίζεται όσο το δυνατόν πιο πιστά η κόκκινη καμπύλη.
- Η κόκκινη καμπύλη αναπαρίσταται ως το διάνυσμα των συντελεστών:
- $[a_1 \ a_2 \ a_3 \ a_4]$
- Στην πράξη χρησιμοποιούνται ποιο «έξυπνες» οικογένειες συναρτήσεων βάσης, π.χ. ορθογώνιες, wavelets κ.τ.λ.

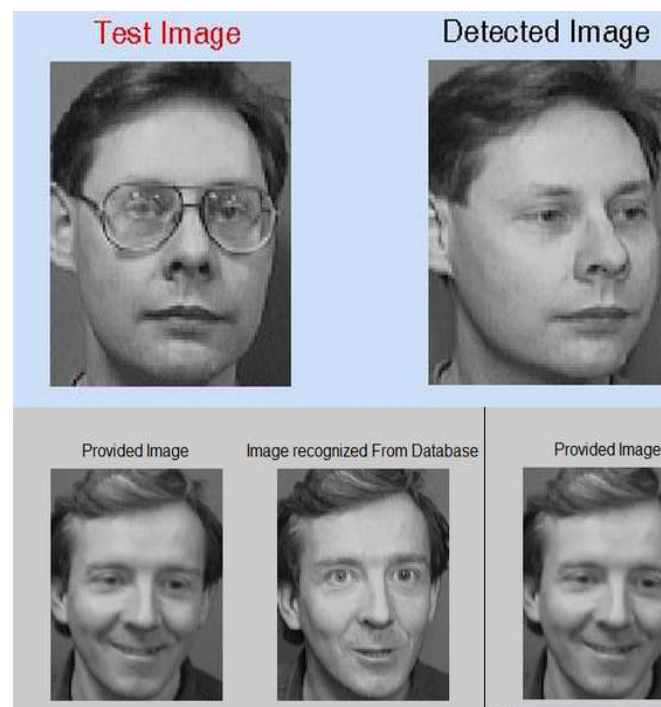
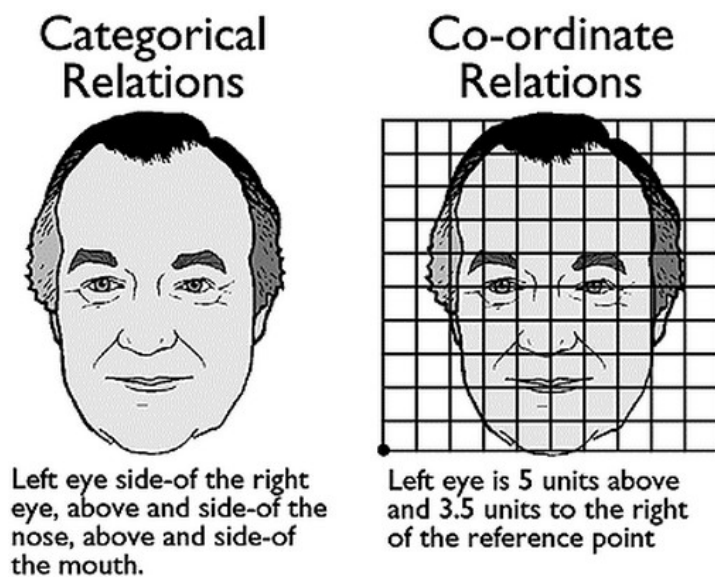
Διανυσματική αναπαράσταση κυματομορφών

- Δειγματοληψία
- Μετασχηματισμοί
 - Fourier, Laplace, wavelets
- Συναρτήσεις βάσης
- Χρονοσειρές
- Επεξεργασία σήματος



Παράδειγμα-εικόνα

- Αναγνώριση προσώπου



Εφαρμογές

- Αναγνώριση εκφράσεων προσώπου (Facial Expression Recognition)

