

Εισαγωγή στην Μηχανική Εκμάθηση

**Τεχνικές επιβλεπόμενης και μη επιβλεπόμενης
εκμάθησης**

Δρ. Δημήτρης Καστανιώτης

<http://www.upcv.upatras.gr/personal/kastaniotis/>

Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας, της Μηχανικής
και των Μαθηματικών – STEM στην Εκπαίδευση 2017

Μηχανική Εκμάθηση και Υπολογιστική Όραση



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Στόχοι του μαθήματος

- Γνωριμία με την επιστήμη της Μηχανικής εκμάθησης
- Παρουσίαση των τεχνικών επιβλεπόμενης και μη-επιβλεπόμενης εκμάθησης
- Διαισθητική περιγραφή μέσα από παραδείγματα
- Παρουσίαση της βασικής ορολογίας και των διαθέσιμων εργαλείων

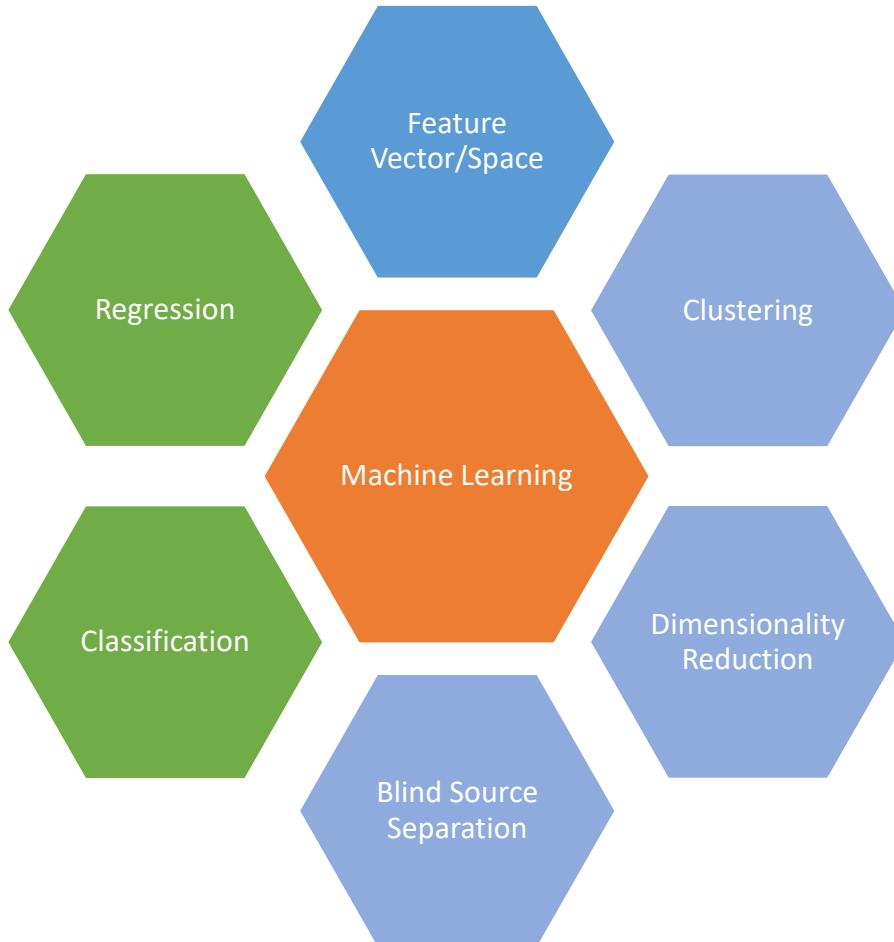
Τι θα γνωρίζετε μετά το τέλος του μαθήματος

- Θα μπορείτε να διατυπώσετε ένα πρόβλημα ως πρόβλημα μηχανικής εκμάθησης
- Θα μπορείτε να αναζητήσετε τον κατάλληλο ή τους κατάλληλους αλγορίθμους για ένα πρόβλημα
- Θα είστε σε θέση να οργανώσετε μια πειραματική διαδικασία (συλλογή δεδομένων, σχεδιασμός συστήματος, αξιολόγηση)

Τι θα γνωρίζετε μετά το τέλος του μαθήματος

- Τι είναι η μηχανική εκμάθηση (machine learning)
- Τι είναι τα χαρακτηριστικά διανύσματα (feature vector)
- Τι είναι ο χώρος των χαρακτηριστικών (feature space)
- Τι είναι η παλινδρόμηση και τι η ταξινόμηση (regression and classification)
- Τι είναι η εύρεση συστάδων στα δεδομένα (clustering)
- Τι είναι η ελάττωση διαστατικότητας (dimensionality reduction)

Τι θα γνωρίζετε μετά το τέλος του μαθήματος



Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας,
της Μηχανικής και των Μαθηματικών – STEM στην
Εκπαίδευση 2017

Μέρος 1^ο

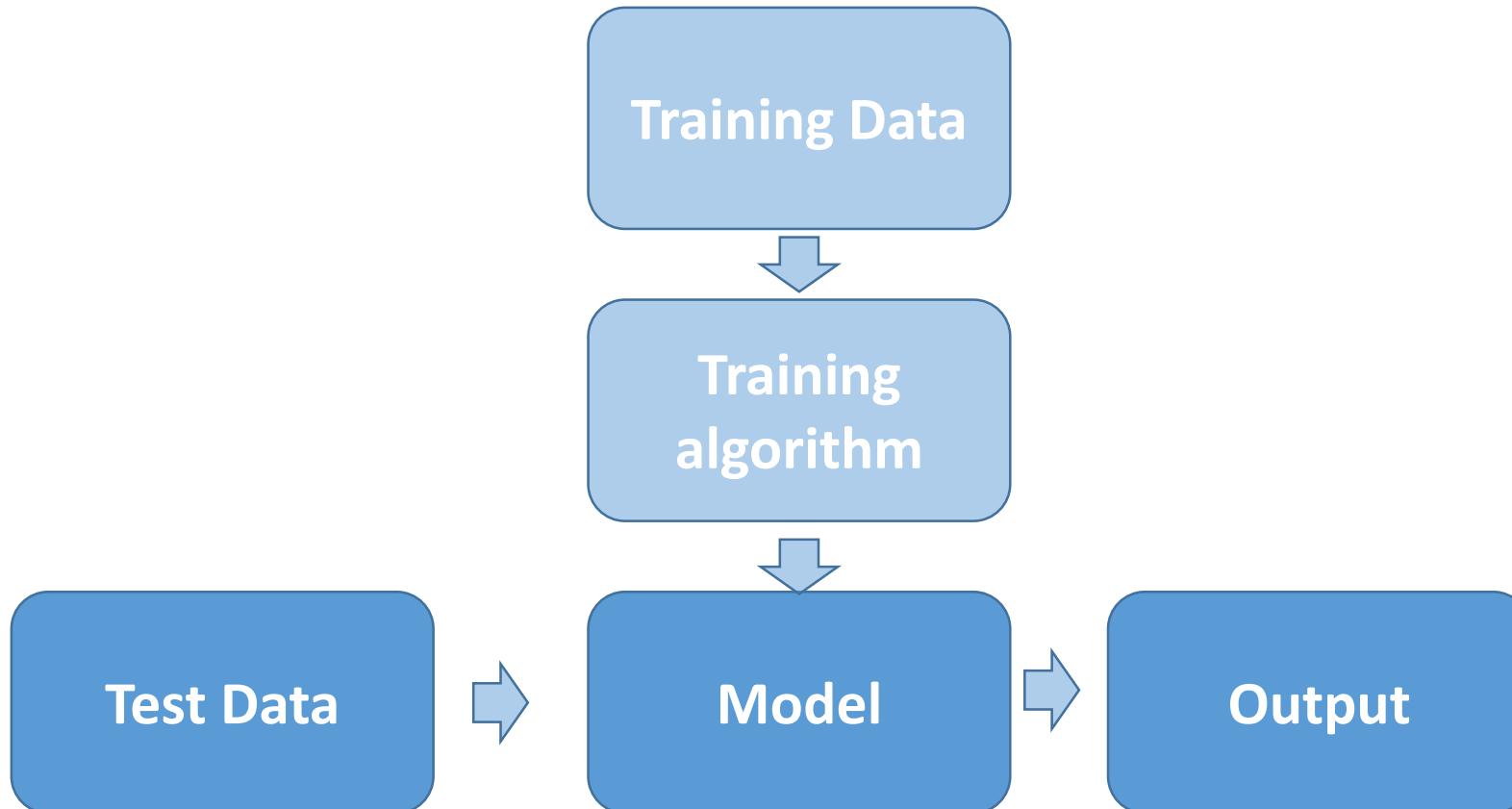
Εισαγωγή στην μηχανική εκμάθηση

Χώρος χαρακτηριστικών, Μη-επιβλεπόμενη και
Επιβλεπόμενη εκμάθησης

Τι είναι η μηχανική εκμάθηση

- Αποτελεί μια διαδικασία όπου ένα υπολογιστικό σύστημα αποκτά εμπειρία παρατηρώντας ένα σύνολο από δεδομένα
- **Το σύστημα αυτό, μετά την απόκτηση εμπειρίας είναι σε θέση να εκτελεί μια λειτουργία χωρίς να είναι σαφώς προγραμματισμένο**
- Μιμείται τη διαδικασία με την οποία ο άνθρωπος μαθαίνει να αλληλοεπιδρά με το περιβάλλον του

Τι είναι η μηχανική εκμάθηση

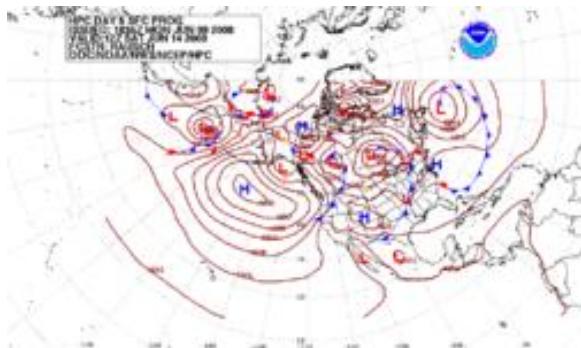


Τι είναι η μηχανική εκμάθηση

Chabots

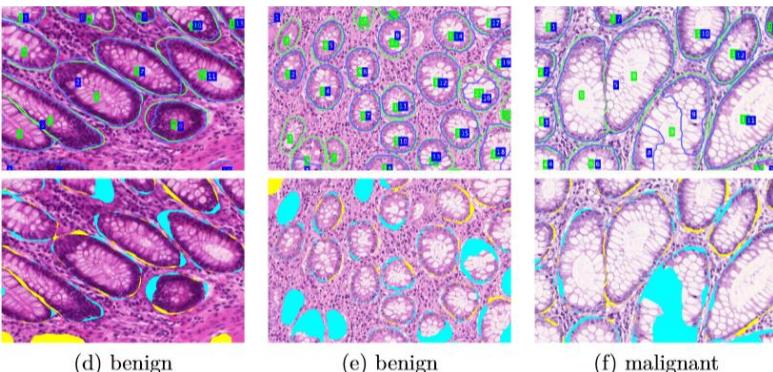
Αυτόνομα οδηγούμενα αυτοκίνητα

Πρόβλεψη καιρού



Πηγή: https://en.wikipedia.org/wiki/Weather_forecasting

Ανίχνευση παθολογίας σε ιστό

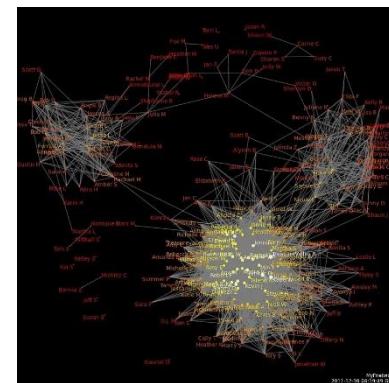


Πηγή: <https://news.developer.nvidia.com/using-cuda-and-machine-learning-to-detect-colon-cancer/>



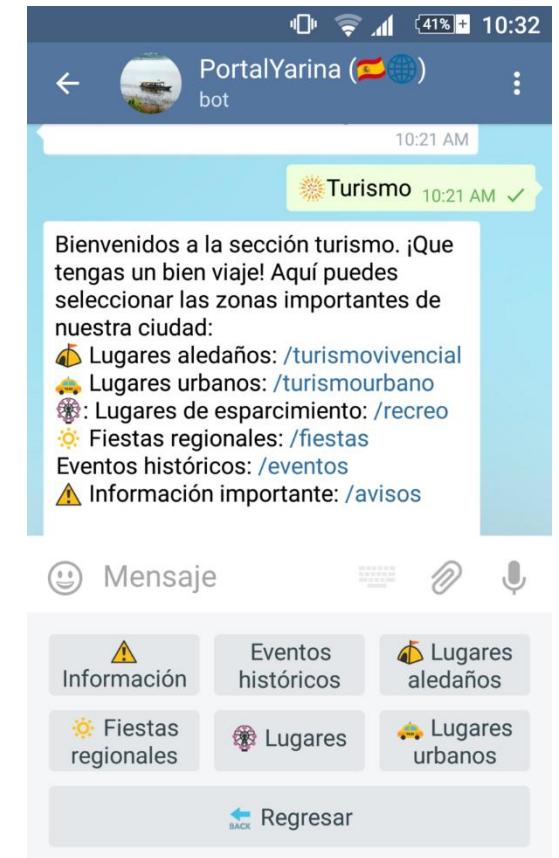
Πηγή: https://en.wikipedia.org/wiki/Autonomous_car

Social Network Analysis



https://en.wikipedia.org/wiki/Social_network_analysis

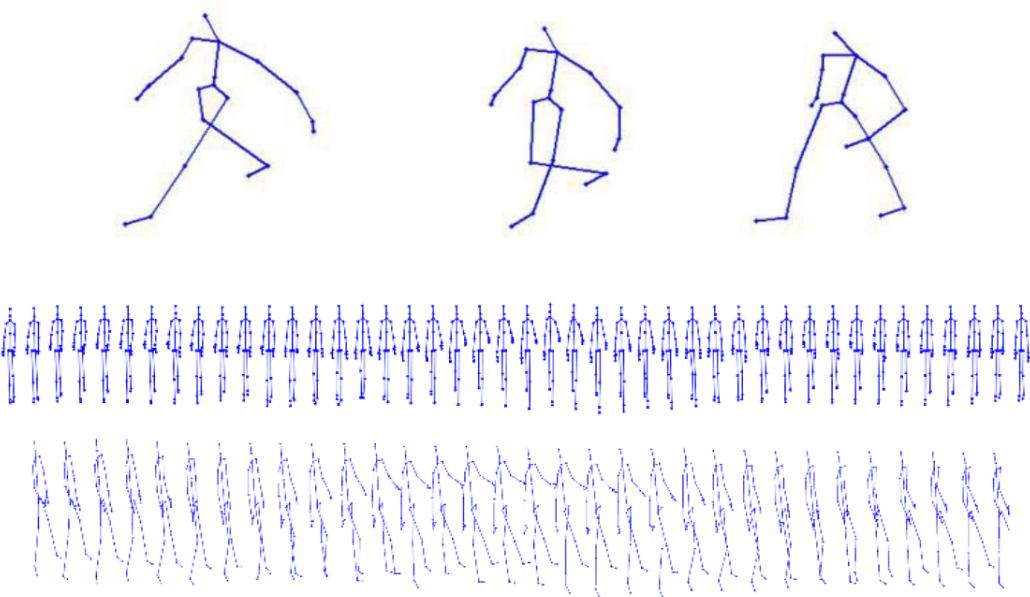
Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας,
της Μηχανικής και των Μαθηματικών – STEM στην
Εκπαίδευση 2017



Πηγή: <https://en.wikipedia.org/wiki/Chatbot>

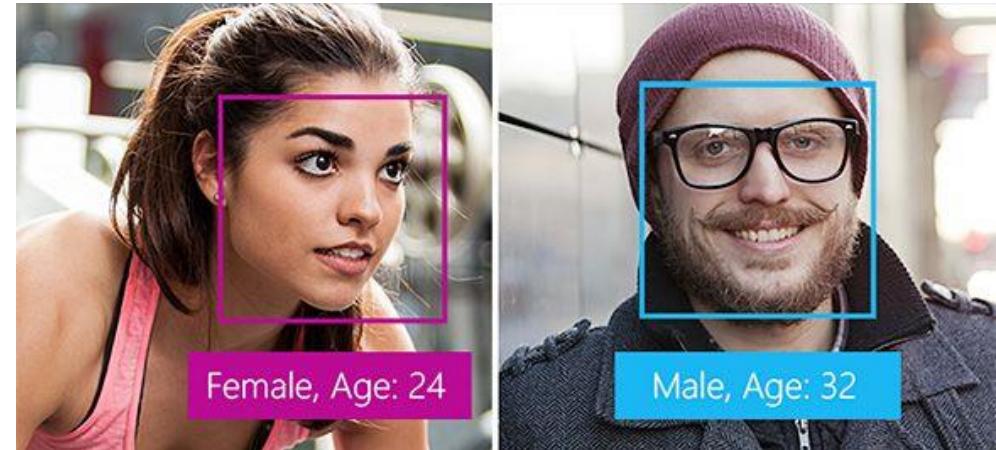
Τι είναι η μηχανική εκμάθηση

Εκτίμηση πόζας σώματος- αναγνώριση βάδισης



Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας,
της Μηχανικής και των Μαθηματικών – STEM στην
Εκπαίδευση 2017

Εντοπισμός και ανάλυση προσώπων



Πηγή: <https://docs.microsoft.com/en-us/azure/cognitive-services/face/overview#face-recognition>

Εκτίμηση πόζας προσώπου



Τι είναι η μηχανική εκμάθηση

Topic Modeling (Uncover the hidden thematic structure from a corpus of documents)

“Arts”	“Budgets”	“Children”	“Education”
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center’s share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

Figure 8: An example article from the AP corpus. Each color codes a different factor from which the word is putatively generated.



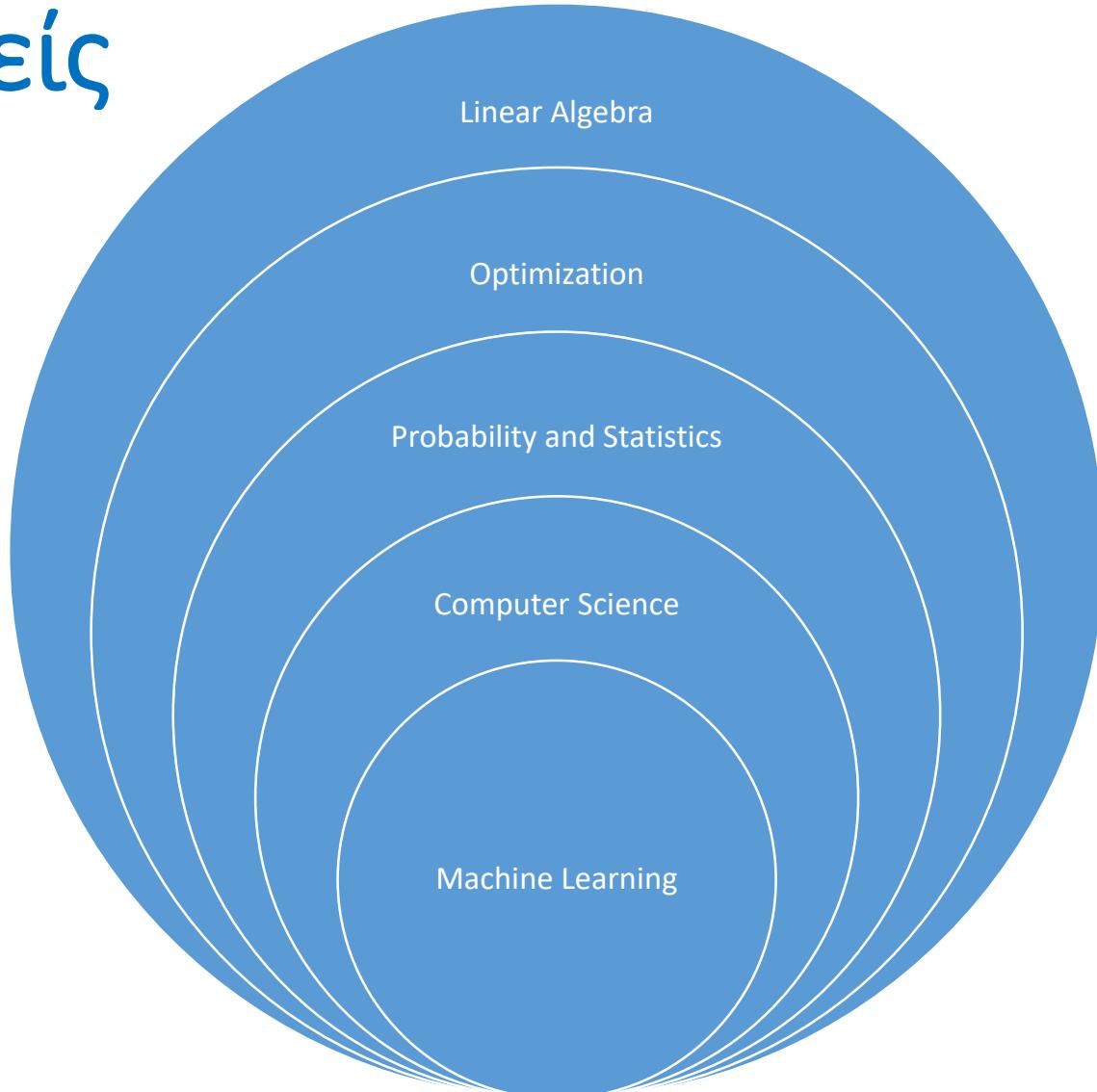
Image source: David M. Blei, et al., Latent Dirichlet Allocation, Journal of Machine Learning Research 3 (2003) 993-1022

Τι είναι η μηχανική εκμάθηση

- Speech to Text
- Text To Speech
- Language translation
- Music/Video Recommendation
- Εξόρυξη πληροφορίας από βάσεις
- Οργάνωση ψηφιακής πληροφορίας
- Spam detection

Επιστημονικοί τομείς

- Γραμμική άλγεβρα
- Πιθανότητες
- Τεχνικές βελτιστοποίησης
- Προγραμματισμός

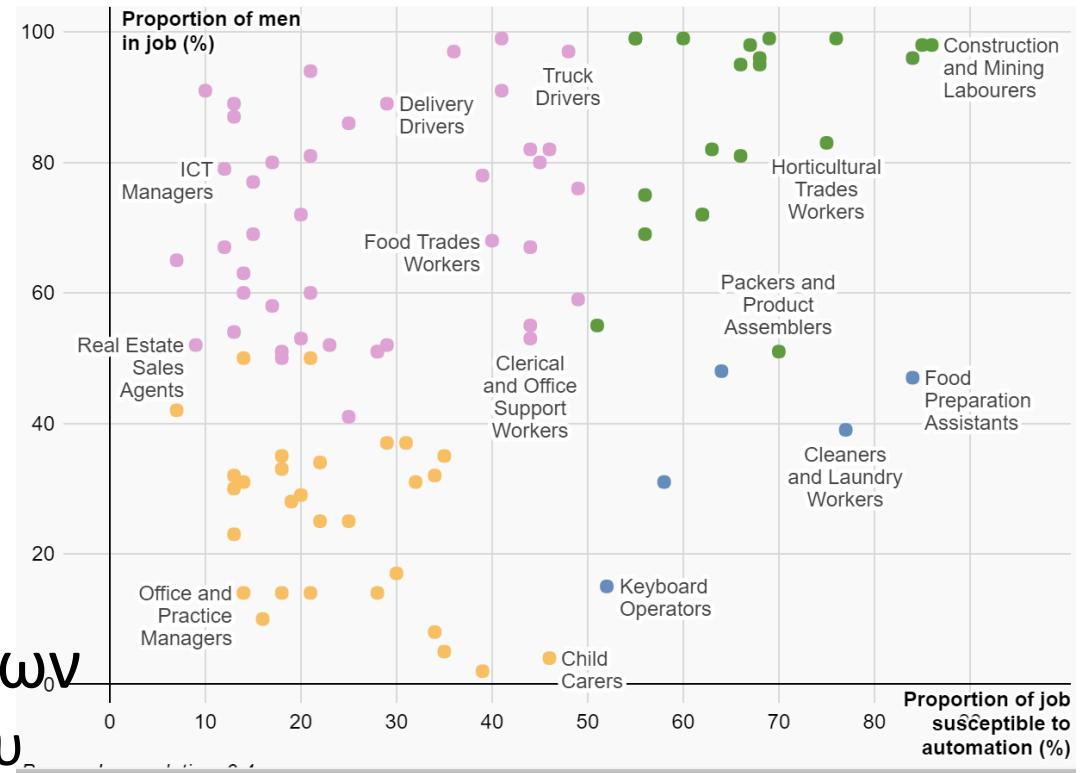


Γιατί μας ενδιαφέρει;

- Ανάλυση των δεδομένων
- Εξόρυξη πληροφορίας
- Εξασφάλιση ποιότητας
- Αυτοματοποίηση διαδικασιών

Γιατί μας ενδιαφέρει;

- Αυτοματοποίηση διαδικασιών
 - Αποτελεί το νέο ηλεκτρισμό
- Επαγγελματικός προσανατολισμός
 - Πρέπει να γνωρίζουμε εάν και πόσο θα επηρεάσει διάφορα επαγγέλματα
- Δυνατότητες αξιοποίησης των εργαλείων
 - Ισχυρό εργαλείο στα χέρια του σύγχρονου ανθρώπου (!)

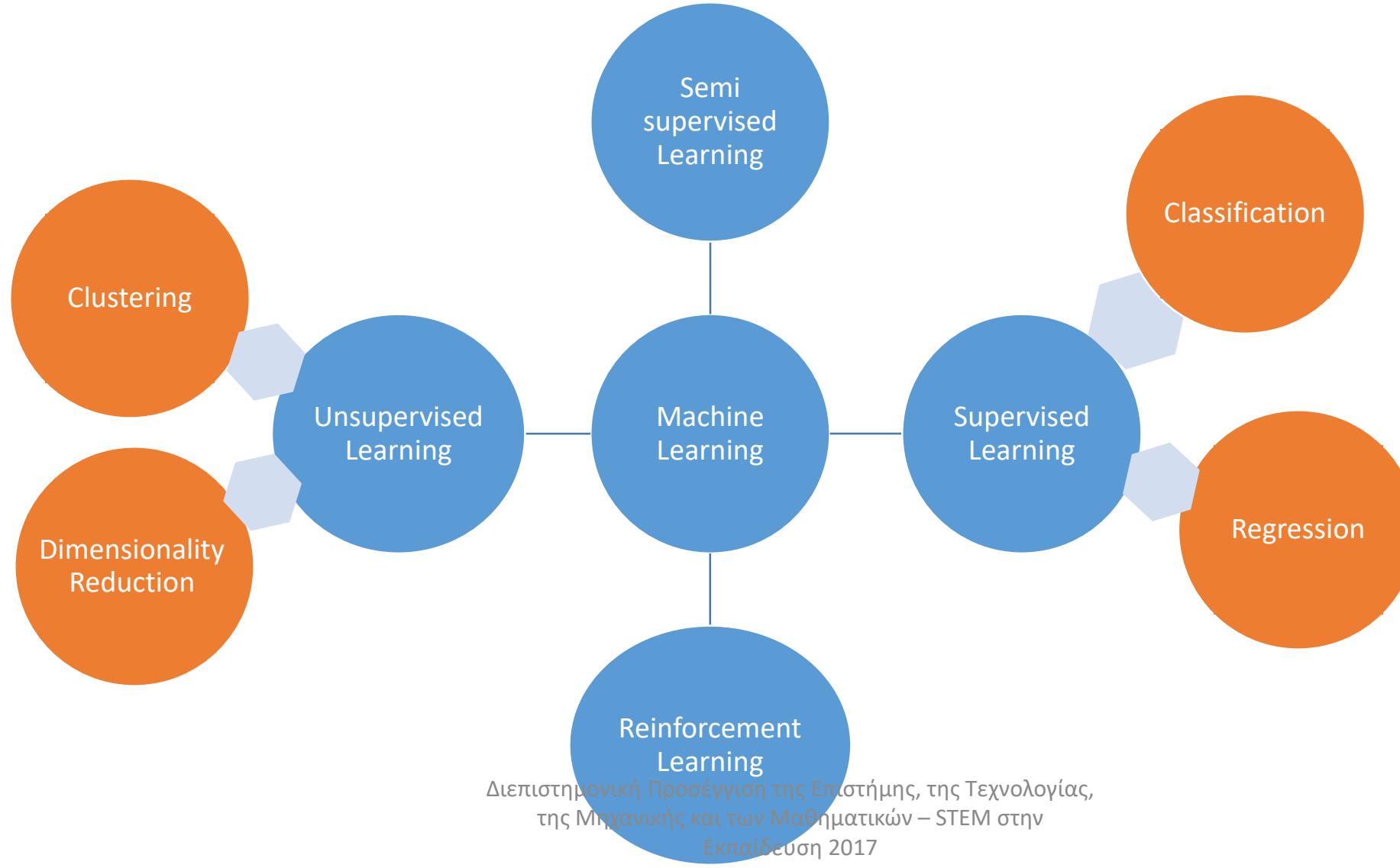


<http://www.abc.net.au/news/2017-08-09/ai-automation-men-and-lower-paid-workers/8741518>

Βασικοί τύποι μηχανικής εκμάθησης

- **Με επιβλεπόμενη (unsupervised)**
 - Τα δεδομένα παρέχονται χωρίς κάποια επιπλέον πληροφορία (χαρακτηρισμό)
 - Ζητούμενο εδώ είναι η εύρεση δομής στα δεδομένα
- **Επιβλεπόμενη (supervised)**
 - Τα δεδομένα παρέχονται ως ζεύγη εισόδου- εξόδου
 - Ζητούμενο εδώ είναι η εύρεση της συνάρτησης που αντιστοιχεί τα διανύσματα εισόδου στα διανύσματα εξόδου
- **Ημίεπιβλεπόμενη (semisupervised)**
 - Λίγα δεδομένα με χαρακτηρισμό και αρκετά δίχως
- **Εξαναγκασμένη εκμάθηση (reinforcement learning)**
 - Κάθε καλή πράξη επιβραβεύεται

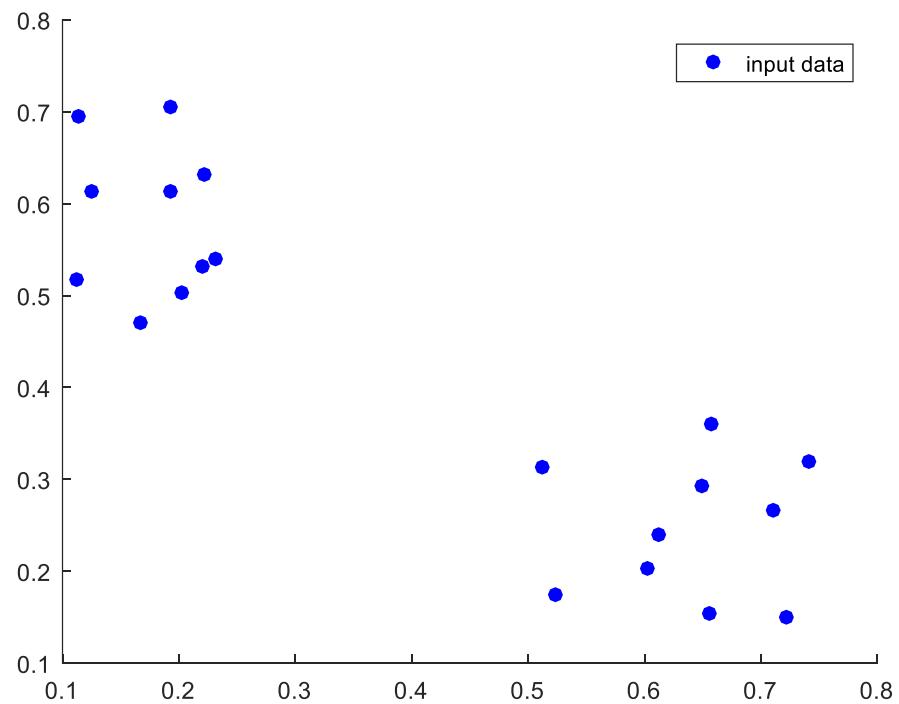
Βασικοί τύποι μηχανικής εκμάθησης



Βασικοί τύποι μηχανικής εκμάθησης

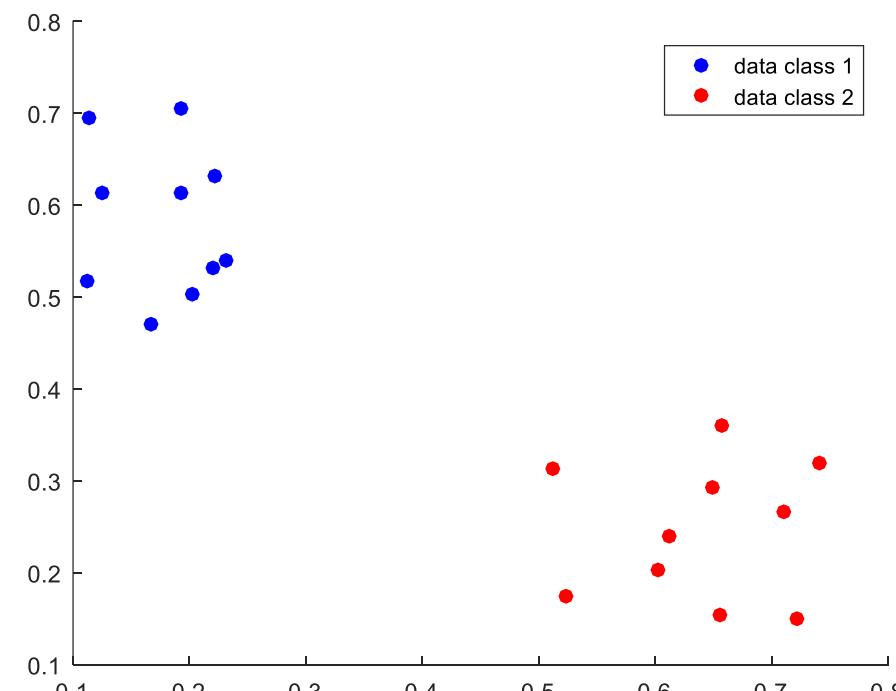
Μη επιβλεπόμενη

Δεν υπάρχει χαρακτηρισμός για τα δεδομένα



Επιβλεπόμενη

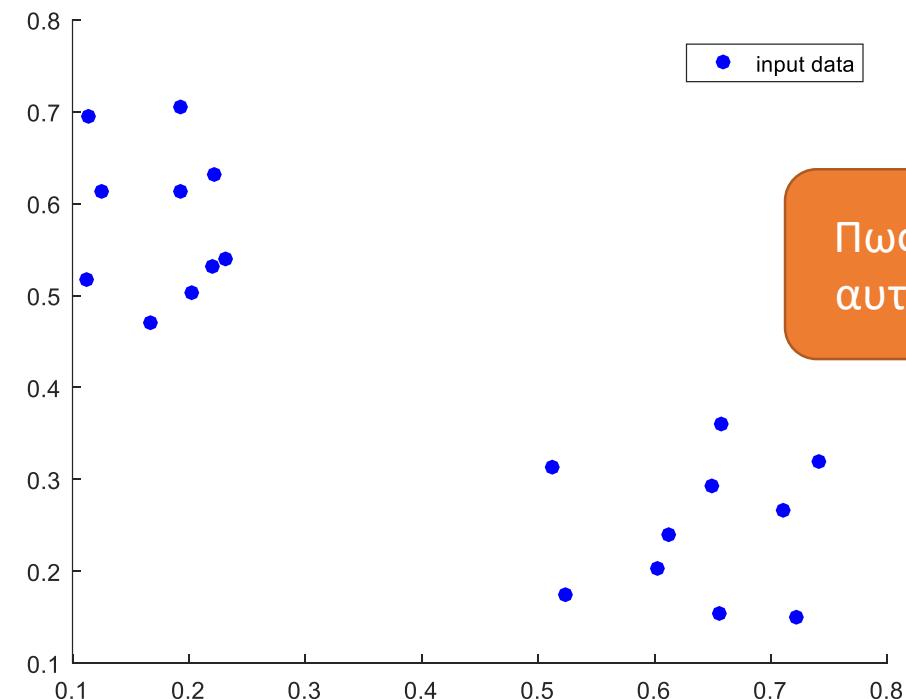
Υπάρχει χαρακτηρισμός για τα δεδομένα



Βασικοί τύποι μηχανικής εκμάθησης

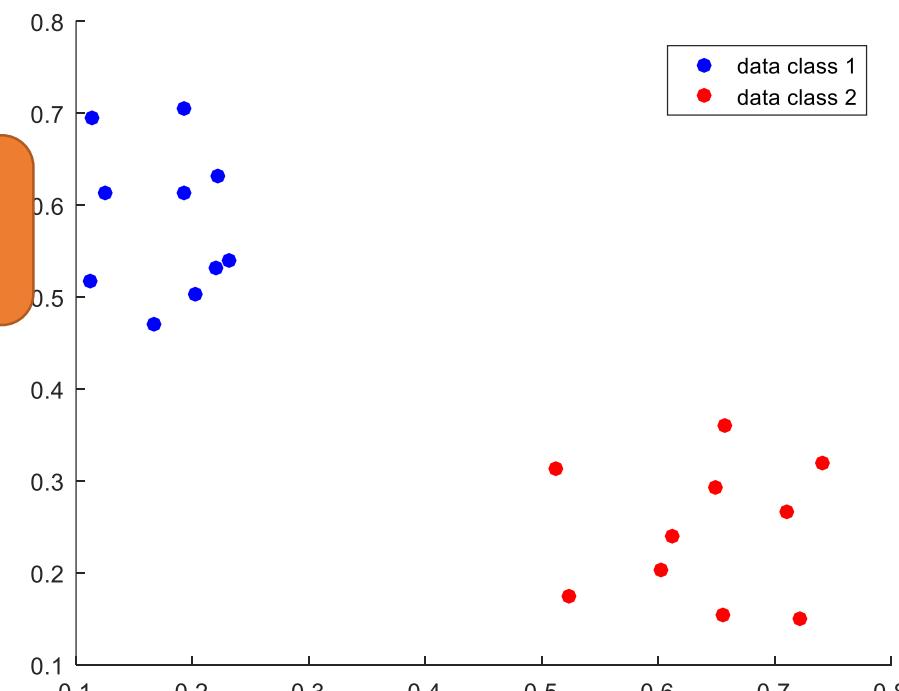
Μη επιβλεπόμενη

Δεν υπάρχει χαρακτηρισμός για τα δεδομένα



Επιβλεπόμενη

Υπάρχει χαρακτηρισμός για τα δεδομένα



Πως προκύπτουν
αυτά τα σημεία;;

Χώρος των χαρακτηριστικών- Χαρακτηριστικά διανύσματα

Δείγματα- Χαρακτηριστικά διανύσματα

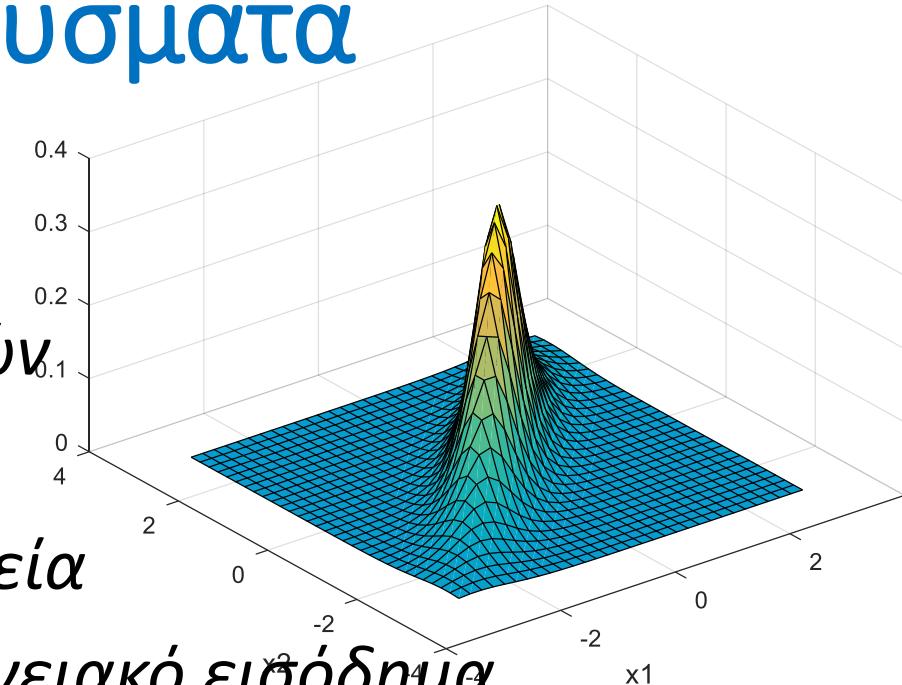
- Δείγματα:

Είναι το σύνολο των παρατηρήσεων/καταγραφών

- **Χαρακτηριστικό διάνυσμα**

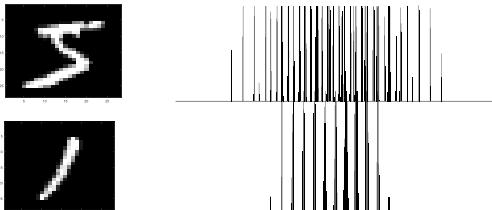
Ένα διάνυσμα που ενσωματώνει διάφορα στοιχεία του δείγματος (π.χ. βαθμολογία μαθητή, οικογενειακό εισόδημα, αριθμός ξένων γλωσσών, ηλικία γονέων)

- **Τι προσπαθούμε να επιτύχουμε με το διάνυσμα αυτό;**



Περισσότερα για τα χαρακτηριστικά διανύσματα

- Δεδομένα εισόδου- Διανύσματα εισόδου- Εξαγωγή χαρακτηριστικών
 - Δημιουργούν τον χώρο των χαρακτηριστικών
 - Πως προκύπτουν;



Τα εικονοστοιχεία χρησιμοποιούνται ως χαρακτηριστικά

Τι κερδίζουμε με την εξαγωγή χαρακτηριστικών;

Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας, της Μηχανικής και των Μαθηματικών – STEM στην Εκπαίδευση 2017

Συχνότητα εμφάνισης λέξεων

Αναγνώριση βάσης με μεραρχικά συσσωρευμένους περιγραφικούς [85]

κάθε καρέ κατεγραφής να έχουμε μια τοπική συμπαράσταση της βάσημασης όχι ως προς το χώρο αλλά το χρόνο. Ένας τρόπος για να εξχειρίσει την πληροφορία αυτή είναι να αξιοποιηθεί η τροχιά επιλεγμένων μελών κατά τη βάση. Πιο συγκεκριμένα, η απαιχούσα από το χώρο της πόλεως στο χώρο των ιστογραμμάτων και η διατήρηση της τοπικότητας είναι εφικτό να οδηγήσουν σε πολύ σποδιακές περιγραφές κανές να αντιτεθέλουν στη λεπτομερεία που απαιχούται για το πρόβλημα της αναγνώρισης βάσισης.

Δεύτερον, η κατανομή των τοπικών ιστογραμμάτων μπορεί να συνοψίσεται σε ένα καθολικό περιγραφή. Βασιζόμαστε σε προδιαρχικά ευρήματα [Legou10], [Ferronini10], [Zhou10], [Reddy14] που δείχνουν ότι οι αναπαραστάσεις αυτές, απόδοσην εξαιρετικά σε ένα πλήθος διαφορετικών εφεργάνων. Πιο συγκεκριμένα, στην παρούσα εργασία παρουσιάζουμε μια διαδικασία κωδικοποίησης η οποία χαρακτηρίζεται από την ικανότητα της να παρχει υψηλού επιπέδου πληροφορίας [Kastaniotis16c], [Kastaniotis13]. Πιο συγκεκριμένα, η διαδικασία αυτή συνδέεται τα σφύλια των αριθμών ανεπιφαντάσιμων μαζί με τις τερνίκιες συσπερρευστές υπολογιστή, οδηγώντας σε μια διανυσματική αναπαράσταση με μεγάλη διακρισιμότητα.

Πρόχειται, μέχρι στιγμής, ένα πλήθος εφεργάνων έχει προσπαθήσει να ενισχύσει τις τερνίκιες δημιουργίες μεσάνες επιπέδου περιγραφών με μοντέλα εκμήδησης λεκός και σράζες αναπαράστασης, τα οποία είναι κατά τα άλλα συλλόγουν καλύτερα την πληροφορία της πολλαπλότητας των δεδουλεύματων. Οι τερνίκιες που χρησιμοποιούν συντελετές αραιής αναπαράστασης, βασίζονται με μια διαδικασία «επλογής» χωρίς ώστε να οδηγήσουν σε μια καθολική περιγραφή της πληροφορίας του σήματος που αναδύονται [YangJianchao09]. Ωστόσο μέσα από τη διαδικασία επλογής όλη η πληροφορία κατανοείται χάνεται. Για το λόγο αυτού η συγγραφέας στην εργασία [Reddy14] πρότεινεν μια μέθοδο στην οποία η τεχνική VLAD [Legou10] θα λειτουργήσει στο χώρο της αραιής αναπαράστασης. Η μέθοδος αυτή ωστόσο παρουσιάζει κάποιους σημαντικούς περιορισμούς. Πρώτον, χρησιμοποιεί την μέθοδο k-means για να μάθει ένα λεκό στο χώρο των αριθμών συντελετών. Δεύτερον, ταξινομεί τους συντελετές αυτούς σε ένα μέλλοντερο κέντρο χρησιμοποίωντας την Euclidean διαστάση. Τρίτον, δεν λαμβάνει άμεσα υπόψη την εγγενή γεωμετρία των αριθμών συντελετών καθώς και το μοντέλο αναφοράς τους. Επίσης η διαδικασία αυτή δεν λέγεται ότι αξιονύμη με απεκίνηση χωρίς όμως να αξιονύμεται τη γεωμετρία της απεικόνισης αυτής.

Σε μια άλλη προσέγγιση, οι Long και [Long15] βελτίωσαν τη διαδικασία κωδικοποίησης του VLAD προτείνονταν την εκμήδηση ενός λεκόν το οποίο θα βρισκόταν στο πρόβλημα της ταξινόμησης. Επίσης ενισχύοντας το μοντέλο του VLAD προσθέτοντας στατιστικές, ροτέιν υψηλότερης τάξης. Σε τέτοιο, αυτές οι προσθήμες, δι-

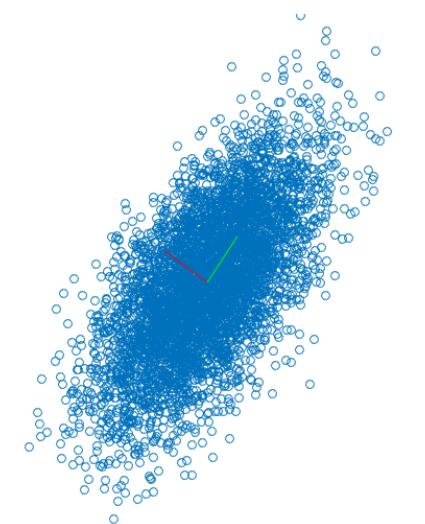
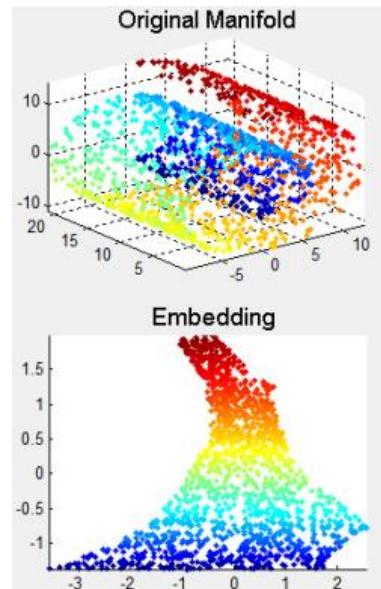
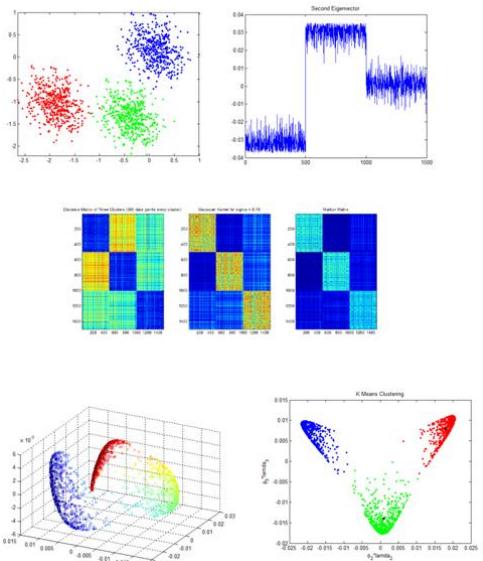
Περισσότερα για τα χαρακτηριστικά διανύσματα

- Τα χαρακτηριστικά μπορεί να παρέχονται/εξάγονται σε:
 - *Πίνακες*
 - *Διανύσματα*
 - *Γράφους*
 - *Εικόνες*
 - *Ακολουθίες εικόνων/ video*
 - *Χαρακτήρες*
 - *Πολυδιάστατα κανάλια*
 - *Μονοδιάστατά σήματα*