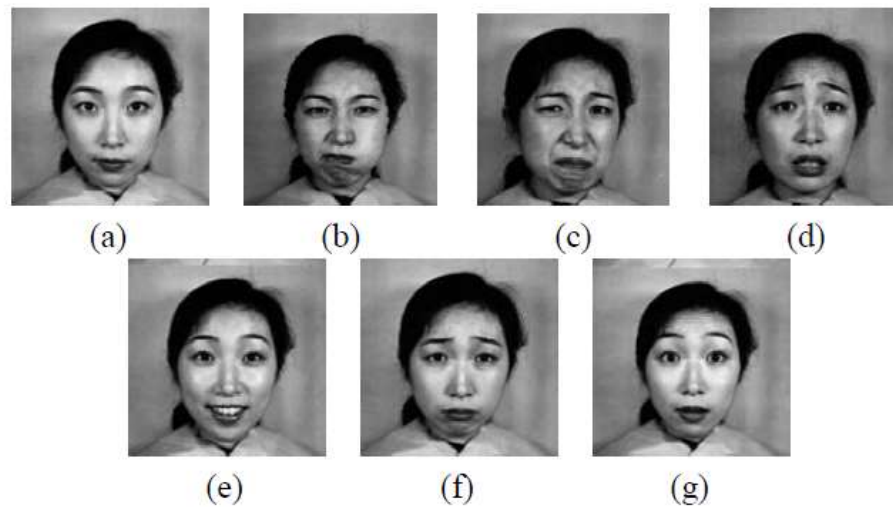


FIGURE 4

Seven different facial expressions from the *JAFFE* benchmark including (a) “neutral”, (b) “angry”, (c) “disgust”, (d) “fear”, (e) “happy”, (f) “sadness”, (g) “surprise”.

Δευτερογενείς εφαρμογές

- Ρομποτική όραση
- Νανο-ρομποτική
- Ανεύρεση επιζώντων
- Εντοπισμός τρομοκρατών
- Εκτίμηση κλινικής Εικόνας Ασθενών
-

video

- Προσθέτουμε μια ακόμη διάσταση που είναι ο χρόνος
- Ένα video είναι μια χρονοσειρά από εικόνες
- Κάθε στοιχείο της χρονοσειράς ονομάζεται frame (πλαίσιο)
- Για κάθε πλαίσιο εφαρμόζονται οι τεχνικές επεξεργασίας εικόνας
- Η παράλληλη επεξεργασία είναι σημαντικό βοήθημα καθώς κάθε πλαίσιο μπορεί να θεωρηθεί ανεξάρτητο από το άλλο στις περισσότερες εφαρμογές

Διανυσματική Αναπαράσταση κειμένου

D_1 : The quick brown fox jumped over the lazy dog.

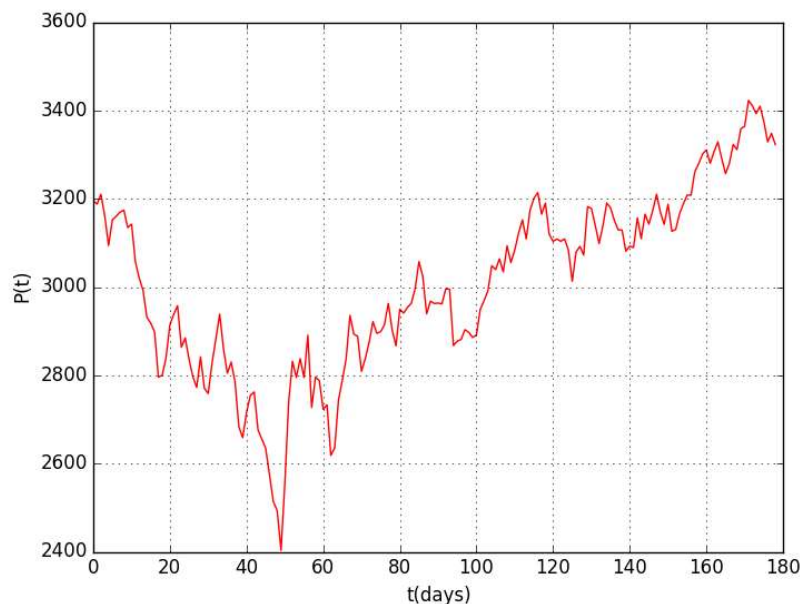
D_2 : The dog hunts a fox.

		the	quick	brown	fox	jumped	over	lazy	dog	hunts	a
D_1	[	2	1	1	1	1	1	1	1	0	0]
D_2	[	1	0	0	1	0	0	0	1	1	1]