Μη-επιβλεπόμενη μηχανική εκμάθηση

Μη-Επιβλεπόμενη εκμάθηση

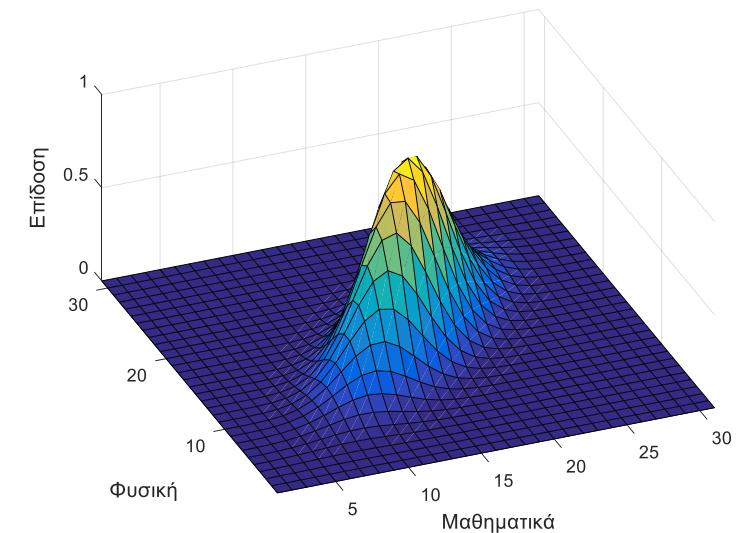
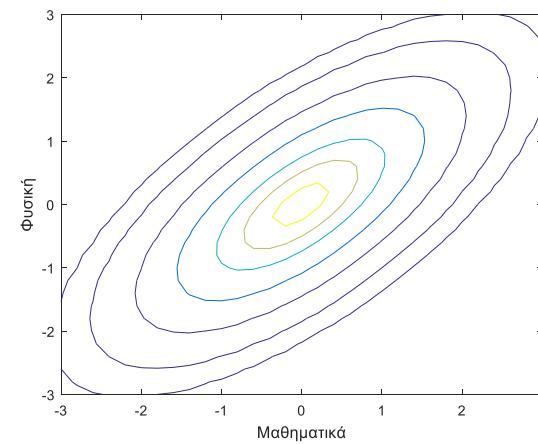
- Εύρεση δομής στα δεδομένα
- Μας ενδιαφέρει η σχέση μεταξύ των δειγμάτων



Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας,
της Μηχανικής και των Μαθηματικών – STEM στην
Εκπαίδευση 2017

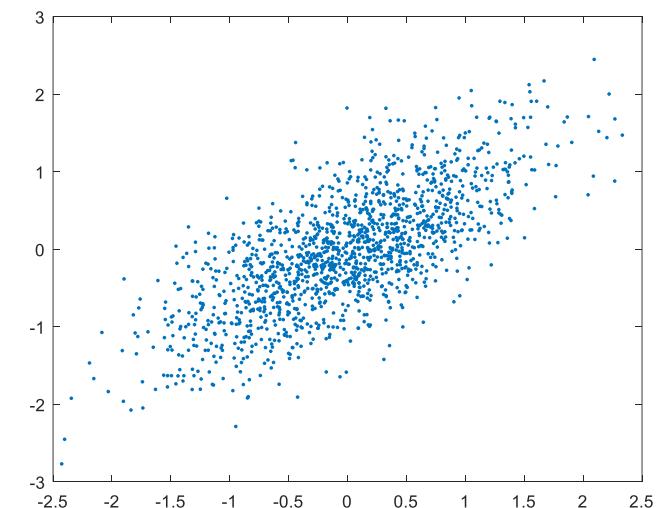
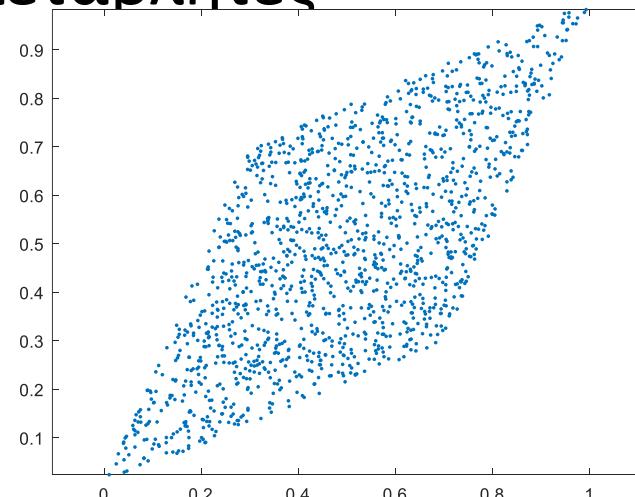
Μη επιβλεπόμενη εκμάθηση

- Ελάττωση διαστατικότητας
- **Κάποια χαρακτηριστικά μπορεί να είναι περιττά**
 - Συσχέτιση → Η γνώση για ένα χαρακτηριστικό μας δίνει πληροφορία για ένα άλλο χαρακτηριστικό
- Στο γράφημα δεξιά οι δύο μεταβλητές είναι συσχετισμένες



Μη επιβλεπόμενη εκμάθηση

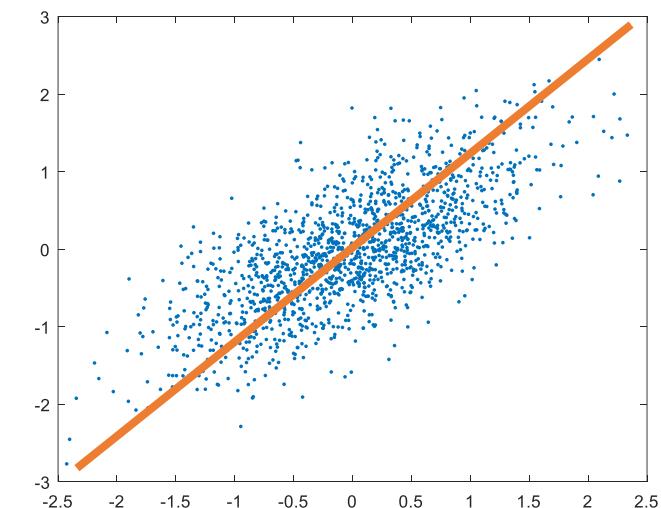
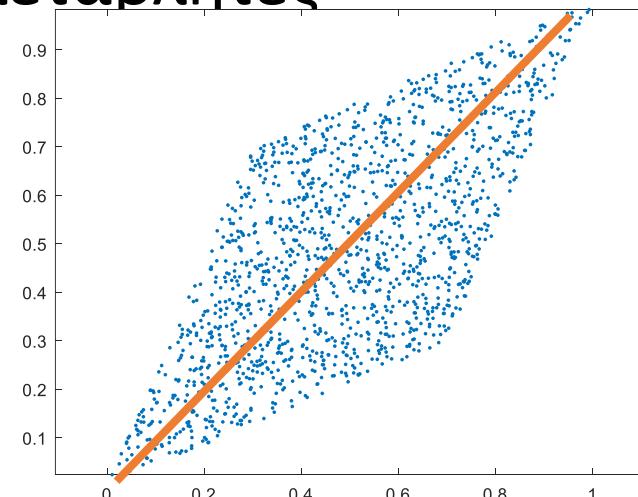
- Ελάττωση διαστατικότητας
- **Κάποια χαρακτηριστικά μπορεί να είναι περιττά**
 - Συσχέτιση → Η γνώση για ένα χαρακτηριστικό μας δίνει πληροφορία για ένα άλλο χαρακτηριστικό
- Στο γράφημα δεξιά οι δύο μεταβλητές είναι συσχετισμένες



Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας,
της Μηχανικής και των Μαθηματικών – STEM στην
Εκπαίδευση 2017

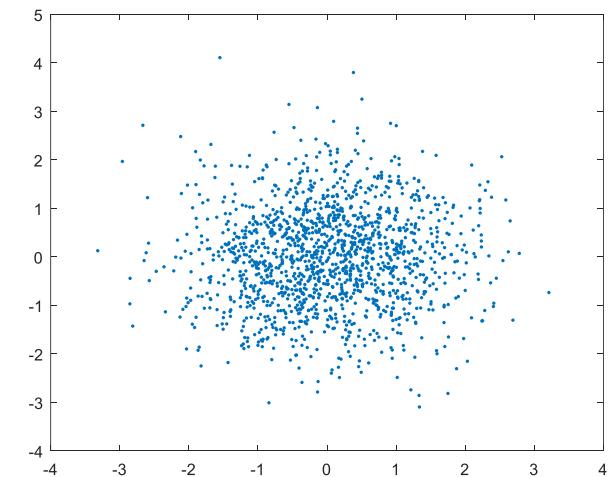
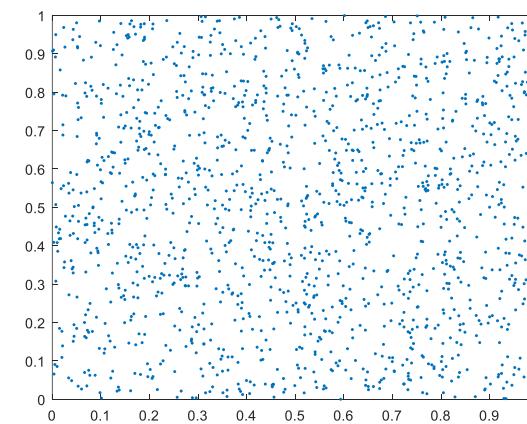
Μη επιβλεπόμενη εκμάθηση

- Ελάττωση διαστατικότητας
- **Κάποια χαρακτηριστικά μπορεί να είναι περιττά**
 - Συσχέτιση → Η γνώση για ένα χαρακτηριστικό μας δίνει πληροφορία για ένα άλλο χαρακτηριστικό
- Στο γράφημα δεξιά οι δύο μεταβλητές είναι συσχετισμένες



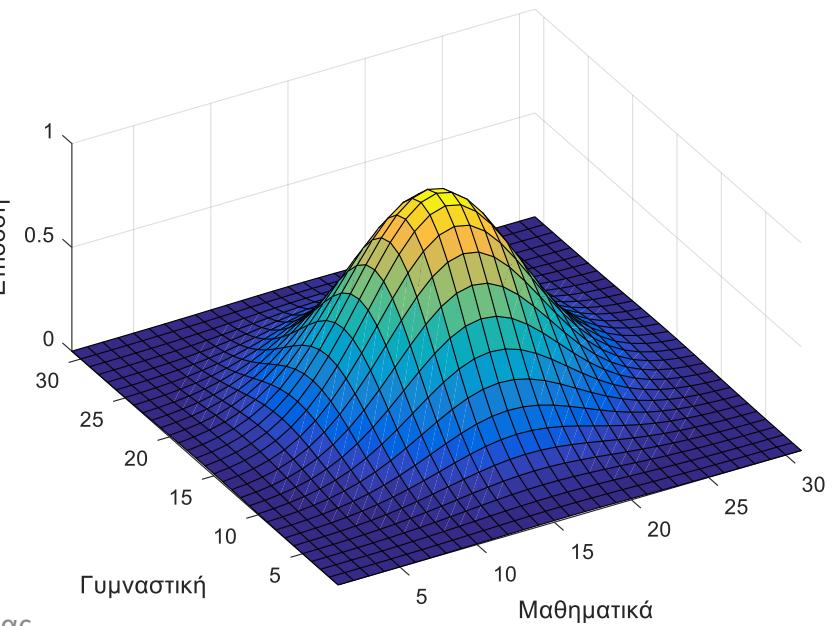
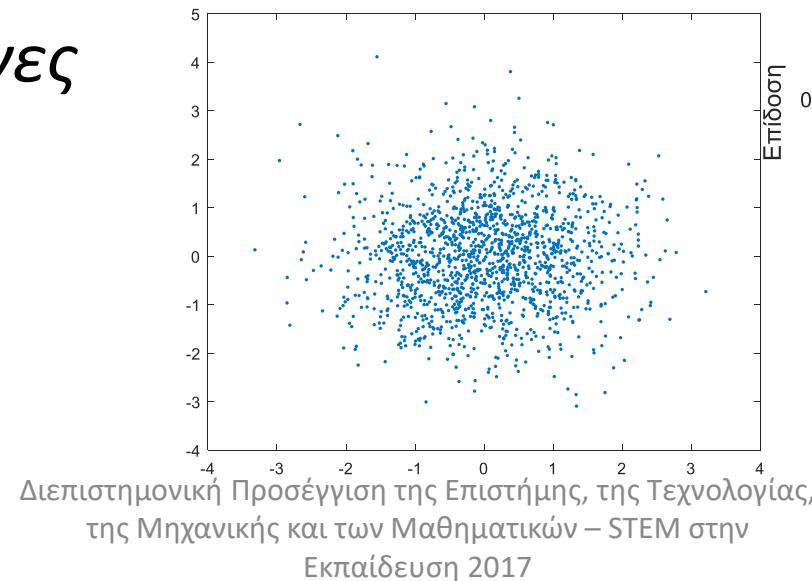
Μη επιβλεπόμενη εκμάθηση

- Ελάττωση διαστατικότητας
- **Κάποια χαρακτηριστικά μπορεί να είναι περιττά**
 - Συσχέτιση → Η γνώση για ένα χαρακτηριστικό μας δίνει πληροφορία για ένα άλλο χαρακτηριστικό
- Στο γράφημα δεξιά οι δύο μεταβλητές **δεν είναι συσχετισμένες**



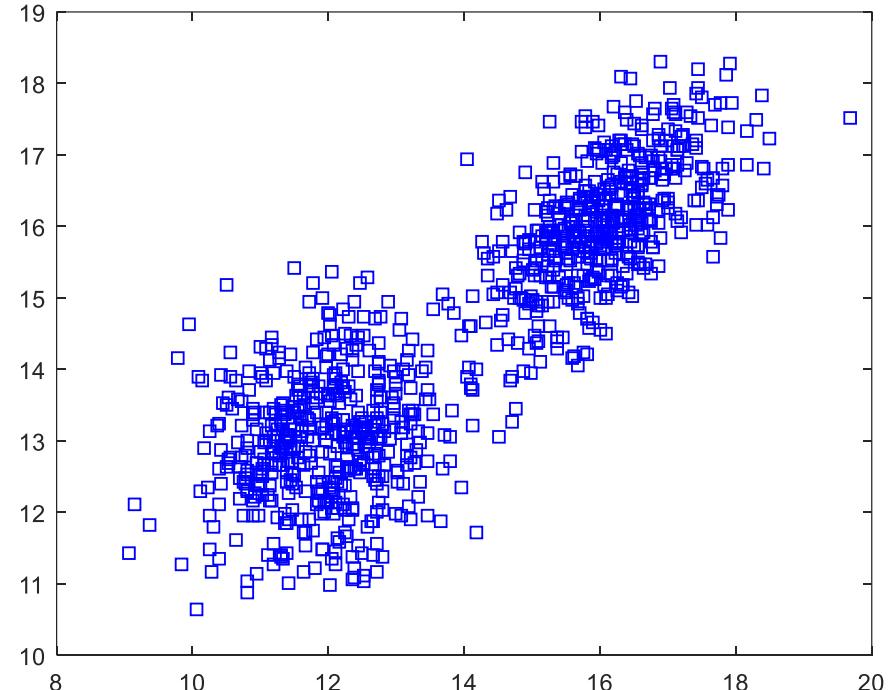
Μη επιβλεπόμενη εκμάθηση

- Ελάττωση διαστατικότητας
- **Κάποια χαρακτηριστικά μπορεί να είναι περιττά**
 - Συσχέτιση → Η γνώση για ένα χαρακτηριστικό μας δίνει πληροφορία για ένα άλλο χαρακτηριστικό
- Στο γράφημα δεξιά οι δύο μεταβλητές **δεν είναι συσχετισμένες**



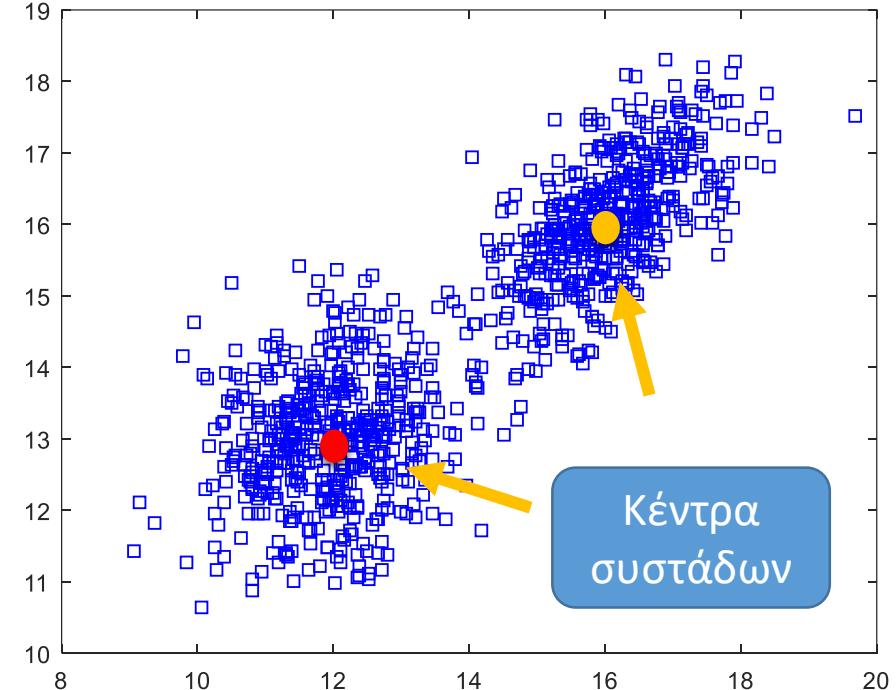
Μη επιβλεπόμενη εκμάθηση

- Επίδοση στα μαθηματικά και τη φυσική μαθητών της **Θεωρητικής** και της **Θετικής** Κατεύθυνσης.
- Μπορούμε εύκολα να αντιληφθούμε την ύπαρξη δύο συστάδων
- Επίσης παρατηρούμε ότι υπάρχει μια συσχέτιση στην επίδοση μεταξύ των μαθημάτων (περισσότερο στου μαθητές της Θετικής κατεύθυνσης)



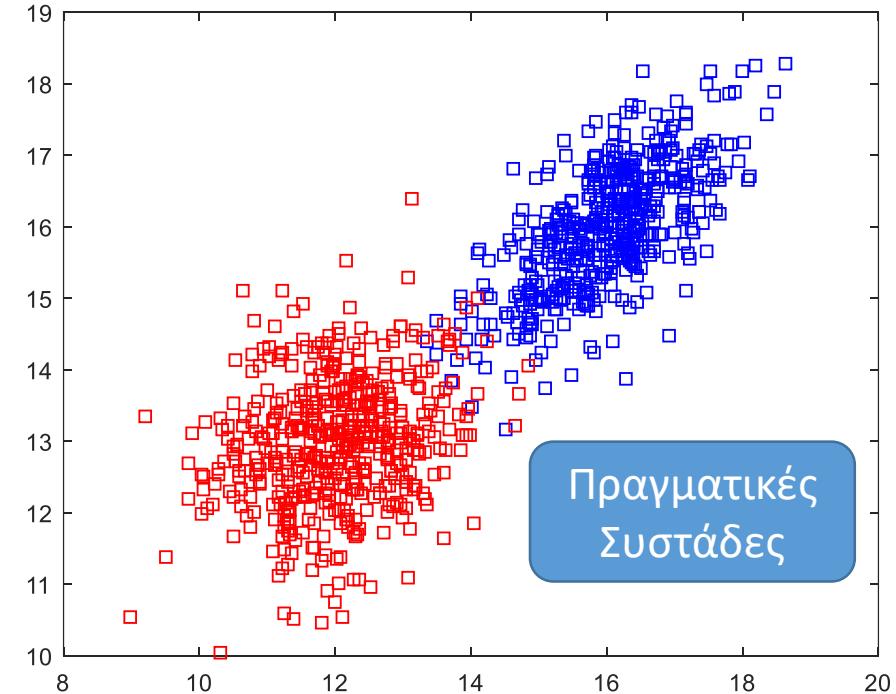
Μη επιβλεπόμενη εκμάθηση

- Επίδοση στα μαθηματικά και τη φυσική μαθητών της **Θεωρητικής** και της **Θετικής** Κατεύθυνσης.
- Σε ένα τυπικό πρόβλημα εύρεσης συστάδων θέλουμε να εντοπίσουμε τις δύο ομάδες μελετώντας τη σχέση μεταξύ των δειγμάτων



Μη επιβλεπόμενη εκμάθηση

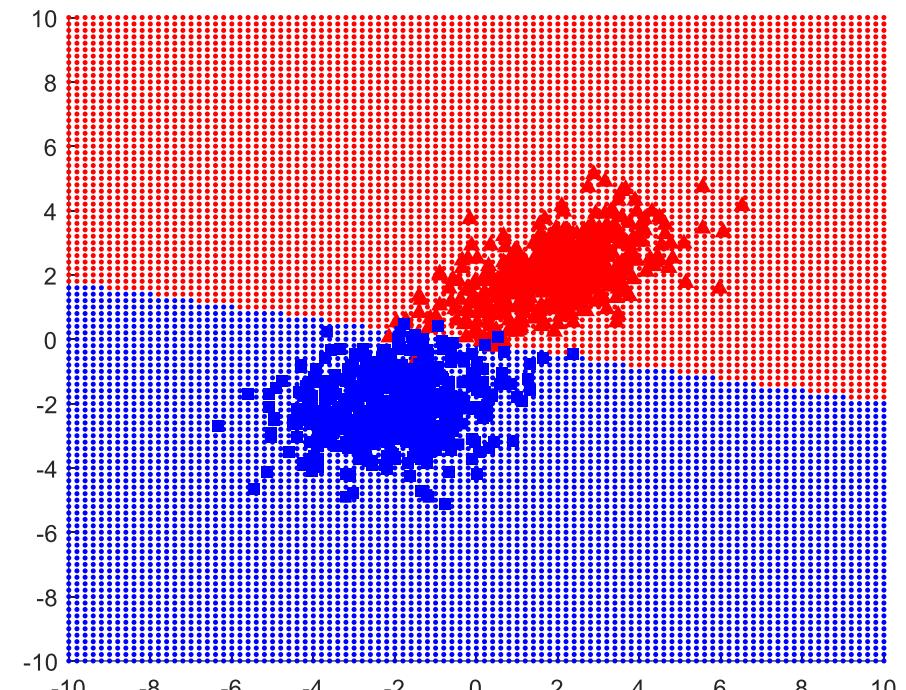
- Επίδοση στα μαθηματικά και τη φυσική μαθητών της **Θεωρητικής** και της **Θετικής** Κατεύθυνσης.
- Σε ένα τυπικό πρόβλημα εύρεσης συστάδων θέλουμε να εντοπίσουμε τις δύο ομάδες μελετώντας τη σχέση μεταξύ των δειγμάτων



Επιβλεπόμενη μηχανική εκμάθηση

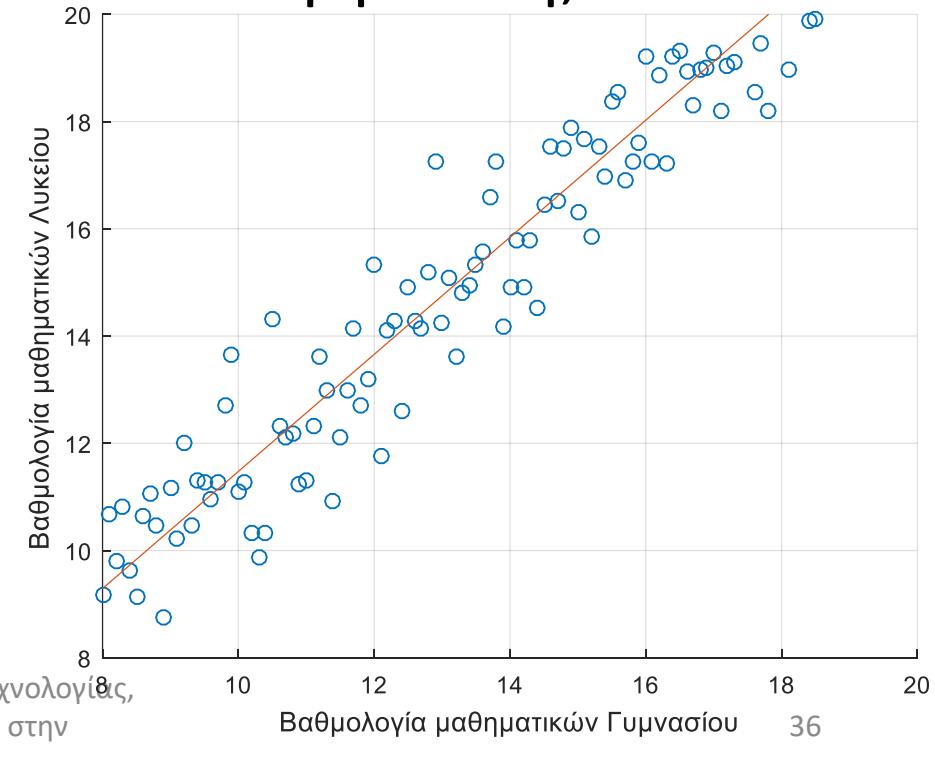
Επιβλεπόμενη εκμάθηση- Ταξινόμηση

- **Προβλήματα ταξινόμησης**
- Εάν θέλαμε να ταξινομήσουμε ένα μαθητή σε μια κατεύθυνση βάσει της βαθμολογίας του θα έπρεπε να βρούμε ένα τρόπο να διαχωρίσουμε τους πληθυσμούς
- Για το σκοπό αυτό θα χρειαζόμασταν ένα σύνολο δειγμάτων και τον χαρακτηρισμό τους



Επιβλεπόμενη εκμάθηση- Παλινδρόμηση

- **Προβλήματα παλινδρόμησης**
- Εάν θέλαμε να εκτιμήσουμε την αναμενόμενη επίδοση ενός μαθητή στα μαθηματικά γνωρίζοντας τη βαθμολογία του στη φυσική;
- Εάν θέλαμε να εκτιμήσουμε το βαθμό της πρώτης τάξης μαθηματικών του λυκείου από τις βαθμολογίες μαθηματικών του γυμνασίου;



Λογισμικό και πλατφόρμες για μηχανική εκμάθηση

Λογισμικό και πλατφόρμες για μηχανική εκμάθηση

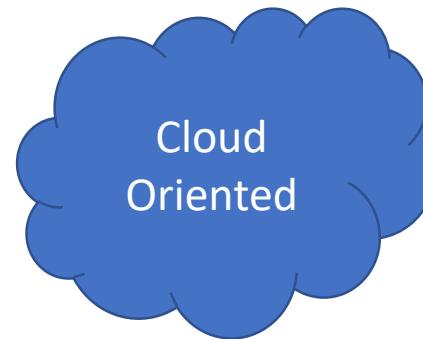
- Matlab/Octave
- *Python*
- *Lua*
- *Julia*
- *Shogun*
- *R Programming Language*



- *Lua/Torch*
- *TensorFlow*
- *Scilab*
- *mlpack*
- *dlib*

Λογισμικό και πλατφόρμες για μηχανική εκμάθηση

- Microsoft Azure ML Studio
- Google Cloud AI
- Amazon AWS Artificial Intelligence
- OpenML
- Pattern



Γλώσσες προγραμματισμού

- Python
- Matlab
- C++
- Java
- Lua
- Julia

Σύνοψη πρώτου μέρους

- Μηχανική εκμάθηση (Ορισμός και εφαρμογές)
- *Προβλήματα μη-επιβλεπόμενης εκμάθησης*
 - *Εύρεση συστάδων στα δεδομένα*
- *Προβλήματα επιβλεπόμενης εκμάθησης*
 - *Παλινδρόμηση, Ταξινόμηση*
- *Γλώσσες προγραμματισμού- πλατφόρμες*

Μέρος 2^ο

Μη-επιβλεπόμενη εκμάθηση

Τεχνικές μη-επιβλεπόμενης εκμάθησης

- Η μέθοδος k-means
- Τεχνικές φασματικής τμηματοποίησης γράφου
- Τεχνικές ελάττωσης διαστατικότητας
- Τεχνικές απεικόνισης δεδομένων υψηλής διαστατικότητας και μεγάλου αριθμού δειγμάτων
- Τεχνικές εκτίμησης πηγαίων σημάτων (blind source separation)

Η μέθοδος k-means

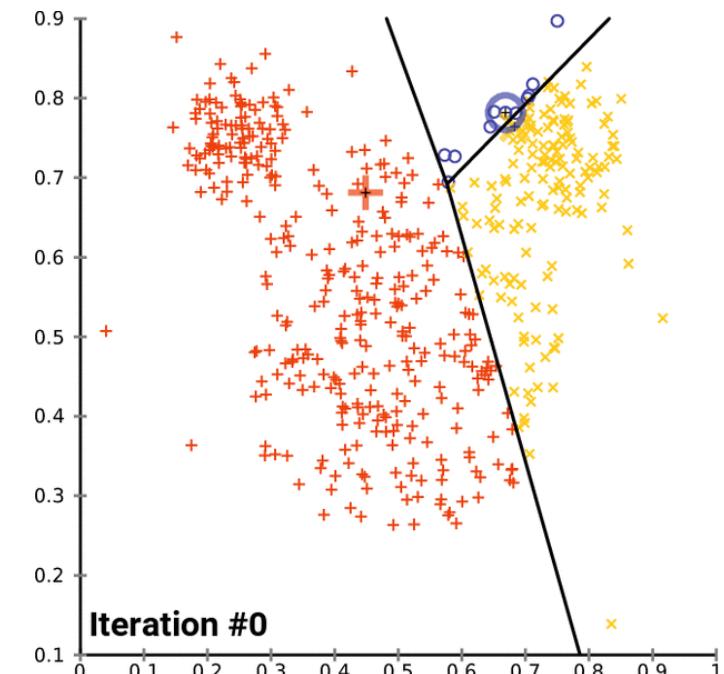
- Εντοπίζει αυτόματα τις υπάρχουσες συστάδες
- Επαναληπτική διαδικασία

Πλεονεκτήματα:

- Εύκολος αλγόριθμος, ικανοποιητικά αποτελέσματα

Μειονεκτήματα:

- Δύσκολο να εκτιμηθεί ο αριθμός των κέντρων



Η μέθοδος k-means

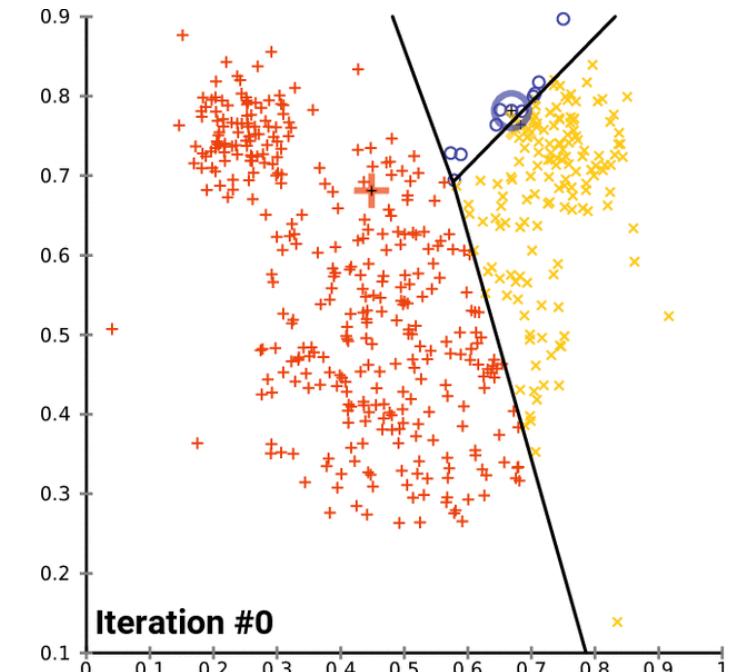
- Εντοπίζει αυτόματα τις υπάρχουσες συστάδες
- Επαναληπτική διαδικασία

Πλεονεκτήματα:

- Εύκολος αλγόριθμος, ικανοποιητικά αποτελέσματα

Μειονεκτήματα:

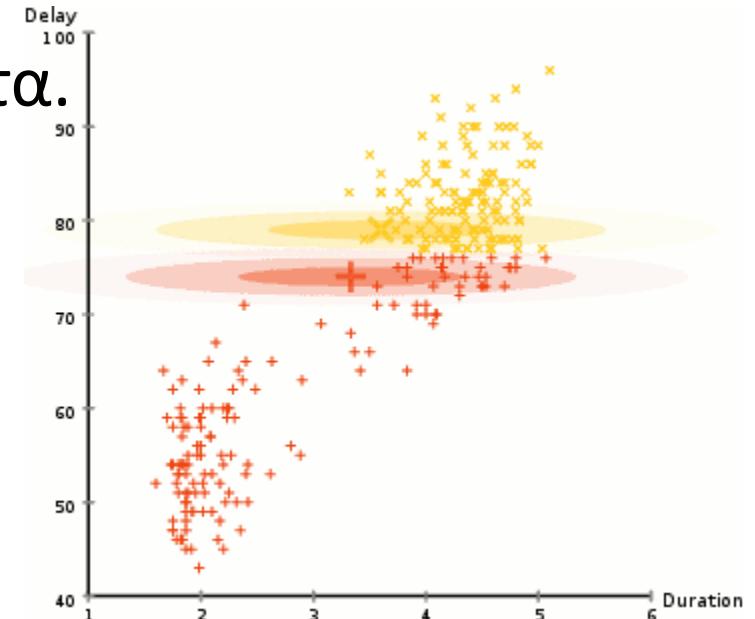
- Δύσκολο να εκτιμηθεί ο αριθμός των κέντρων



Καμία υπόθεση για την κατανομή των δειγμάτων

Expectation–maximization algorithm

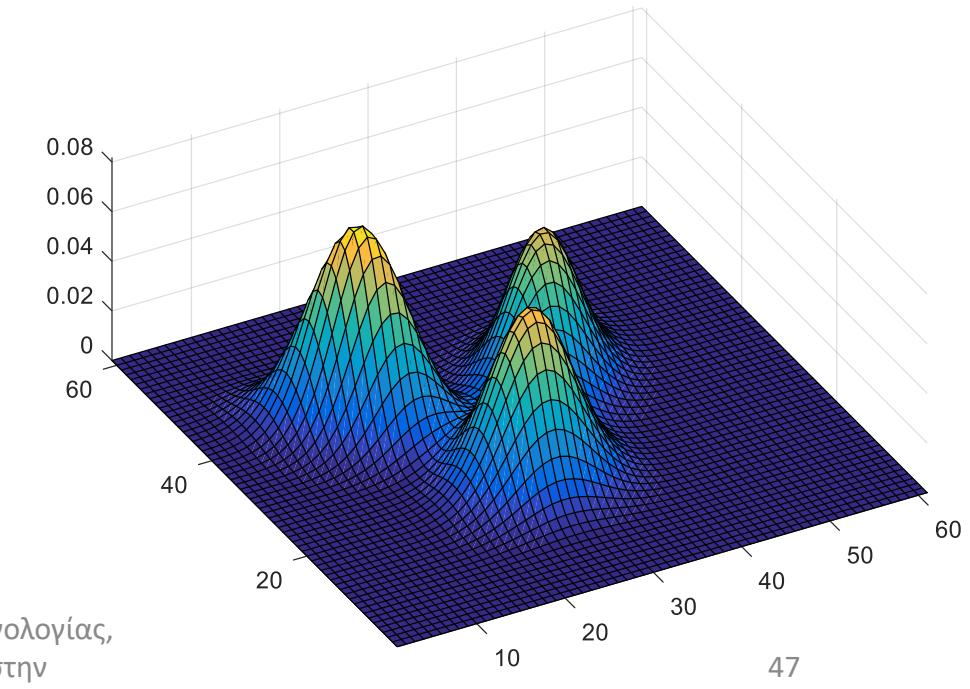
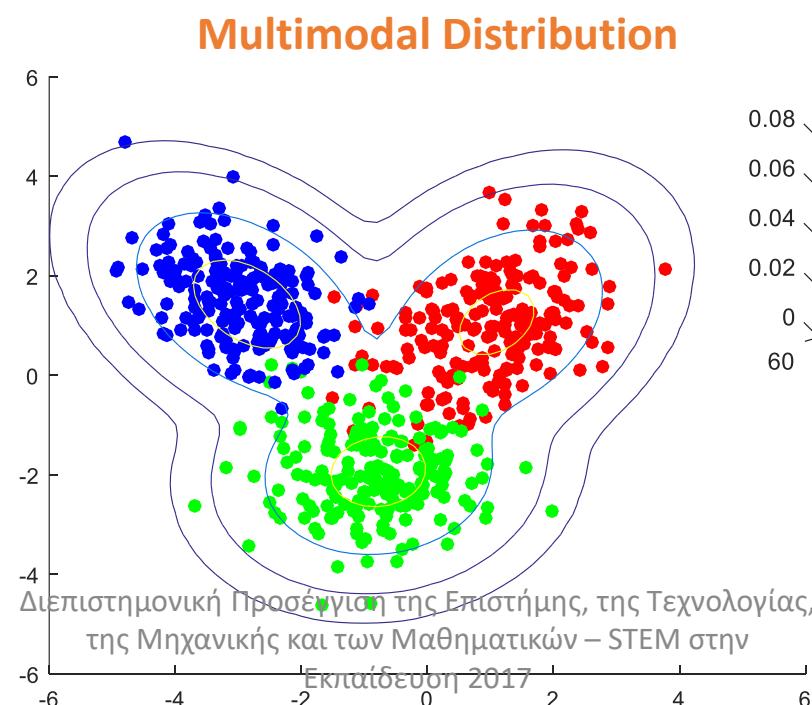
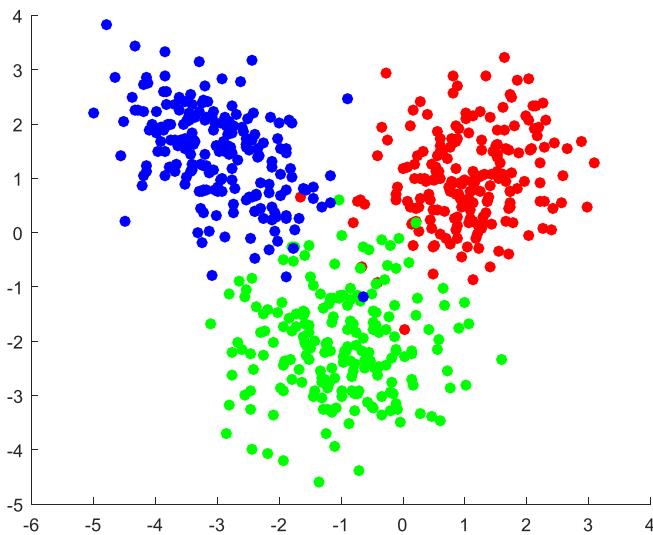
- Επαναληπτική διαδικασία με την οποία μπορούμε να εκτιμήσουμε τις παραμέτρους ενός μοντέλου (κατανομής).
- Επαναληπτική διαδικασία με δύο βασικά βήματα.
 - Ανάθεση του κάθε σημείου στην κατανομή ανήκει (μέγιστη πιθανότητα)
 - Ενημέρωση των παραμέτρων της κατανομής



Πηγή: https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm

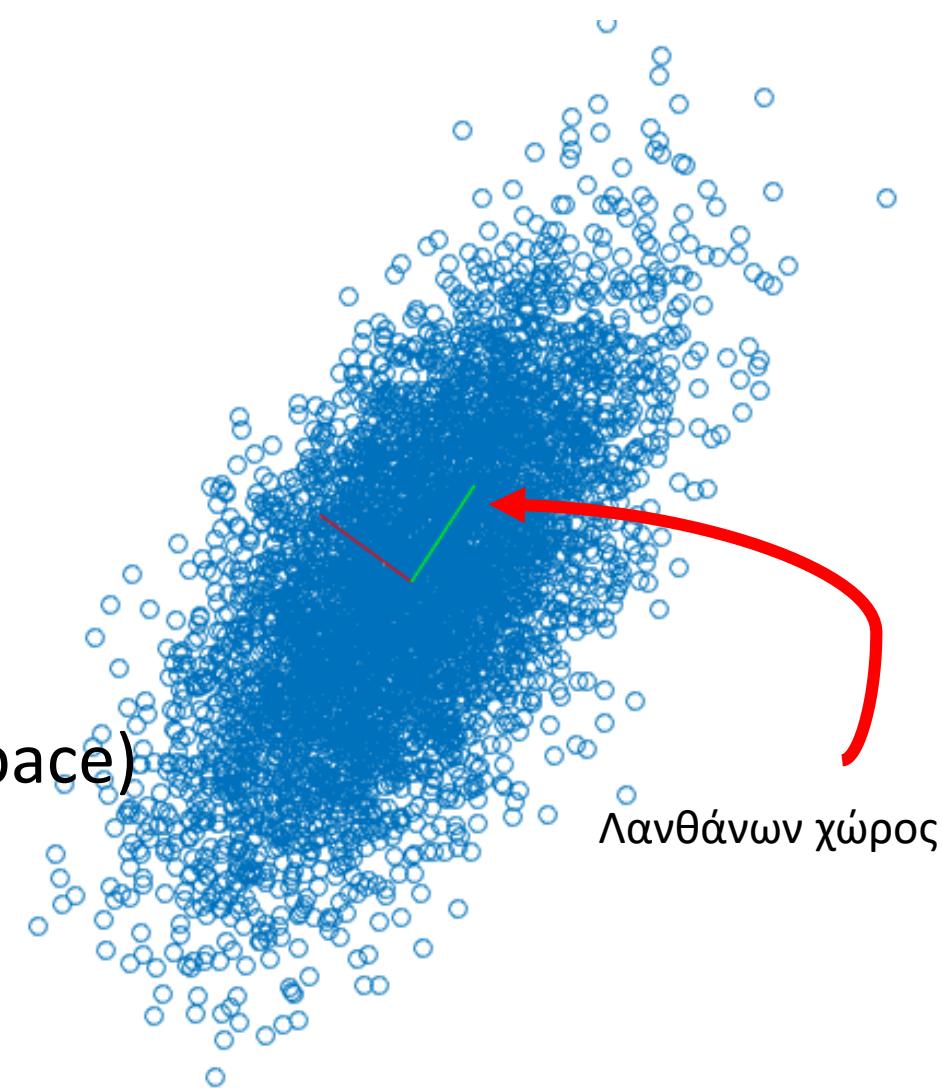
Gaussian-Mixture Models

- Η κανονική κατανομή, σε πολλές περιπτώσεις δεν μπορεί να αναπαραστήσει πιστά τα δεδομένα
- Ο συνδυασμός Gaussian κατανομών μπορεί να περιγράψει πολύπλοκες κατανομές.



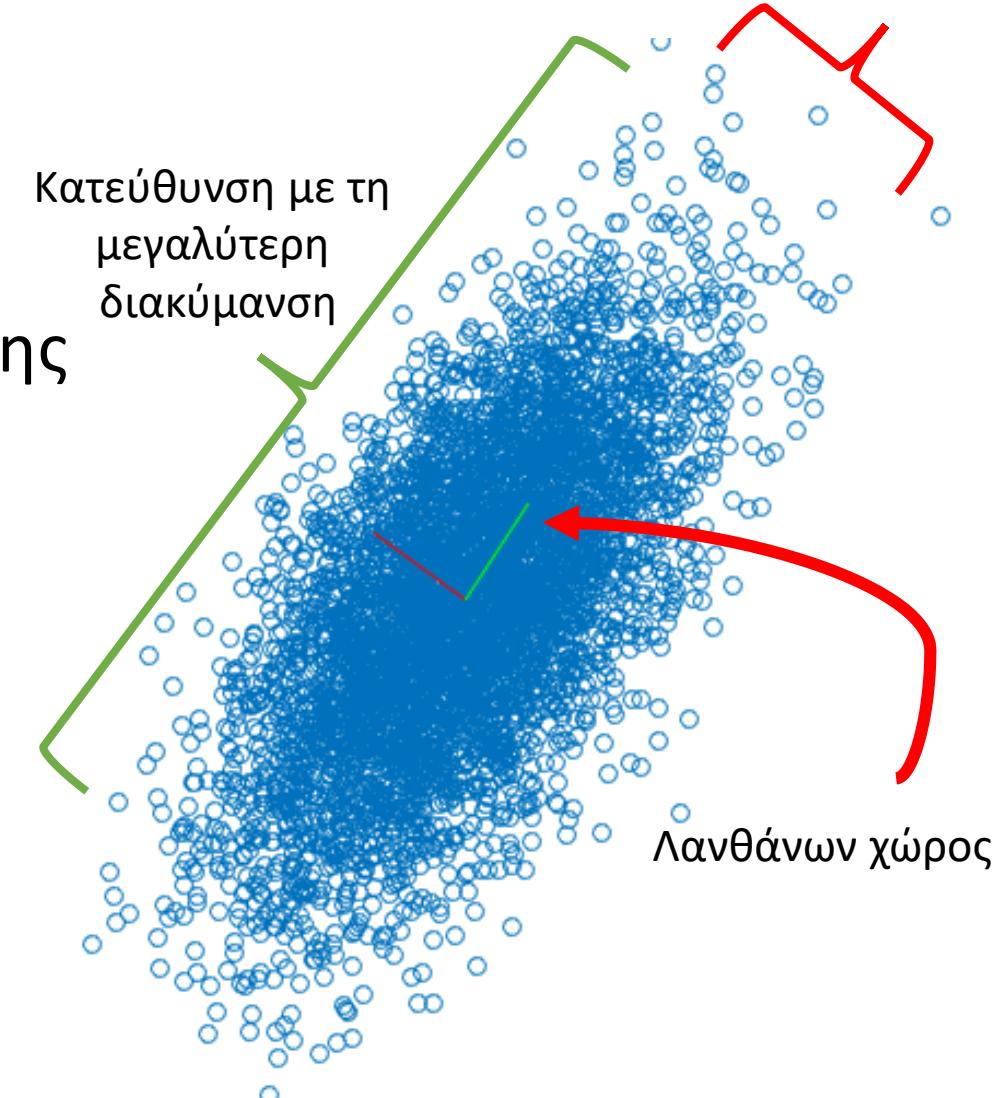
Ελάττωση διαστατικότητας

- Principal Component Analysis (PCA)
- Αποτελεί μια γραμμική τεχνική ελάττωσης διαστατικότητας
- Απεικόνιση σε ένα λανθάνων χώρο (latent space)



Ελάττωση διαστατικότητας

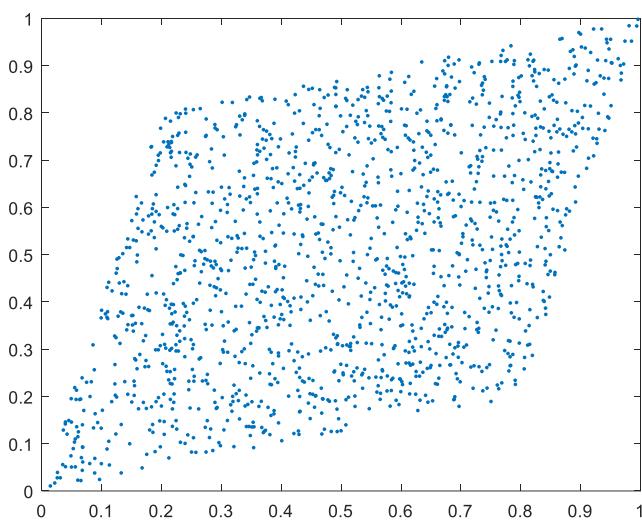
- Principal Component Analysis (PCA)
- Ιδιοδιανύσματα του πίνακα συνδιακύμανσης
- Διατήρηση των αξόνων στους οποίους τα δεδομένα έχουν μεγάλη διακύμανση
- Αποσυσχετίζει τα δεδομένα
- Ορθοκανονικό σύστημα συντεταγμένων



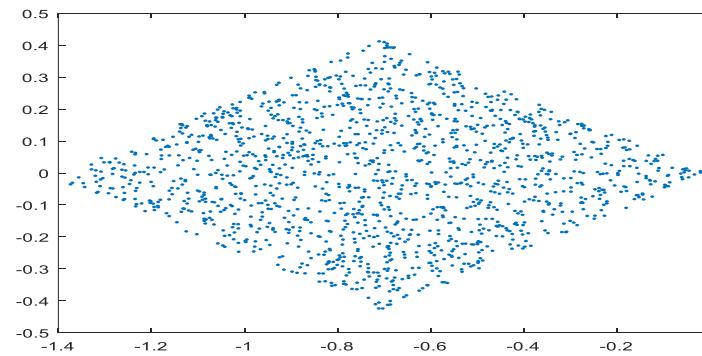
Ελάττωση διαστατικότητας

- Principal Component Analysis (PCA)

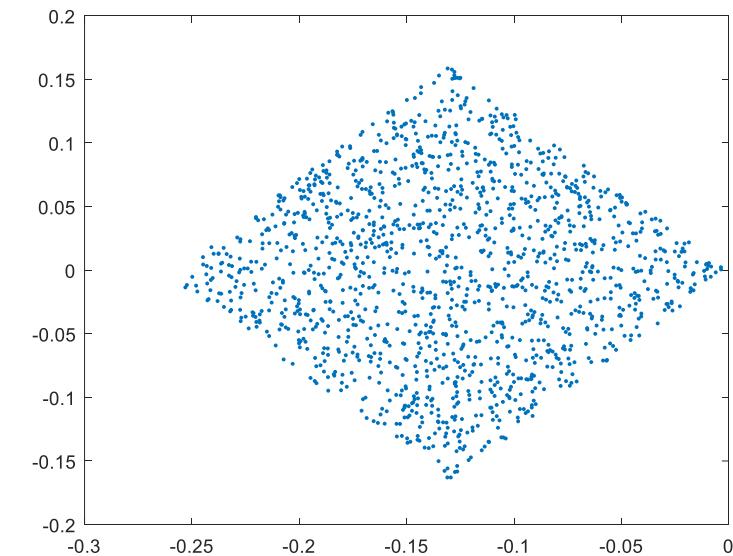
Uniform data Correlated



Uniform data decorrelated with PCA



Whitening



Ελάττωση διαστατικότητας

- **Non-Negative Matrix Factorization (NMF)**
- Η PCA μπορεί να μην αποτελεί μια διαισθητική λύση καθώς τα ιδιοδιανύσματα μπορεί να λάβουν αρνητικές τιμές
- **Επίσης οι άξονες είναι ορθογώνιοι**
- Η NMF παρέχει μια ελκυστική λύση καθώς βρίσκει ένα λανθάνων χώρο όπου οι άξονες είναι αποκλειστικά θετικοί (καθώς και οι συντελεστές ανακατασκευής)
- Επίσης οι άξονες δεν είναι υποχρεωτικό να είναι ορθογώνιοι

Ελάττωση διαστατικότητας

- **Non-Negative Matrix Factorization (NMF)**
- Βρίσκει εφαρμογές σε:
 - Ελάττωση διαστατικότητας
 - Ομαδοποίηση κειμένων – Εύρεση θεματικών δομών
- Κυρίως όπου τα δεδομένα αναπαρίστανται με θετικές τιμές (π.χ. εικόνες)