- Παράγωγες, σύνθετες λέξεις ειδικά σε διάφορες γλώσσες όπως Ελληνικά, Γερμανικά
- Προκύπτουν αραιοί πίνακες
- Διανύσματα πολλών διαστάσεων
- Απαιτείται προ επεξεργασία
 - π.χ αφαίρεση ορισμένων συχνά εμφανιζόμενων λέξεων (The, a, an κτλ)
 - Steaming (ρίζα της λέξης)
 - Dimensionality reduction
- Συχνότητα εμφάνισης ή Εντροπία $H(S) = -\sum_{i=1}^N p_i \log_b(p_i)$

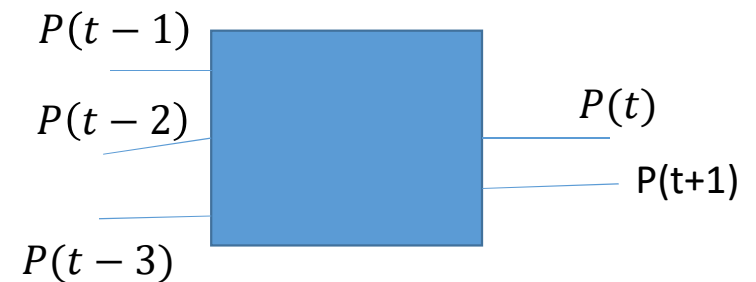
Χρονοσειρές

- $P(t)$, $t=1,2,3,\dots,N$ π.χ. Δείκτης χρηματιστηρίου



Αναπαράσταση στο πεδίο του χρόνου

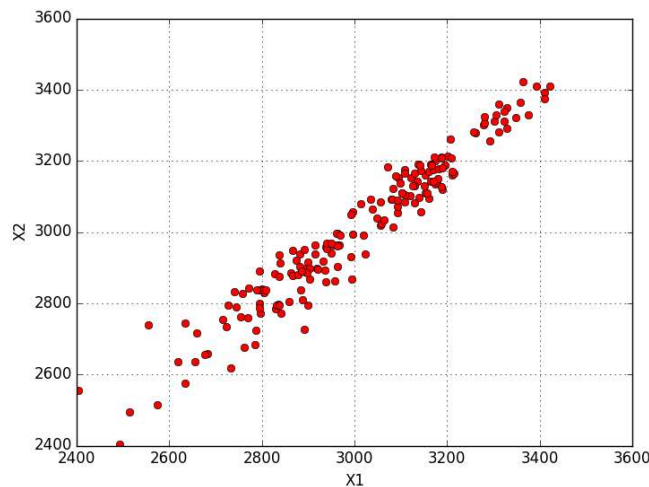
2003-01-07 3195.02
2003-01-08 3187.88
2003-01-09 3210.27
2003-01-10 3160.99
2003-01-13 3094.09
2003-01-14 3152.29
2003-01-15 3160.13
2003-01-16 3169.82
2003-01-17 3174.03



D1: 3094,3152,3160, 3169,3174

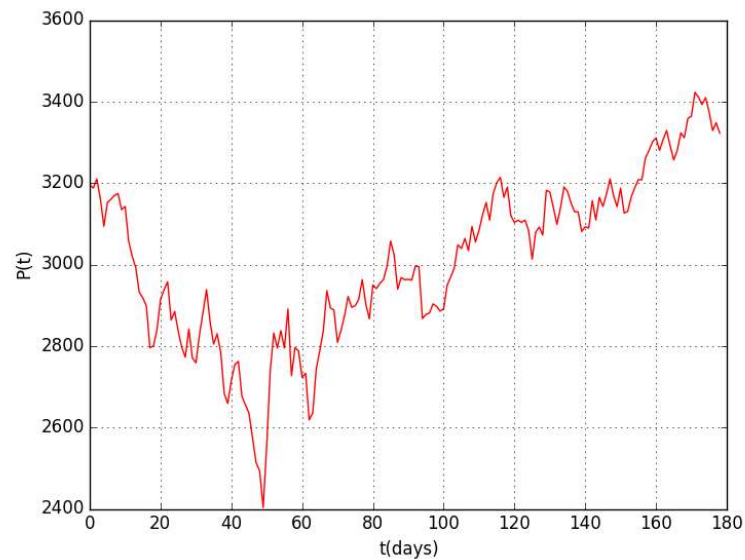
Αναπαράσταση στο πεδίο των εισόδων

- Για κάθε χρονική στιγμή
- $\mathbf{X}_t = [P(t-3), P(t-2), P(t-1), P(t)] = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{y}]$

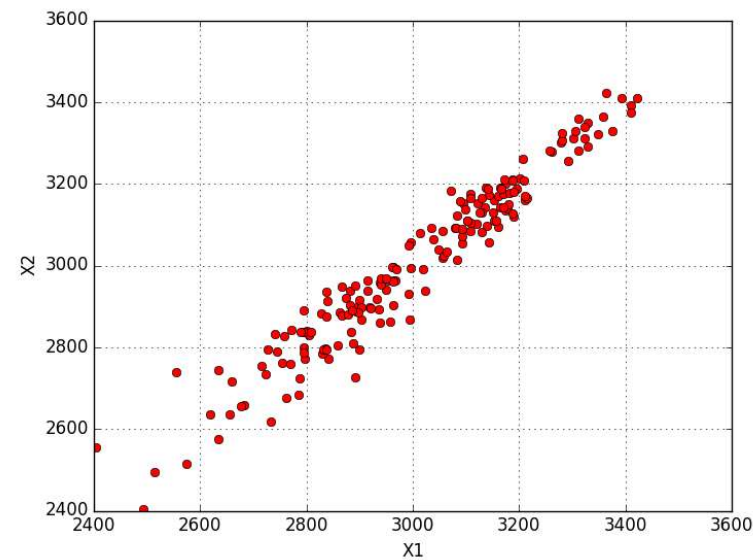


2003-01-07 [3195.02 3187.88 3210.27 3160.99]
2003-01-08 [3187.88 3210.27 3160.99 3094.09]
2003-01-09 [3210.27 3160.99 3094.09 3152.29]
2003-01-10 [3160.99 3094.09 3152.29 3160.13]
2003-01-13 [3094.09 3152.29 3160.13 3169.82]
2003-01-14 [3152.29 3160.13 3169.82 3174.03]
2003-01-15 [3160.13 3169.82 3174.03 3134.66]
2003-01-16 [3169.82 3174.03 3134.66 3142.59]
2003-01-17 [3174.03 3134.66 3142.59 3056.93]

Αναπαράσταση χρονοσειράς



Πεδίο του χρόνου
(μη στάσιμη)

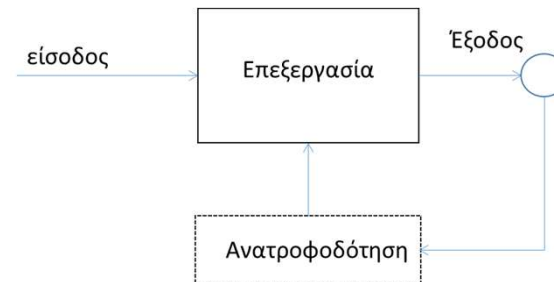
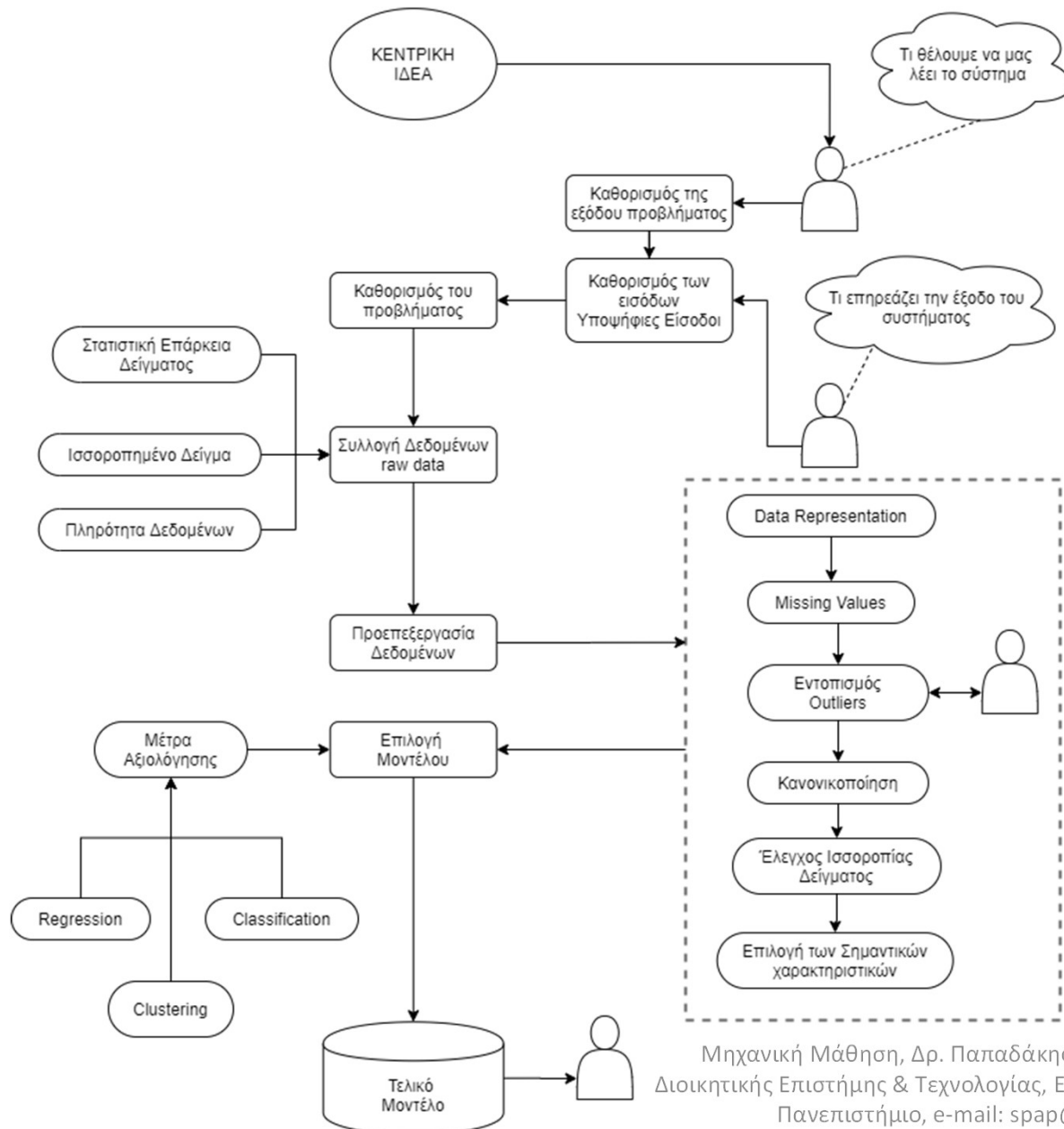