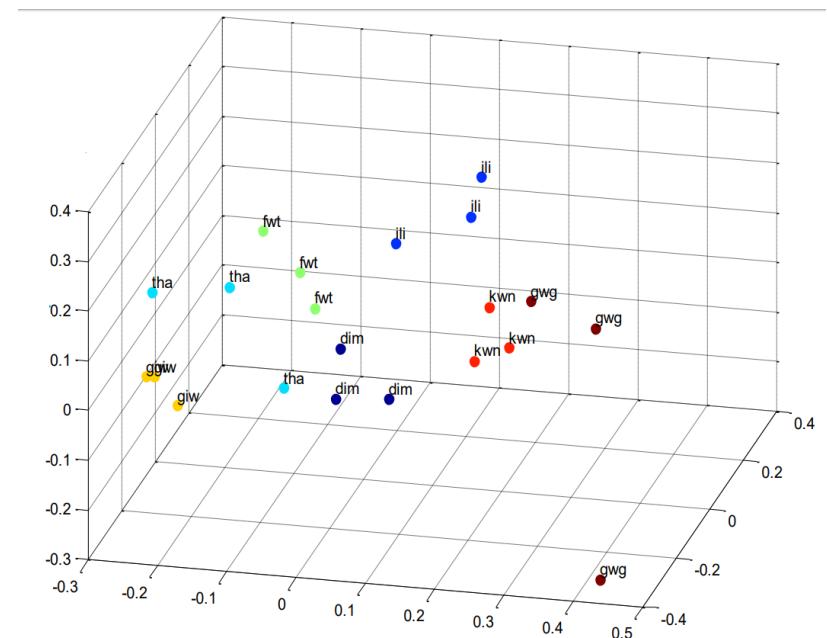
Ελάττωση διαστατικότητας

- Non-Negative Matrix Factorization (NMF)



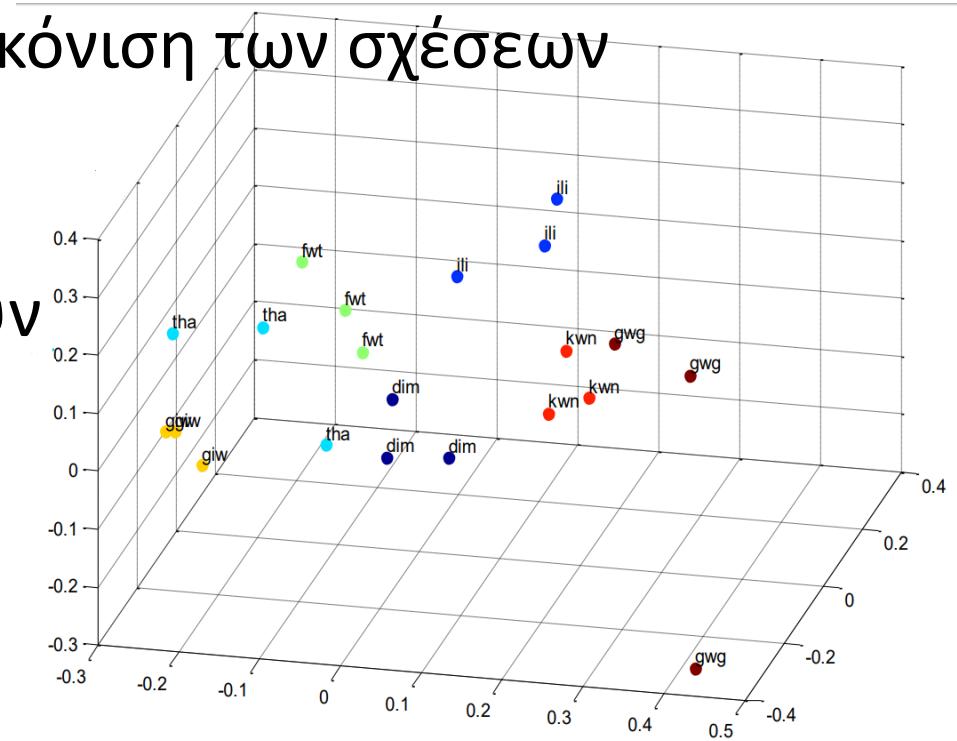
Ελάττωση διαστατικότητας

- Απεικόνιση των σχέσεων μεταξύ δειγμάτων που βρίσκονται σε διαστάσεις $>> 2$ στο επίπεδο
- Multidimensional scaling
 - Classical, Metric, Non-Metric, Generalized
- Τα σημεία στον λανθάνων χώρο διατηρούν (όσο είναι εφικτό) τις αποστάσεις του αρχικού χώρου
- Συνήθως δύο η τρεις διαστάσεις αρκούν



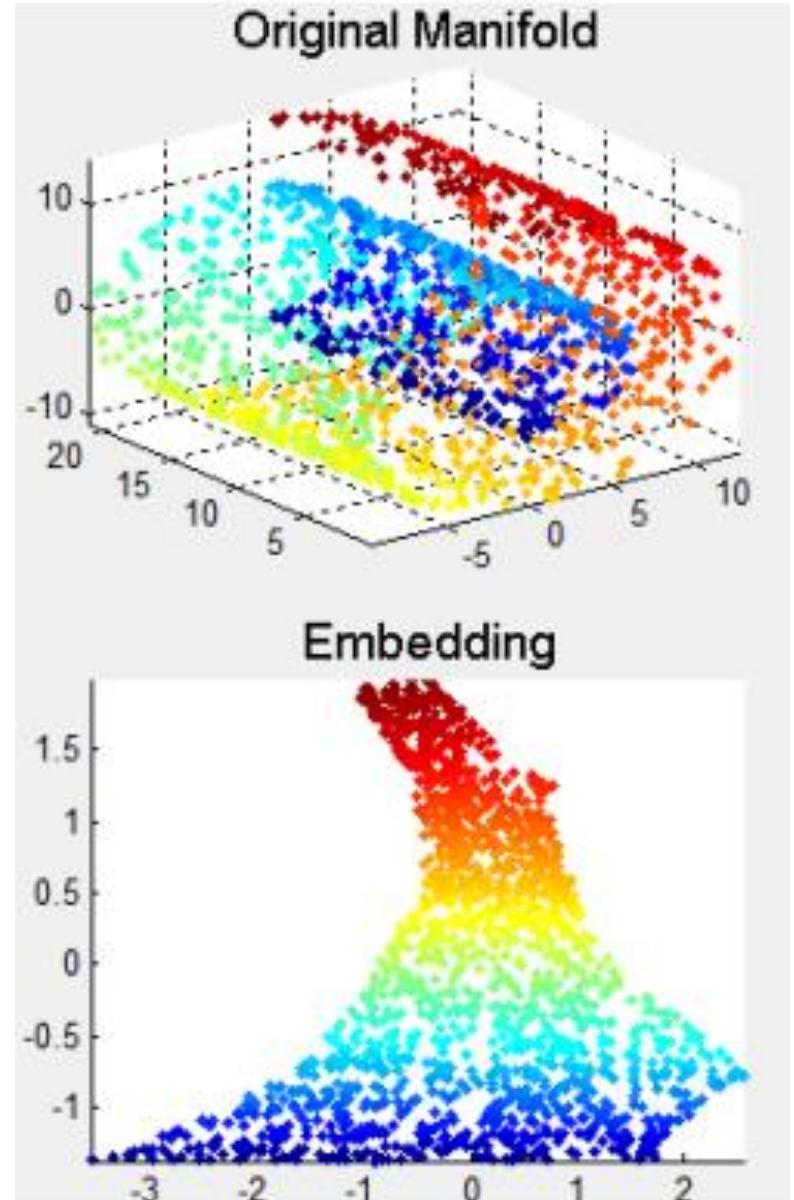
Ελάττωση διαστατικότητας

- Classical Multidimensional scaling (Principal Coordinates Analysis)
- Εξαιρετικά χρήσιμο εργαλείο για την απεικόνιση των σχέσεων δεδομένων σε χαμηλές διαστάσεις
- Τα σημεία στον λανθάνων χώρο διατηρούν τις αποστάσεις του αρχικού χώρου
- Συνήθως δύο ή τρεις διαστάσεις αρκούν



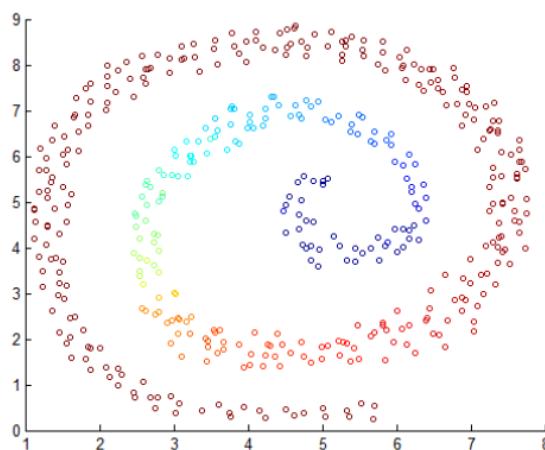
Ελάττωση διαστατικότητας

- Isomap
- Χρησιμοποιούνται γεωδαιτικές αποστάσεις
- Γίνεται αντιληπτή η γεωμετρία των δεδομένων
- Αποτελεί μια **μη γραμμική** τεχνική ελάττωσης διαστατικότητας

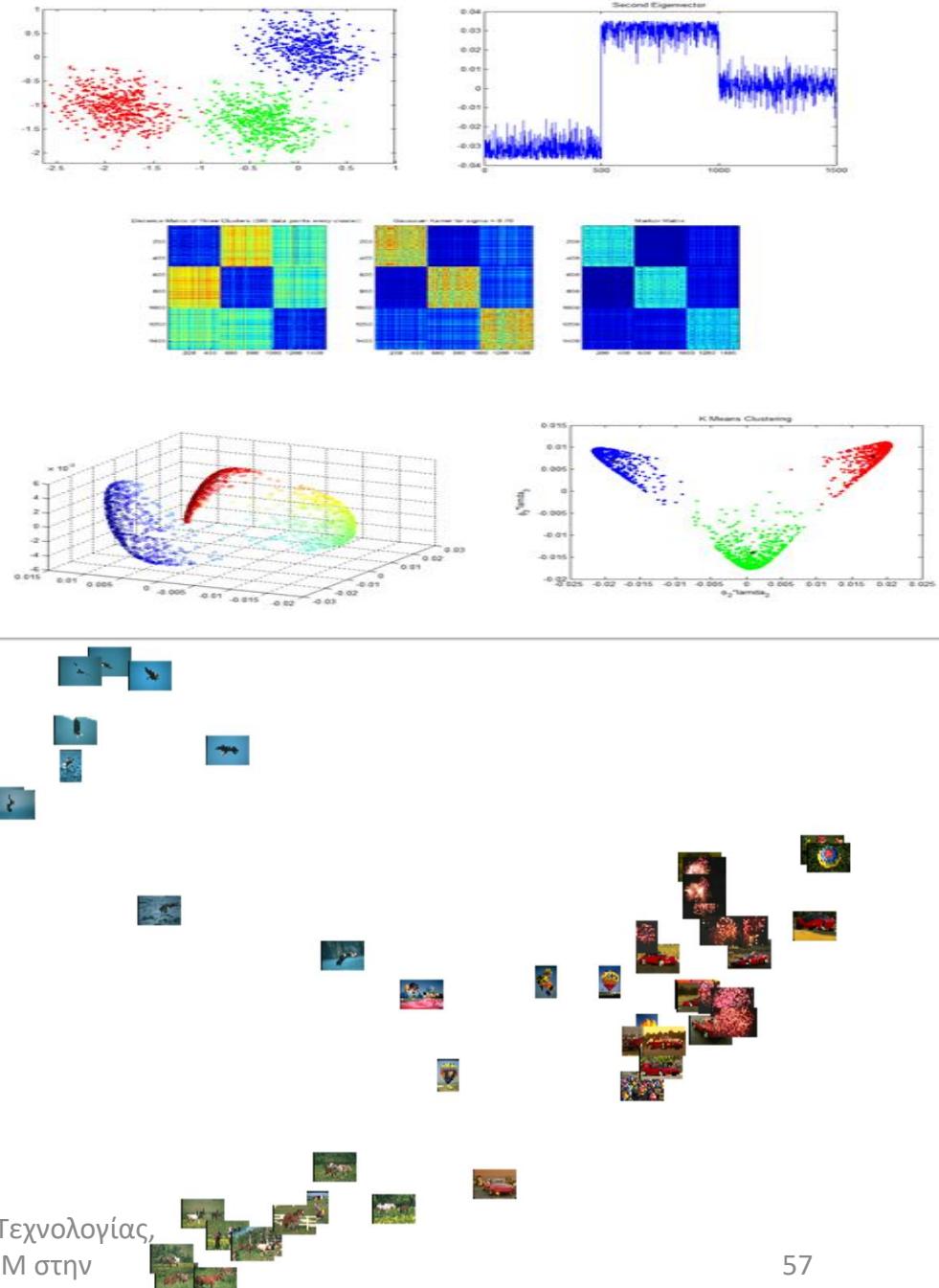


Ελάττωση διαστατικότητας

- Diffusion distance/Maps
- Λαμβάνει υπόψιν την πυκνότητα των δεδομένων
- Εισάγει την έννοια του χρόνου



Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας,
της Μηχανικής και των Μαθηματικών – STEM στην
Εκπαίδευση 2017



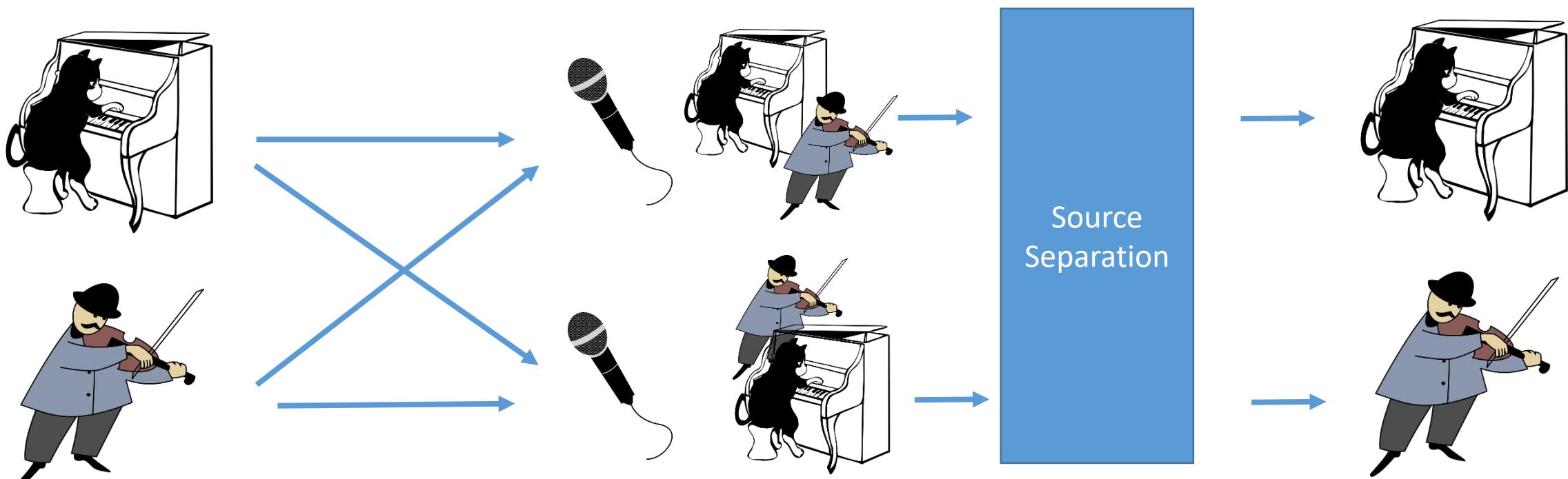
Εύρεση πηγαίων σημάτων

Διαισθητική περιγραφή

- Συνήθως καταγράφουμε ένα σήμα το οποίο προέρχεται από ένα σύνολο πηγών
- Σε πολλές περιπτώσεις δεν μας ενδιαφέρει το σήμα που καταγράφουμε αλλά οι επιμέρους πηγές που το δημιούργησαν
- Πως μπορούμε λοιπόν να ανακτήσουμε τα σήματα πηγής;

Εύρεση πηγαίων σημάτων

Διαισθητική περιγραφή- “Cocktail party problem”



Εύρεση πηγαίων σημάτων

- Independent Component Analysis- ICA
- Τα σήματα αποτελούν μίξη σημάτων πηγής
- Προσπαθεί αν ανακτήσει τα πηγαία σήματα καθώς και τον πίνακα μίξης
- **Βασίζεται κυρίως στο γεγονός ότι τα πηγαία δεδομένα είναι “λιγότερο Gaussian” από τα σήματα μίξης (Central Limit Theorem)**

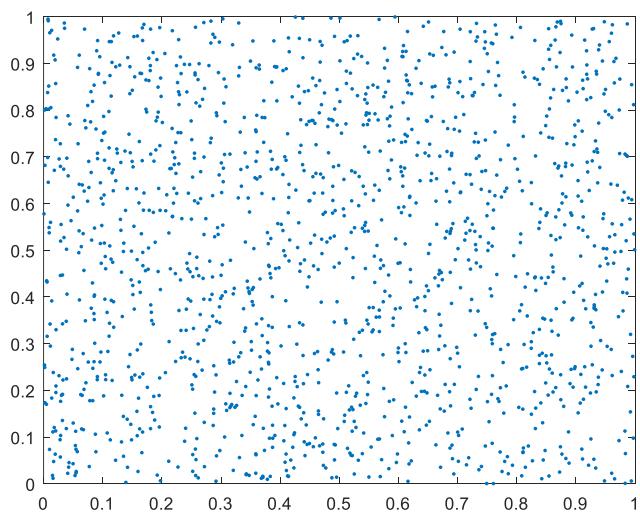
Εύρεση πηγαίων σημάτων

- Επίλυση του προβλήματος με διαφορετικές προσεγγίσεις
- Non- Gaussianity
 - Kyrtosis
 - Negentropy
 - Approximations of negentropy
- Information theory- Minimization of mutual information
- Maximum likelihood estimation

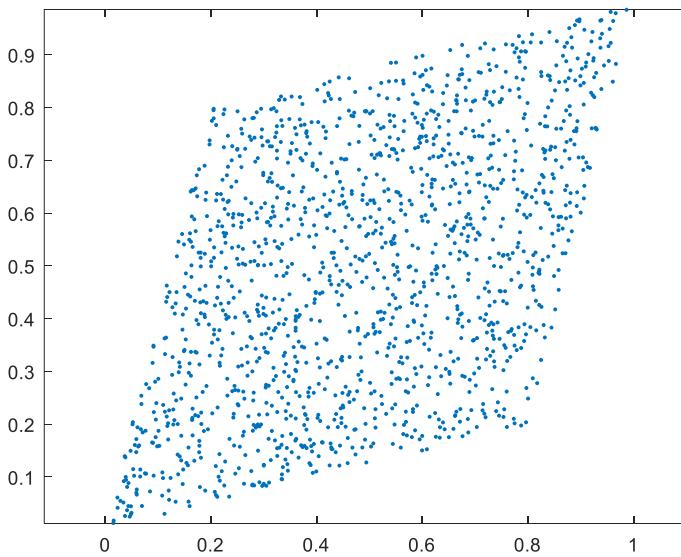
Εύρεση πηγαίων σημάτων

- Independent Component Analysis- ICA

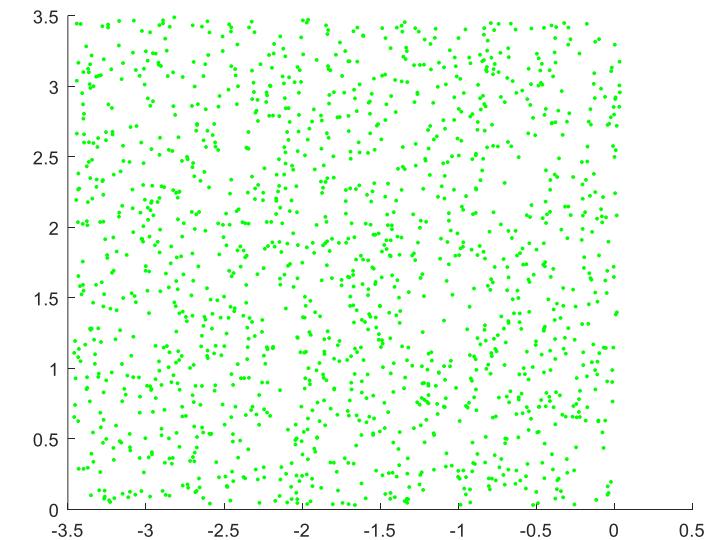
Original



Mixed



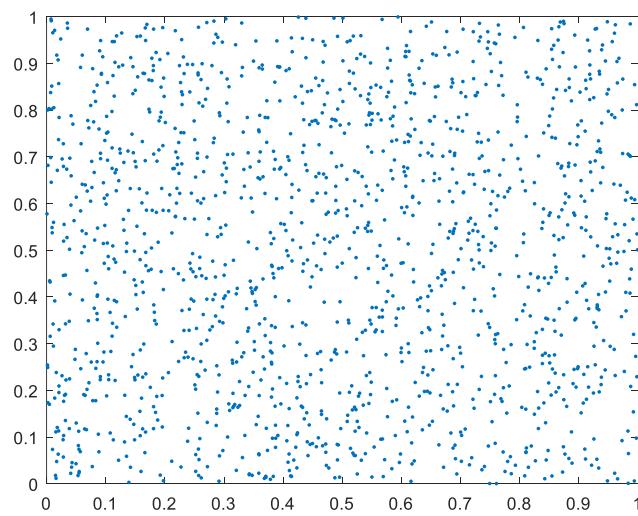
Recovered ICA



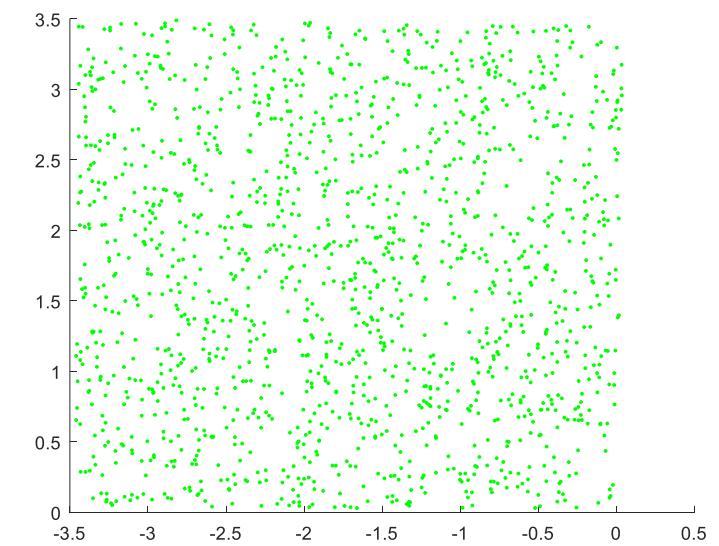
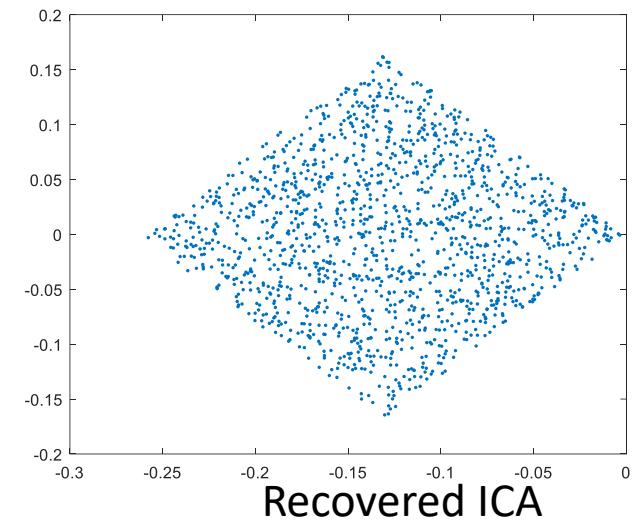
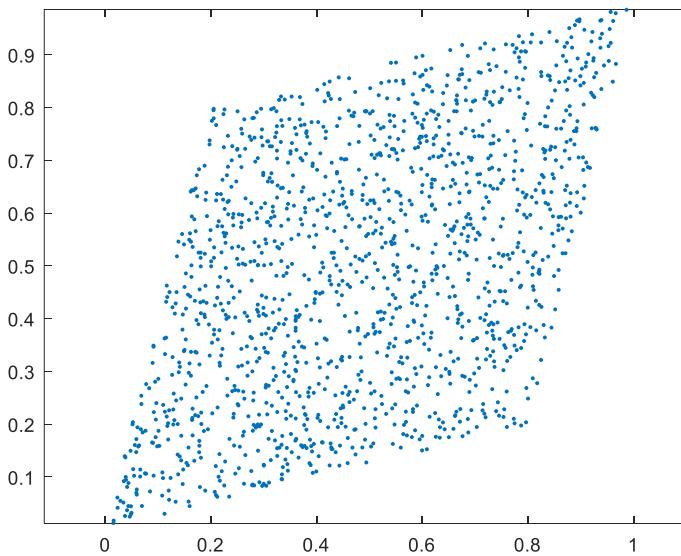
Εύρεση πηγαίων σημάτων

- Independent Component Analysis- ICA vs PCA

Original



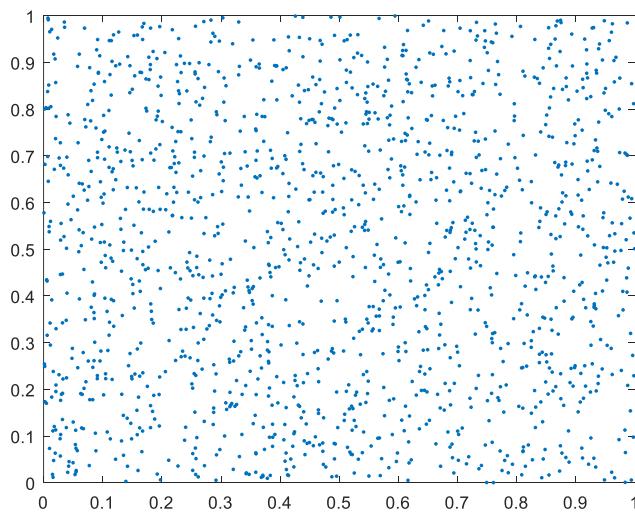
Mixed



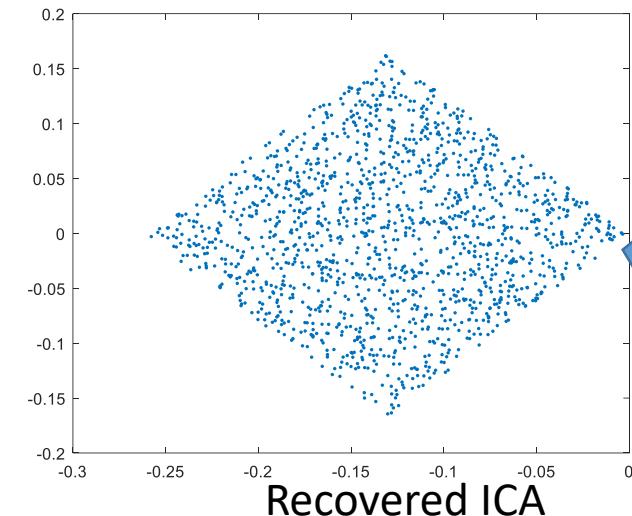
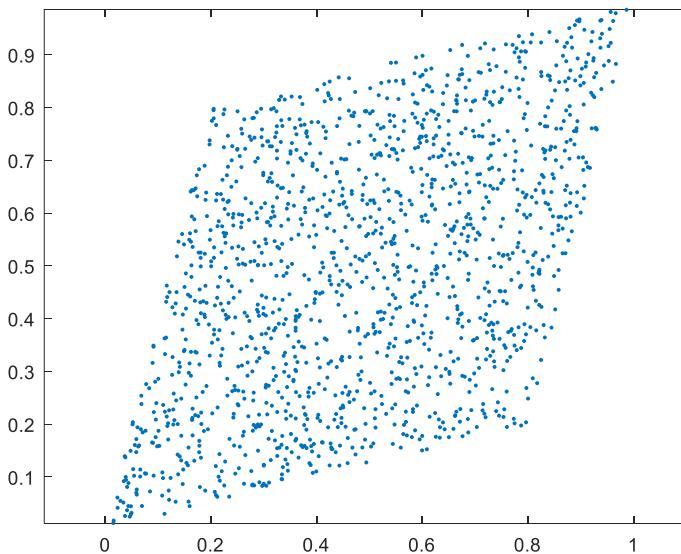
Εύρεση πηγαίων σημάτων

- Independent Component Analysis- ICA vs PCA

Original



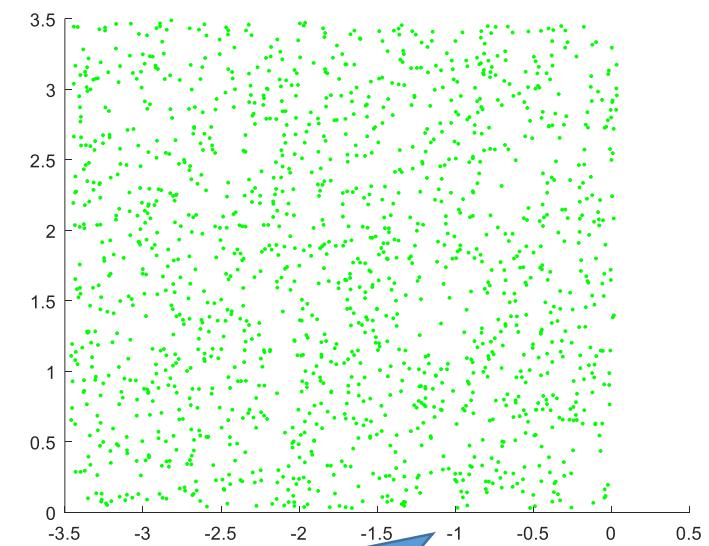
Mixed



Still
Gaussian



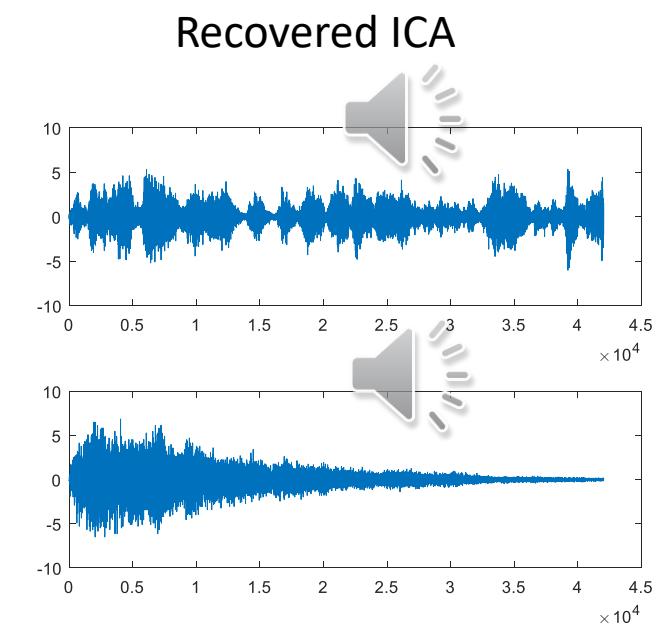
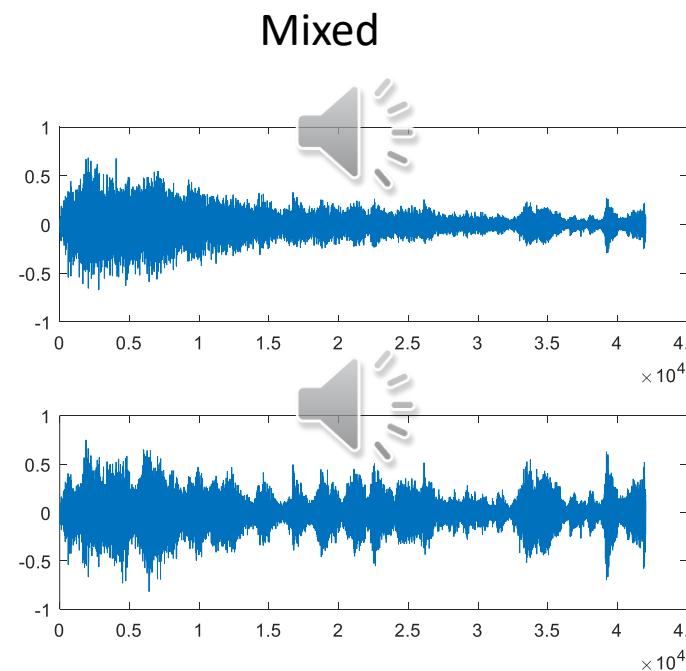
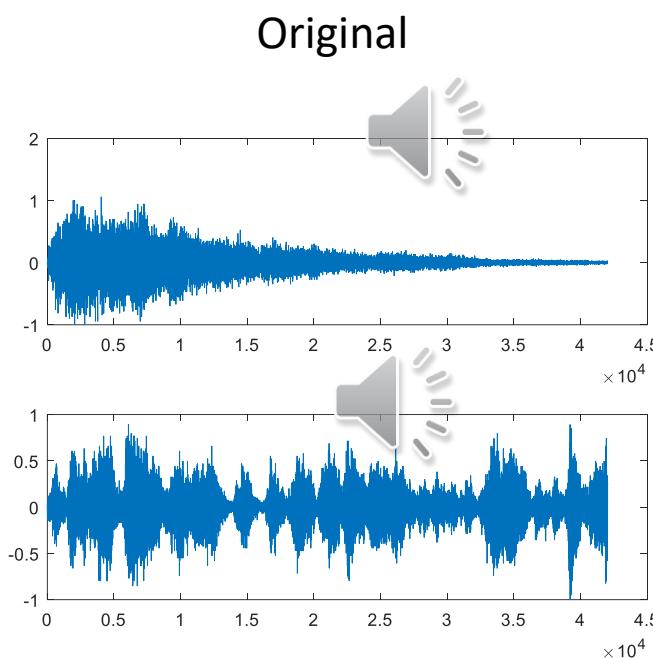
Recovered ICA



Less
Gaussian!

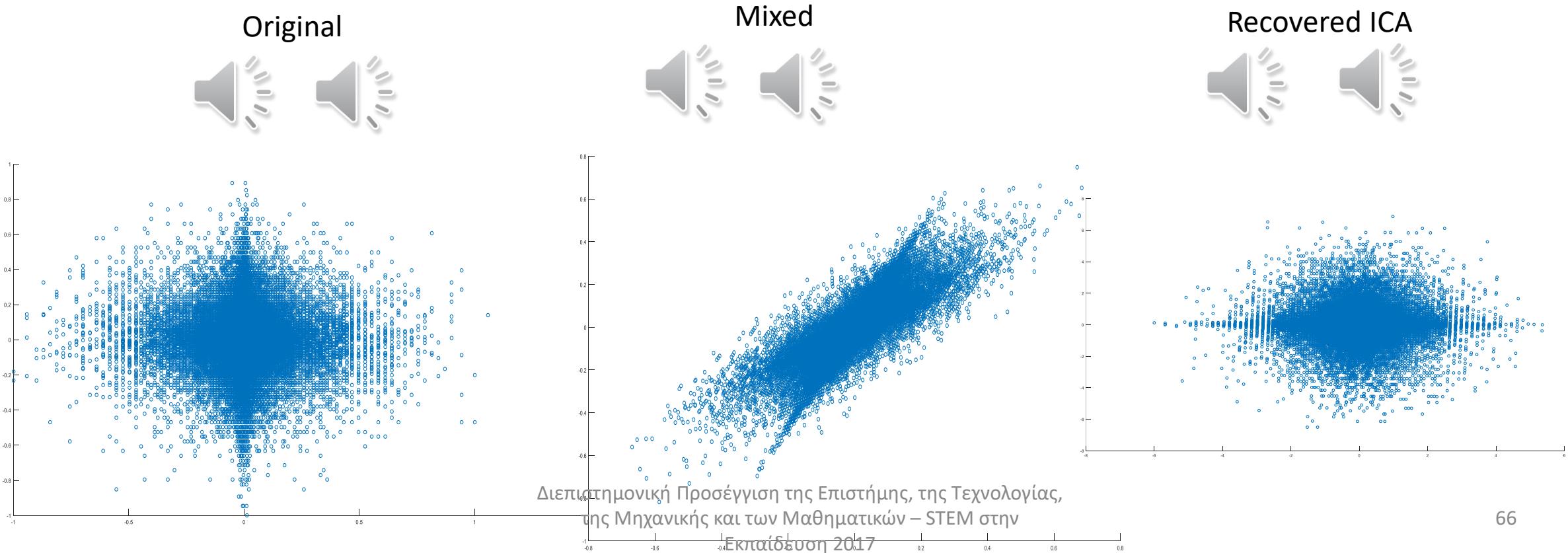
Εύρεση πηγαίων σημάτων

- Independent Component Analysis- ICA
- Ανάκτηση σημάτων ήχου



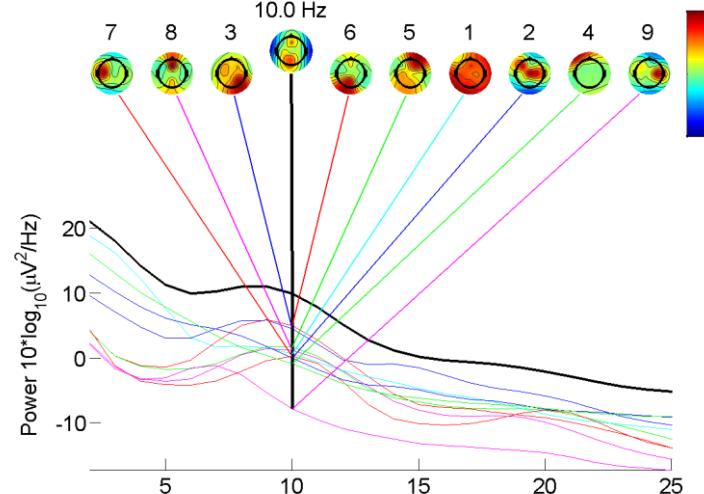
Εύρεση πηγαίων σημάτων

- Independent Component Analysis- ICA
- Ανάκτηση σημάτων ήχου



Εύρεση πηγαίων σημάτων

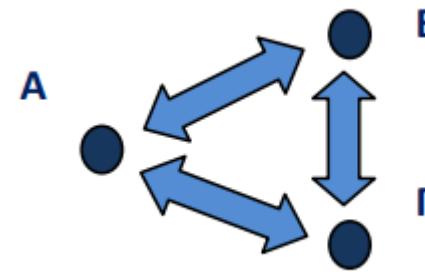
- Independent Component Analysis- ICA
- Εφαρμογές στην ανάλυση σημάτων EEG, ομιλίας, γεωφυσικά σήματα, τηλεφωνία κ.α.



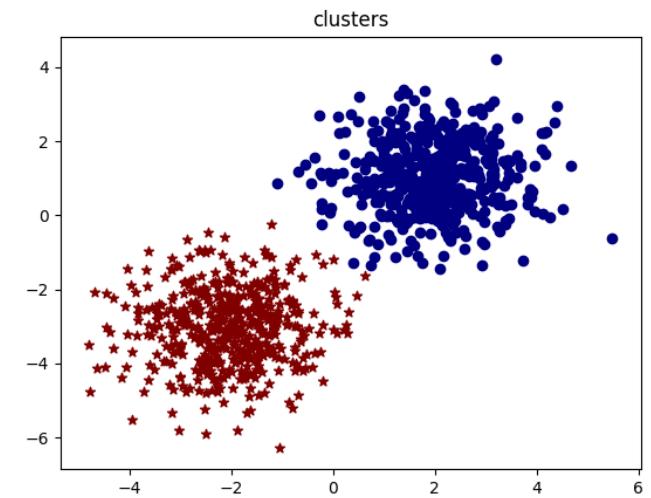
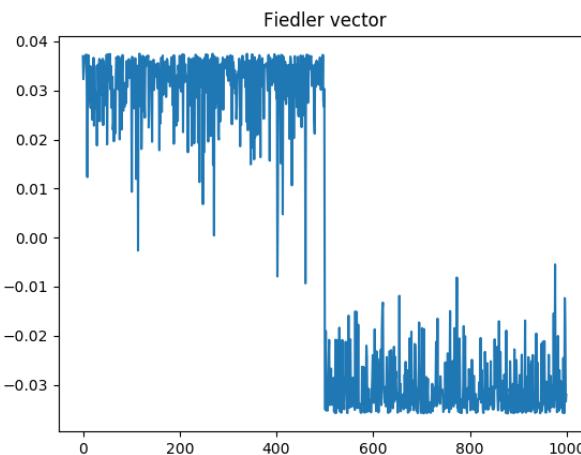
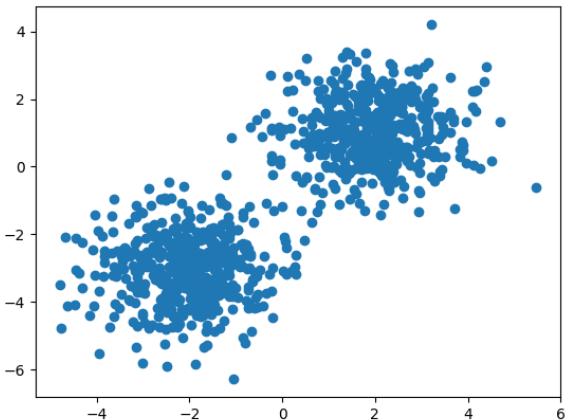
https://en.wikipedia.org/wiki/Independent_component_analysis

Τμηματοποίηση γράφου

- Fiedler Vector



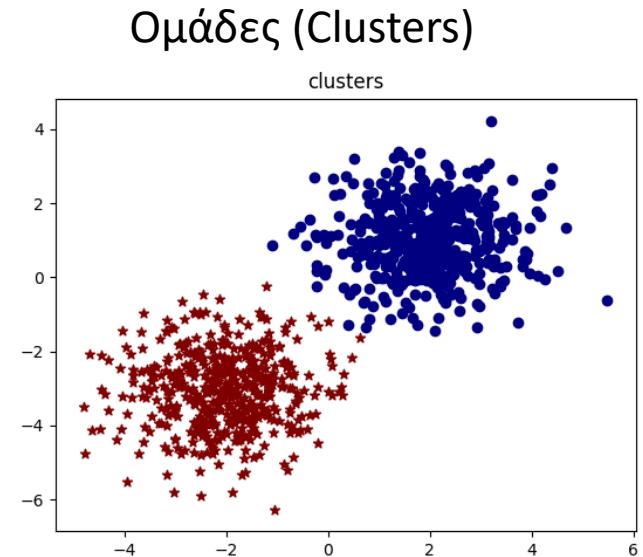
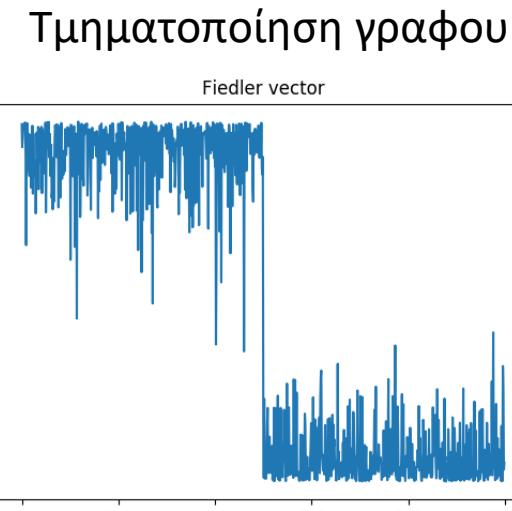
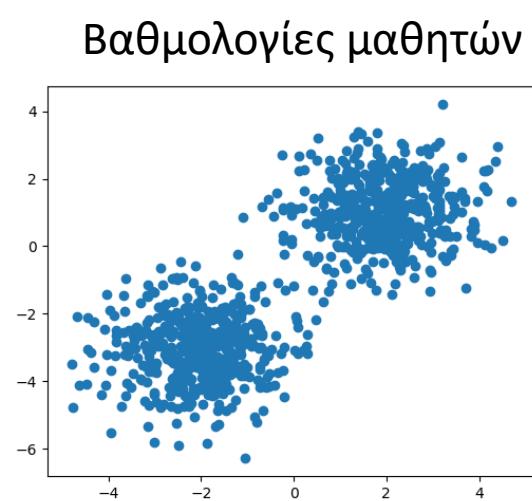
	A	B	Γ
A	0	1	1
B	1	0	1
Γ	1	1	0



<https://github.com/dimkastan/PyTorch-Spectral-clustering>

Εύρεση συστάδων από Γράφους

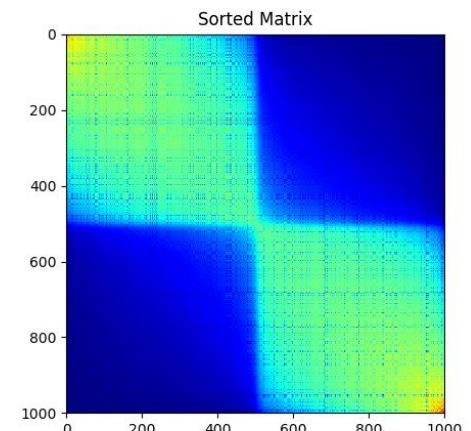
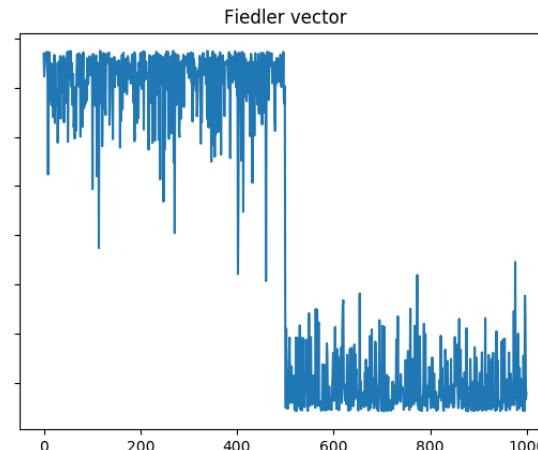
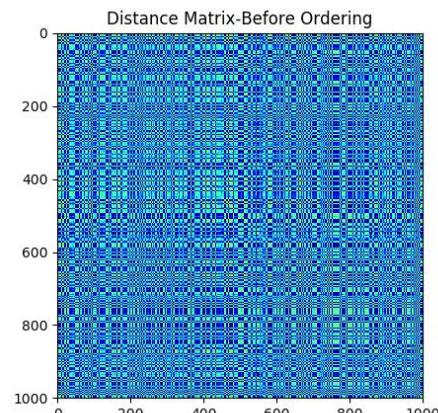
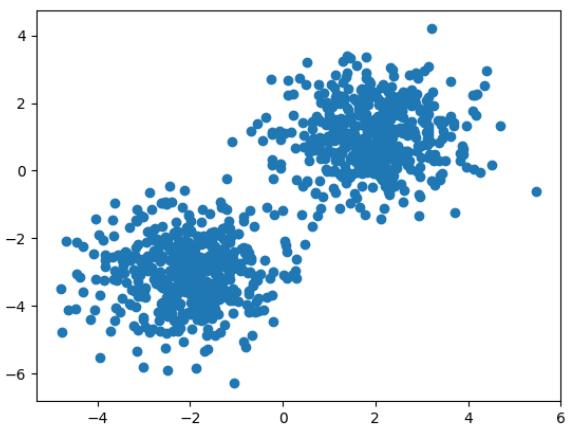
Θέλουμε να «χωρίσουμε» τους μαθητές μιας τάξης βάση της επίδοσης τους στα μαθηματικά και τη φυσική σε δύο ομάδες (Θεωρητική και Θετική κατεύθυνση)



<https://github.com/dimkastan/PyTorch-Spectral-clustering>

Εύρεση συστάδων από Γράφους

Θέλουμε να «οργανώσουμε» τους μαθητές μιας τάξης βάση της επίδοσης τους στα μαθηματικά και τη φυσική



<https://github.com/dimkastan/PyTorch-Spectral-clustering>

Μέρος 3^ο

Επιβλεπόμενη εκμάθηση

Στόχοι του μαθήματος

- Εισαγωγή στη επιβλεπόμενη μηχανική εκμάθηση
- Παρουσίαση των βασικών μεθόδων παλινδρόμησης και ταξινόμησης
- Τεχνικές βελτιστοποίησης
- Σχεδιασμός, αξιολόγηση και χρήση μοντέλων μηχανικής εκμάθησης

Επιβλεπόμενη εκμάθηση I

- Αποτελεί τη διαδικασία εκείνη όπου ένα σύστημα μέσα από μια διαδικασία εκμάθησης αποκτά εμπειρία από ένα σύνολο δεδομένων, ώστε να εκτελεί μια λειτουργία χωρίς να είναι σαφώς προγραμματισμένο
- Το σύστημα αυτό μαθαίνει μια **συνάρτηση απεικόνισης** μεταξύ των δεδομένων εισόδου και των δεδομένων εξόδου.

Στην πράξη:

- Ένα **μοντέλο** προσαρμοζεται στα δεδομένα με στόχο να ικανοποιήσει ενα **κριτήριο**.