Χώρος των εισόδων
(στάσιμη)

Μειονεκτήματα της διανυσματικής αναπαράστασης

- Κάθε δεδομένο πρέπει να έχει **τον ίδιο αριθμό** συνιστωσών το οποίο δεν είναι πάντα εφικτό
 - Αν σε κάποια από τα δεδομένα δεν υπάρχουν όλες οι συνιστώσες τότε έχουμε ελλιπείς μετρήσεις (**Missing Values**)
 - Στρατηγικές αντιμετώπισης των ελλιπών μετρήσεων
- Μπορεί να σχηματιστούν διανύσματα μεγάλης διάστασης με χαρακτηριστικά αλληλοεξαρτώμενα μεταξύ τους ή πλεονάζοντα
 - Κατάρτα των πολλών διαστάσεων (**curse of dimensionality**)
 - Μείωση του πλήθους των διαστάσεων (**dimensionality reduction**)
 - Προσδιορισμός των σημαντικών χαρακτηριστικών
- εναλλακτικά:
 - Λίστες
 - Σύνολα
 - κατανομές

Βασική Διαδικασία Μοντελοποίησης



Μηχανική Μάθηση, Δρ. Παπαδάκης Στέλιος, Τμήμα
Διοικητικής Επιστήμης & Τεχνολογίας, Ελληνικό Μεσογειακό
Πανεπιστήμιο, e-mail: spap@hmu.gr

Παράδειγμα - Άσκηση

- Μια ασφαλιστική εταιρία θέλει να εκτιμήσει αν ένας νέος ασφαλισμένος θα εκδηλώσει αιφνίδιο θάνατο τα επόμενα πέντε χρόνια
- Εκτιμά ότι οι παράγοντες κινδύνου είναι:
 - Αρτηριακή πίεση
 - Καρδιακός ρυθμός (παλμοί)
 - Ύψος
 - Βάρος
 - Χοληστερίνη
 - Κάπνισμα
 - Φύλο
- Τι δεδομένα θα πρέπει να συλλέξει?
- Πως θα δημιουργήσει ένα ευφυές σύστημα?
- Πως θα μετρήσει την επίδοση/αξιοπιστία του συστήματος