Επιβλεπόμενη εκμάθηση II

- **Επιβλεπόμενη εκμάθηση**

- Δίνεται ένα σύνολο δειγμάτων υπό τη μορφή ζευγών εισόδου- εξόδου
- Στόχος μας είναι να μοντελοποιήσουμε τη σχέση μεταξύ των δεδομένων εισόδου και των δεδομένων εξόδου

- Δεδομένα εισόδου

- Predictors

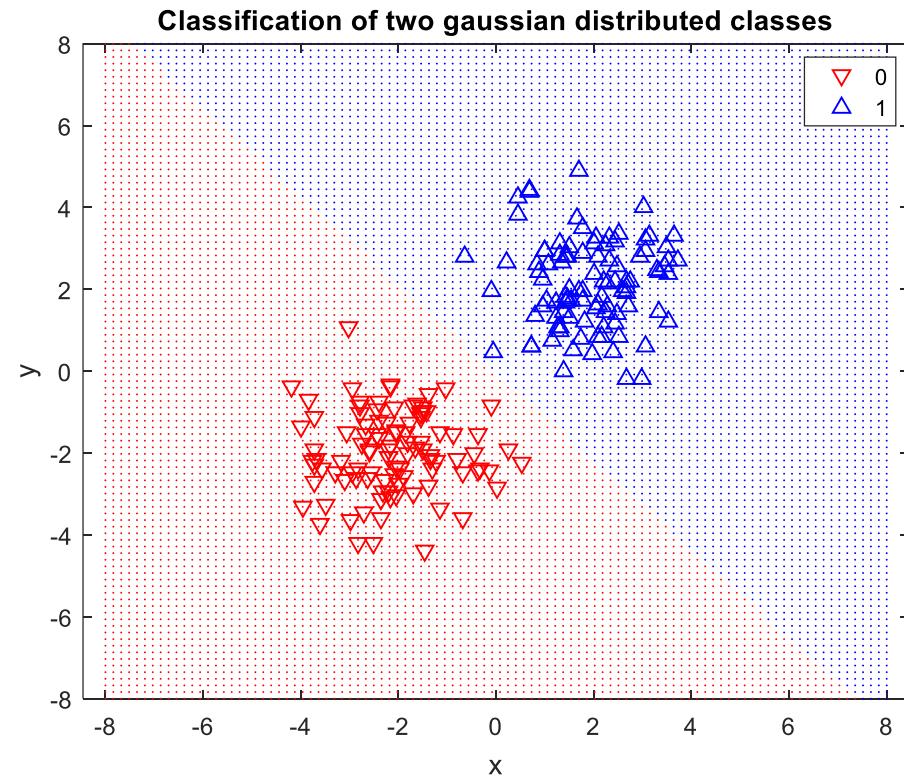
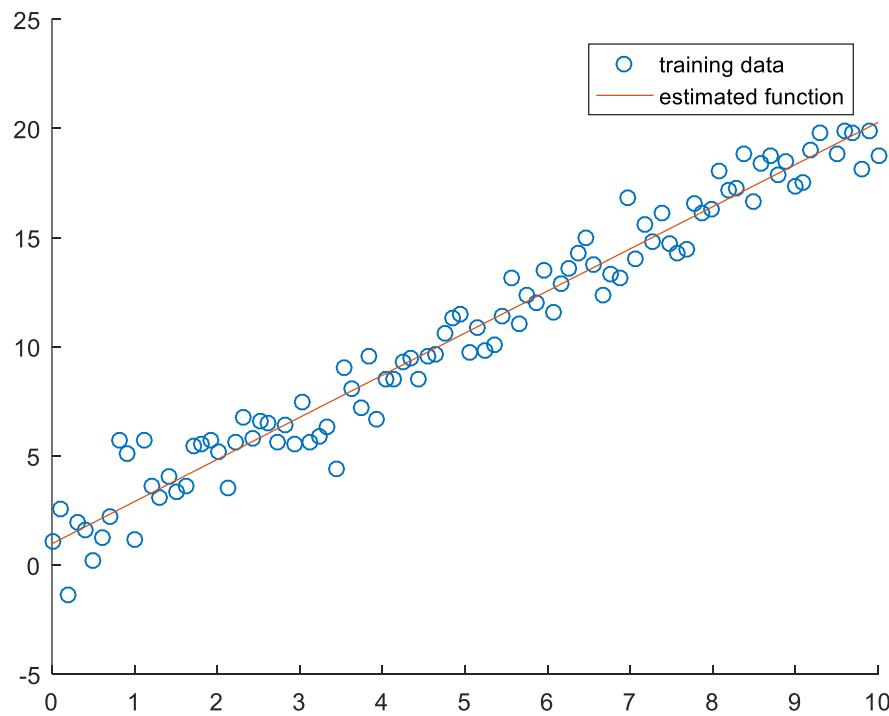
- Δεδομένα εξόδου

- Target qualitative or quantitative Variables (ποιοτικές ή ποσοτικές μεταβλητές)
- Ποσοτικές → **Παλινδρόμηση**
- Ποιοτικές → **Ταξινόμηση**

Επιβλεπόμενη εκμάθηση III



- Παραδείγματα- Παλινδρόμησης: Αριστερά, Ταξινόμησης: Δεξιά



Γραμμική παλινδρόμηση

Γραμμική παλινδρόμηση

- Παρουσίαση του προβλήματος

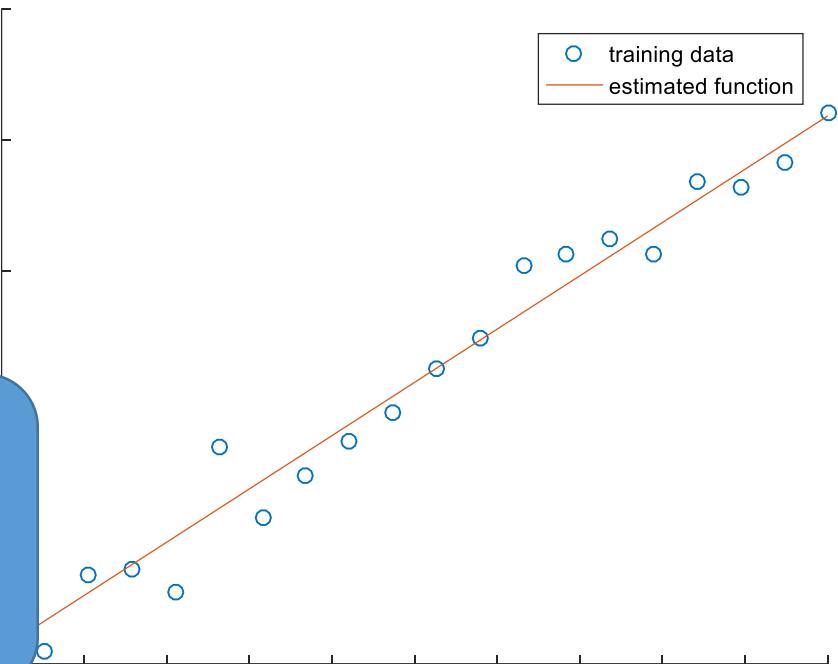
- Αναζητούμε τη σχέση μεταξύ του πρώτου (x) και του δεύτερου διαγωνίσματος (y)

Univariate

x	$Y=f(x)$
0	2.3676
0.5263	0.5003
1.0526	3.4089
1.5789	3.5852
2.1053	2.7585
....
10.0	21.0459

Τα Δεδομένα εισόδου δίνονται σε ζευγάρια της μορφής:
 $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$

Έστω ότι το κάθε διάνυσμα x έχει τον βαθμό του πρώτου διαγωνίσματος και y είναι μια βαθμωτή τιμή με το βαθμό του δεύτερου διαγωνίσματος για N μαθητές



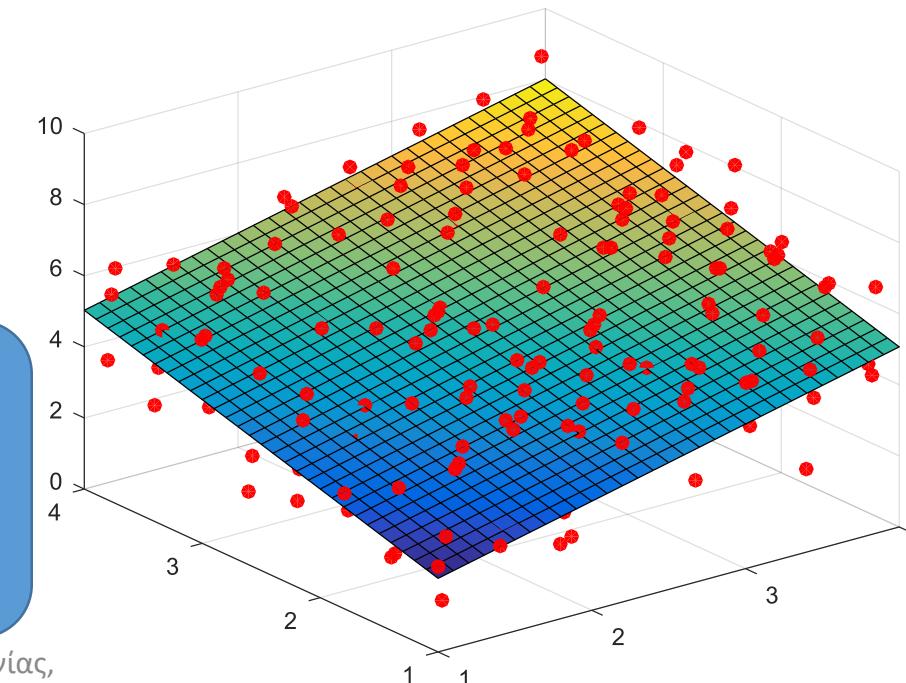
Γραμμική παλινδρόμηση

- Παρουσίαση του προβλήματος

- Αναζητούμε τη σχέση μεταξύ των δύο πρώτων (x) και του τρίτου διαγωνίσματος (y)

Multivariate

x	$Y=f(x)$	
1.1 , 1,8	2.3676	Τα Δεδομένα εισόδου δίνονται σε
2.2, 1,4	4.5003	ζευγάρια της μορφής: $(x_1,y_1),(x_2,y_2),(x_3,y_3),\dots,(x_n,y_n)$
....	
9.5, 8,9	21.0459	Έστω ότι το κάθε διάνυσμα x έχει τους βαθμούς των δύο πρώτων διαγωνισμάτων και y είναι μια βαθμωτή τιμή με το βαθμό του τρίτου διαγωνίσματος



Γραμμική παλινδρόμηση

- Θα ασχοληθούμε με γραμμικά μοντέλα της μορφής

$$y = w_0 \cdot x_0 + w_1 \cdot x_1 + \cdots + w_d \cdot x_d$$

- Για y μια scalar μεταβλητή και για $x_0 = 1$, έχουμε:

$$y = w_0 \cdot 1 + w_1 \cdot x_1 + \cdots + w_d \cdot x_d$$

Όπου σε μορφή πίνακα διατυπώνεται ως εξής:

$$\widehat{Y} = X^T \widehat{W} \quad X \in \mathbb{R}^{d+1}$$

Γραμμική παλινδρόμηση

- Πως υπολογίζουμε τις παραμέτρους του μοντέλου μας;

$$RSS(w) = \sum_{i=1}^n (Y_i - X_i^T W)^2$$

$$RSS(w) = (Y - X^T W)(Y - X^T W)$$

Using normal equations

$$w = (X^T X)^{-1} X^T y$$

Γιατί αυτή και όχι κάποια άλλη,
συνάρτηση κόστους π.χ:

Least Absolute Deviation

$$\sum_{i=1}^n \|Y_i - X_i^T W\|_1$$



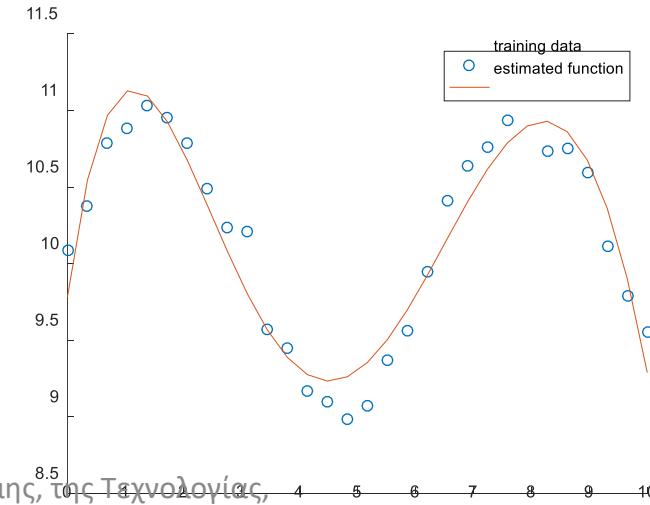
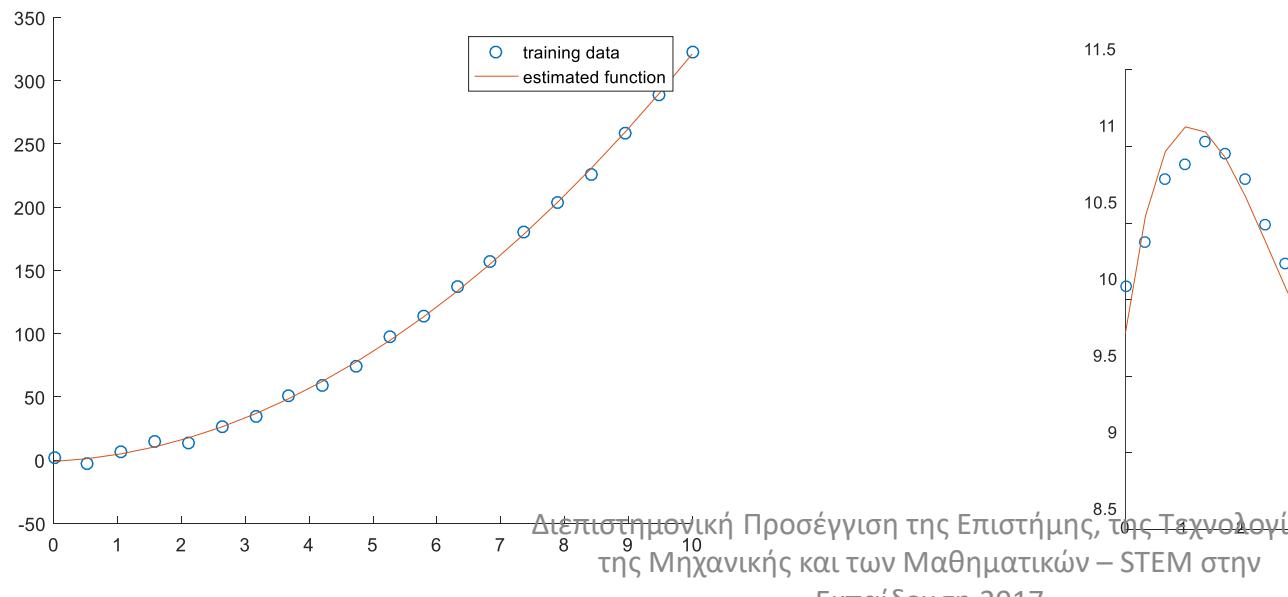
LSE	LAD
Επιρρεπής στους outliers	Εύρωστη στους outliers
Σταθερή λύση και μοναδική	Μη σταθερή και πολλές λύσεις

Γραμμική παλινδρόμηση



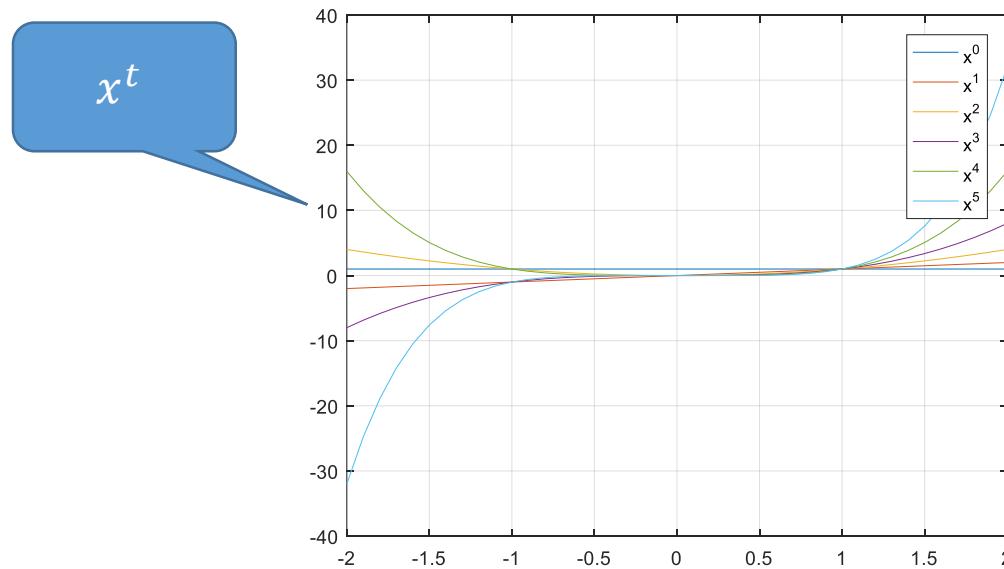
- Μη γραμμική σχέση εισόδου εξόδου;
- Επεκτείνουμε τη βάση μας προσθέτοντας όρους πολυωνύμου
- Γραμμική επέκταση βάσης

$$y(w, x) = w_0 + w_1 \cdot x_1 + w_2 \cdot {x_1}^2 + \dots + w_M \cdot {x_1}^M$$

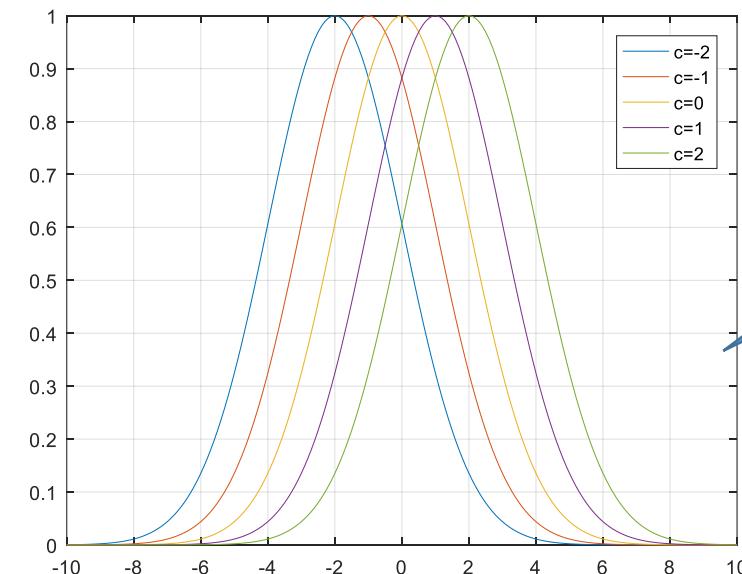


Γραμμική παλινδρόμηση

- Επεκτείνουμε τη βάση μας προσθέτοντας όρους πολυωνύμου
- Τι άλλες βάσεις μπορούμε να χρησιμοποιήσουμε;



x^t



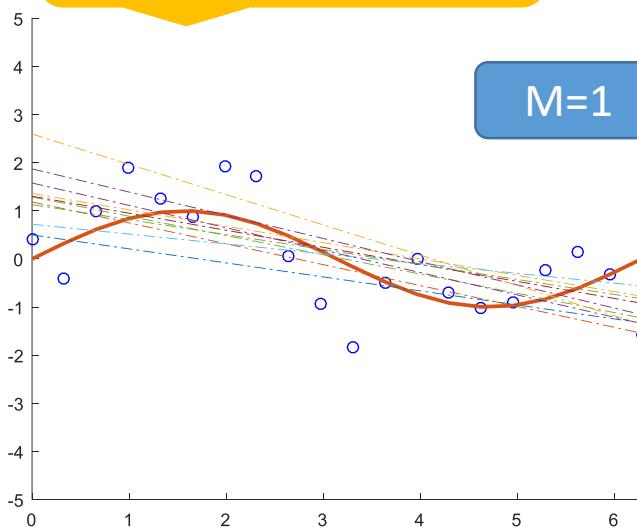
*Radial Basis
functions*

Bias-Variance Tradeoff

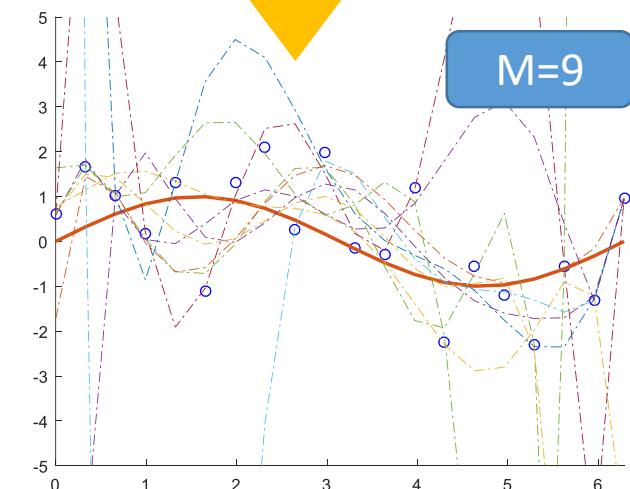
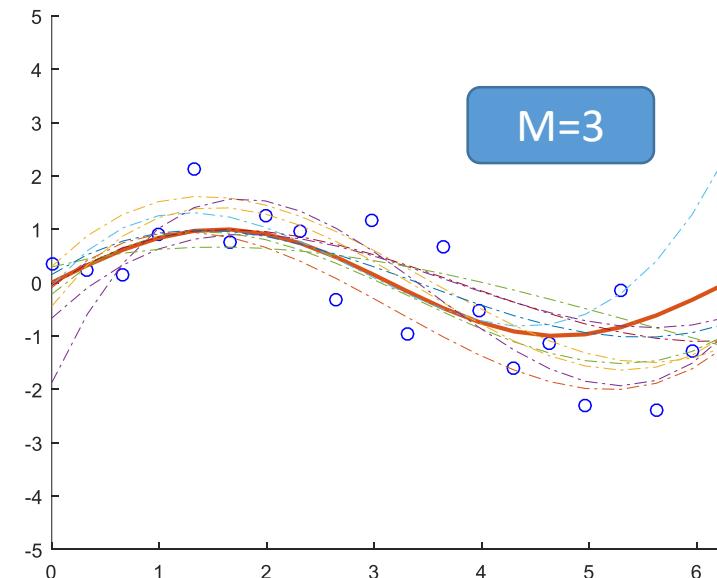


- Σχέση πολυπλοκότητας μοντέλου- σφάλματος
- Τα μισά δείγματα για εκπαίδευση. Πολλές επαναλήψεις

Λάθος υπόθεση για το
μοντέλο μας- BIAS

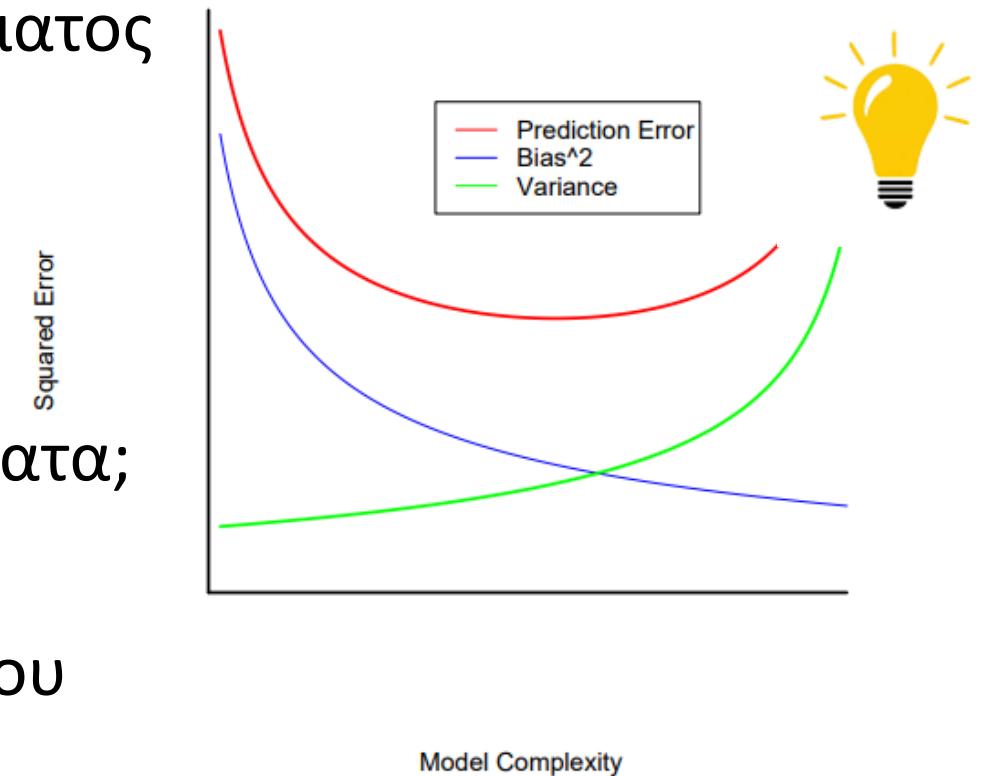


Μεγάλη
πολυπλοκότητα
μοντέλου- Variance



Bias-Variance Tradeoff

- Σχέση πολυπλοκότητας μοντέλου- σφάλματος
- Διαισθητική περιγραφή
- Πόσο καλά θα αναπαραστήσει νέα δείγματα;
- Η εκτίμηση των παραμέτρων του μοντέλου πρέπει να γίνεται με μεγάλη προσοχή



Regularization

- Τα μοντέλα με αυξημένη πολυπλοκότητα τείνουν να προσαρμόζονται ιδανικά στα δεδομένα εκπαίδευσης (increased variance)
- Μεγάλο σφάλμα στα δεδομένα δοκιμής
- Αυτή η κατάσταση μπορεί να αντιμετωπιστεί εισάγοντας κάποιους περιορισμούς στην ενέργεια των παραμέτρων.
- **Δύο βασικές προσεγγίσεις**
 - Ridge Regression
 - Lasso Regression

Regularization



Ridge Regression

$$\hat{w} = \underset{w}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - X_i^T W)^2 + \lambda \|W\|_2$$

Reconstruction error Penalty Term

The diagram shows the Ridge Regression cost function. It consists of two main parts: a sum of squared differences between observed values \$Y_i\$ and predicted values \$X_i^T W\$, and a regularization term \$\lambda \|W\|_2\$. Blue curly braces under the first part are labeled "Reconstruction error" in orange, and blue curly braces under the second part are labeled "Penalty Term" in blue.

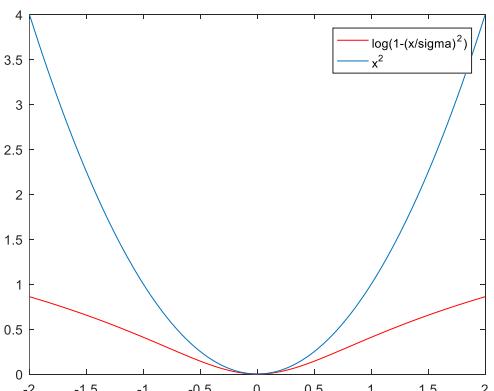
Lasso Regression

$$\hat{w} = \underset{w}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - X_i^T W)^2 + \lambda \|W\|_1$$

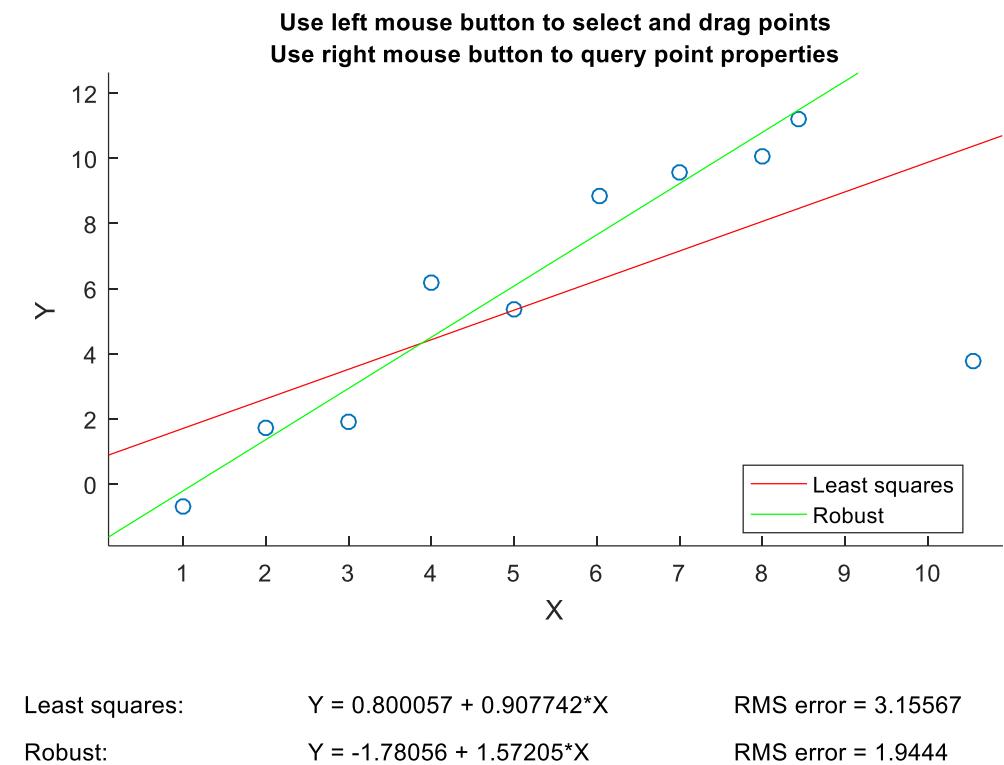
$$w = (\lambda I + X^T X)^{-1} X^T y$$

Robust statistics fitting- outliers

- Πως χειριζόμαστε τους outliers
- Outliers have large residuals
- Suppress them via reweighting
- Δείτε το “robustdemo” του Matlab
- Αλλαγή συνάρτησης κόστους

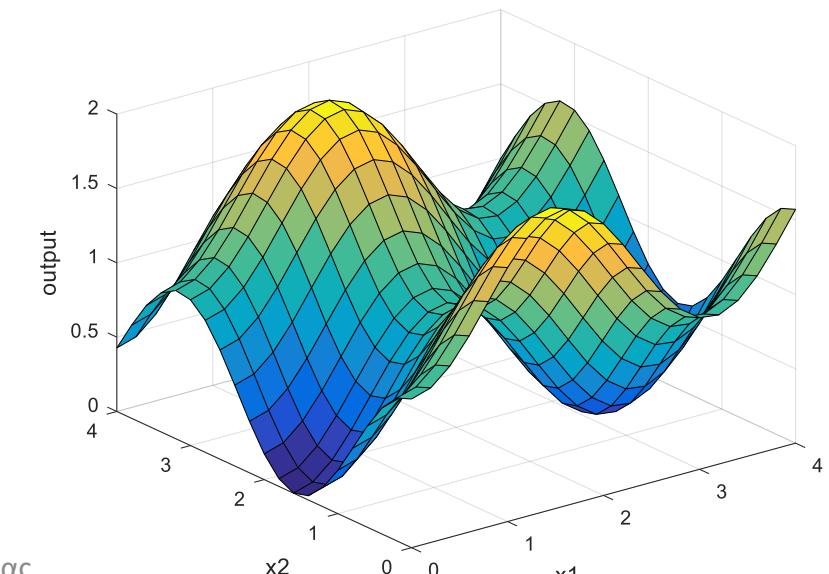
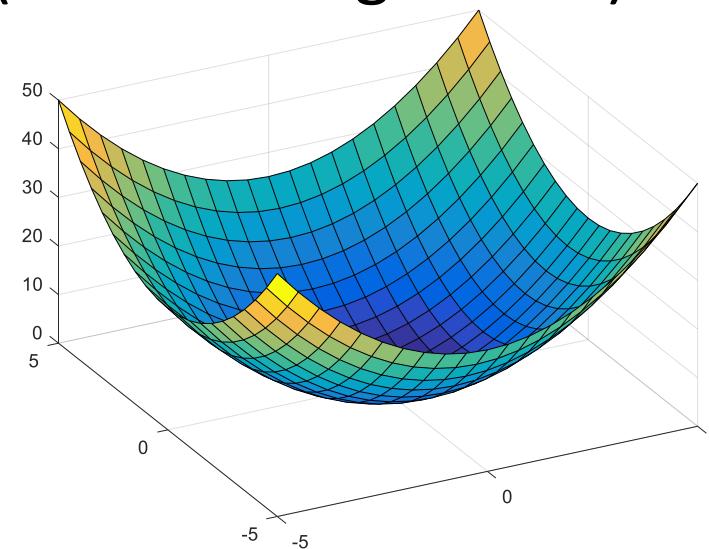
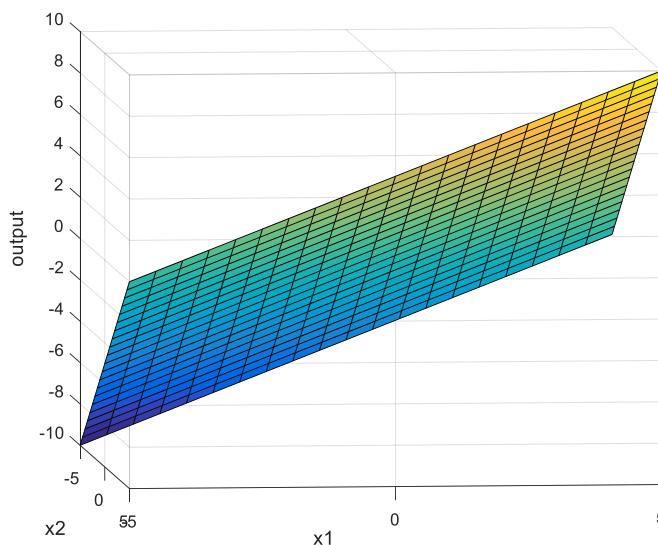


Lorentzian
-Δεν υπάρχει
αναλυτική λύση



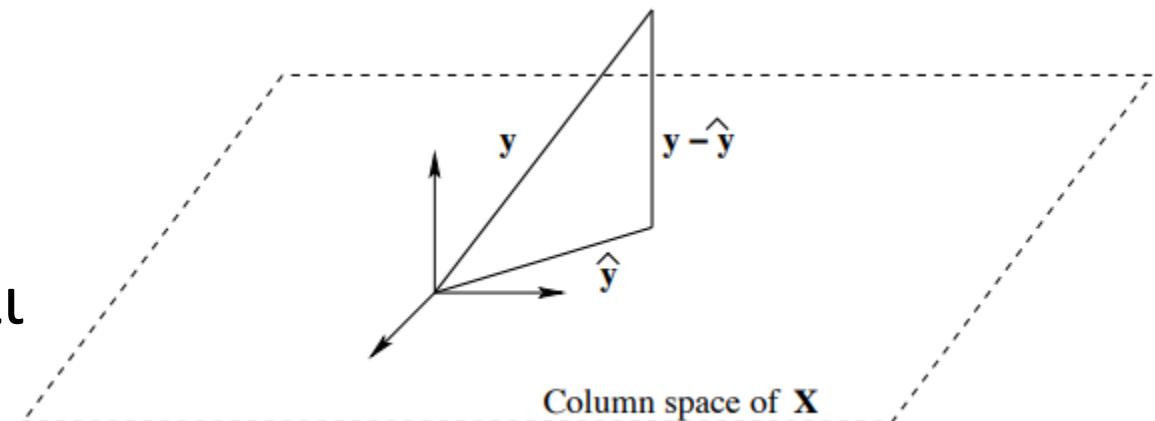
Γραμμική παλινδρόμηση πολλών μεταβλητών (multivariate regression)

- Μέχρι τώρα ασχοληθήκαμε με συναρτήσεις μίας μεταβλητής.
- Όπως αναφέραμε μπορεί να έχουμε πολλαπλές εισόδους ή εξόδους
- Θα παρουσιάσουμε εδώ ένα παράδειγμα όπου η μεταβλητή εισόδου μας ανήκει στον R^2 (bivariate regression)



Γραμμική παλινδρόμηση πολλών μεταβλητών

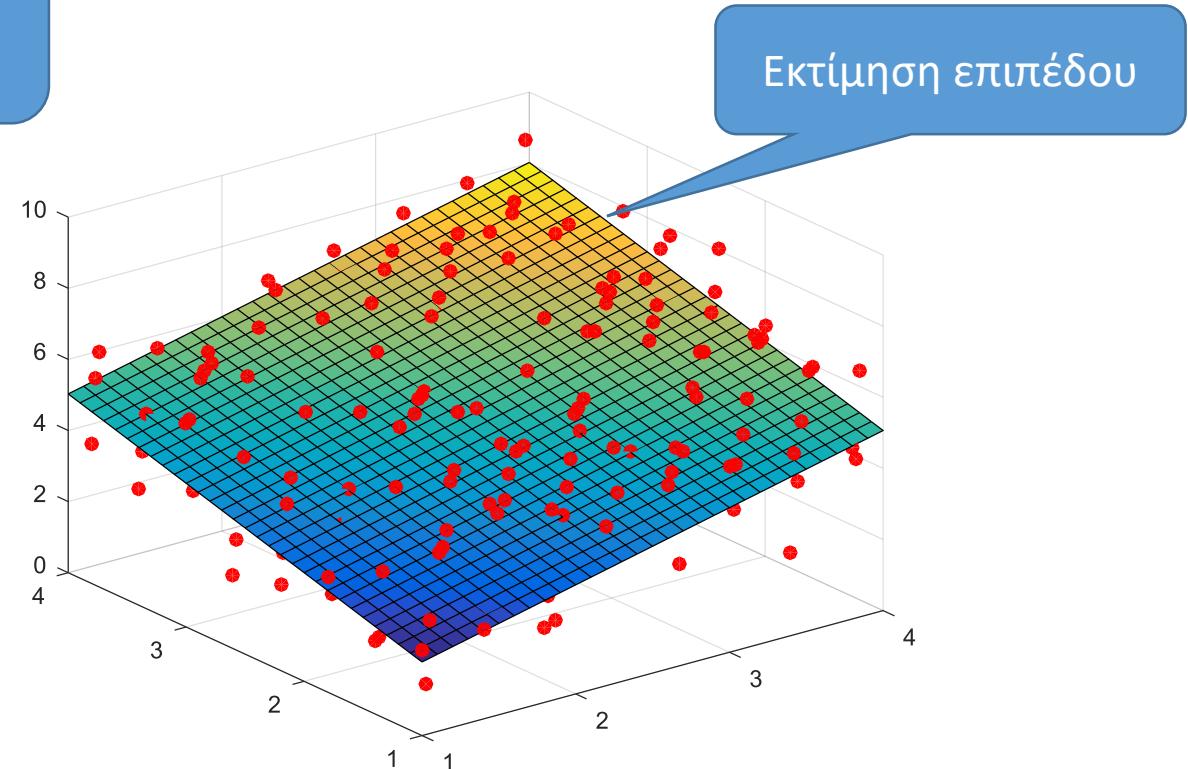
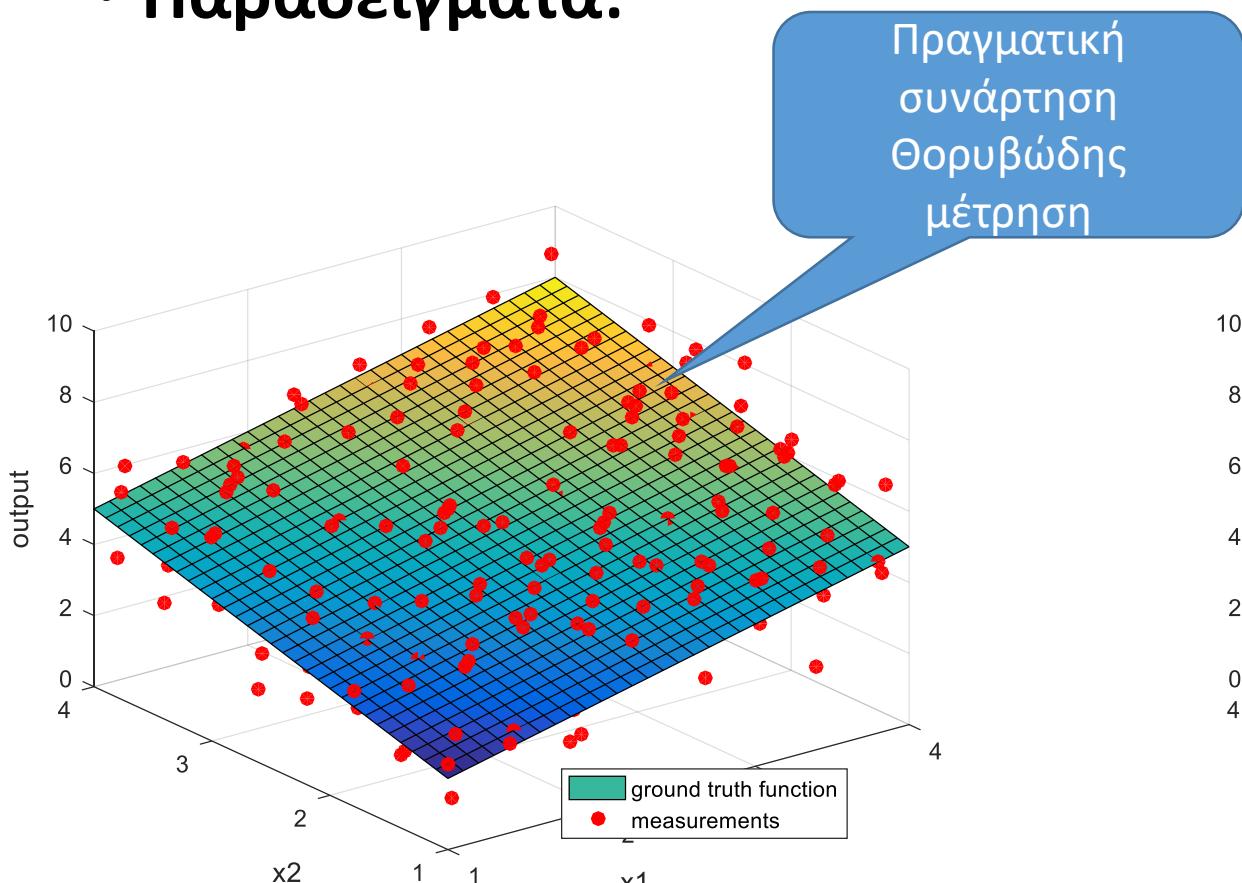
- Γεωμετρική αναπαράσταση
- Έχουμε πει ότι η λύση δίνεται για:
- $w = (X^T X)^{-1} X^T y$
- Η εκτίμηση ενός νέου σημείου:
- $\hat{y} = Xw = X(X^T X)^{-1} X^T y$
- Ο όρος $H = X(X^T X)^{-1}$ ονομάζεται “hat” πίνακας.



Using normal equations- Τα υπόλοιπα (residuals) είναι κάθετα στην επιφάνεια

Γραμμική παλινδρόμηση πολλών μεταβλητών

- ## • Παραδείγματα:



Τεχνικές βελτιστοποίησης

Εισαγωγή στις τεχνικές βελτιστοποίησης

- Όλα τα προβλήματα μηχανικής εκμάθησης έχουν κάποια κοινά χαρακτηριστικά
- Προσπαθούν να ικανοποιήσουν ένα κριτήριο προσαρμόζοντας ένα σύνολο παραμέτρων
- Οι παράμετροι που ικανοποιούν το κριτήριο υπολογίζονται επιλύοντας ένα πρόβλημα βελτιστοποίησης

Εισαγωγή στις τεχνικές βελτιστοποίησης

- Πολλές φορές ζητείται να βρεθεί το ελάχιστο/μέγιστο μιας συνάρτησης
- Στα προβλήματα μηχανικής εκμάθησης, συνήθως αναζητάμε τις παραμέτρους που μεγιστοποιούν ή ελαχιστοποιούν μια συνάρτηση κόστους/πιθανότητας.
- Η λύση των παραμέτρων μπορεί πολλές φορές να βρεθεί με αναλυτικό τρόπο
 - Ωστόσο στις περισσότερες περιπτώσεις έχουμε μη γραμμική σχέση των δεδομένων εισόδου εξόδου
 - Επίσης πολλές φορές ο αναλυτικός τρόπος δεν είναι υπολογιστικά ελκυστικός (μεγάλες βάσεις)

Παρουσίαση του προβλήματος

- Η μηχανική εκμάθηση είναι άρρηκτα συνδεδομένη με την έννοια της βελτιστοποίησης
- Αυτό που επιθυμούμε είναι να ελαχιστοποιήσουμε (ή να μέγιστοποιήσουμε) μια συνάρτηση αναφορικά με ένα σύνολο παραμέτρων.
- Μας επιτρέπει να προσεγγίσουμε με υπολογιστικά ελκυστικό τρόπο προβλήματα όπου ο αριθμός των δειγμάτων είναι μεγάλος ή δεν υπάρχει αναλυτική λύση

Πρόβλημα βελτιστοποίησης

Διατύπωση της λύσης του:

$$RSS(w) = \sum_{i=1}^n (Y_i - X_i^T W)^2$$

Ως πρόβλημα βελτιστοποίησης:

$$\hat{w} = \underset{w}{\operatorname{argmin}} \sum_i^n \|Y_i - X_i^T W\|^2$$

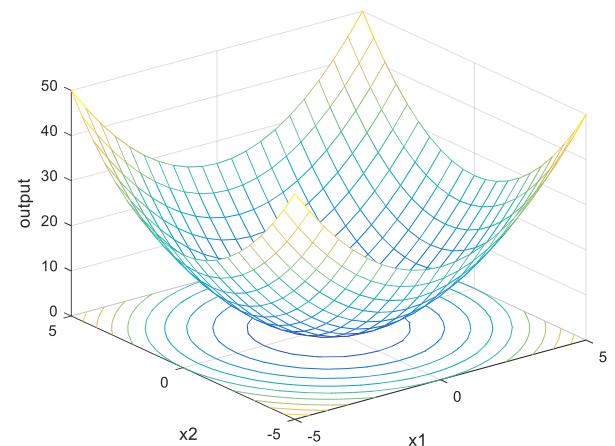
Ως πρόβλημα βελτιστοποίησης με περιορισμούς (constrained):

$$\hat{w} = \underset{w}{\operatorname{argmin}} \sum_i^n \|Y_i - X_i^T W\|^2, \text{ s.t. } \|W\| \leq t$$

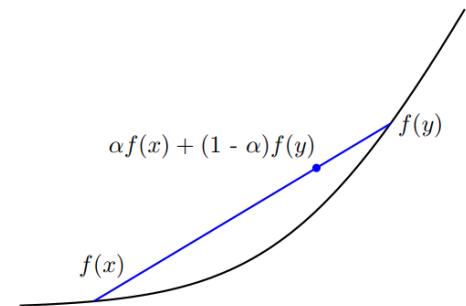
Convex optimization

- Least Squares Loss → Convex
 - *twice differentiable with positive semi-definite matrix*
- Τι μορφή έχει η συνάρτηση κόστους μας;
- Πως μπορούμε να βρούμε τις παραμέτρους που ελαχιστοποιούν την σχέση:

$$\hat{w} = \underset{w}{\operatorname{argmin}} \sum_i^n \|Y_i - X_i^T W\|^2$$



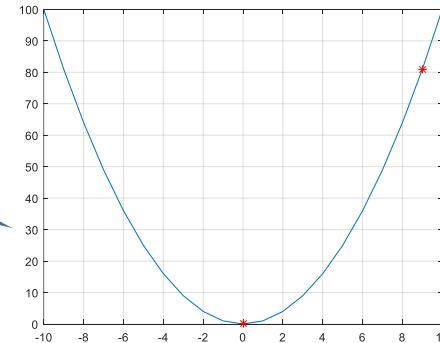
$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$



Newton-Raphson

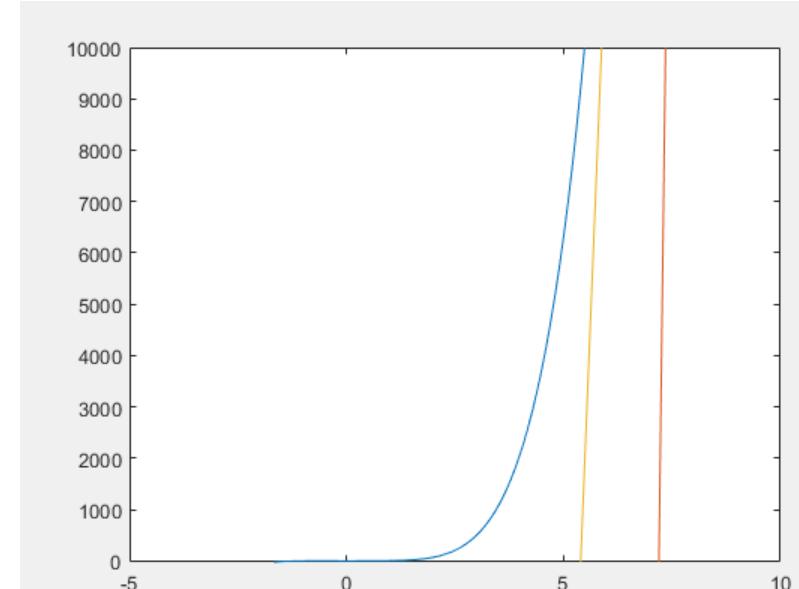
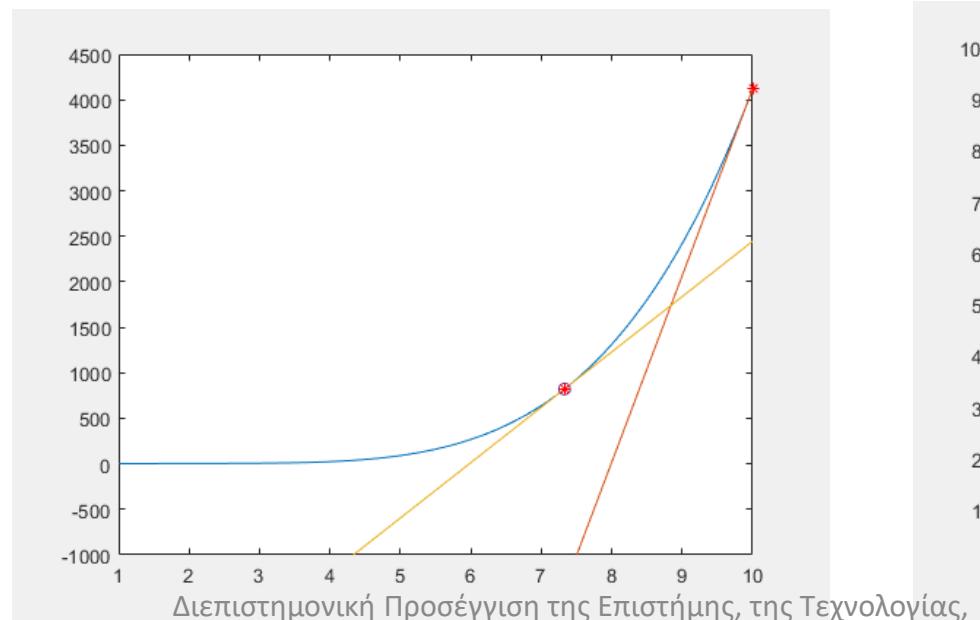
- Επαναληπτική διαδικασία
- Second order method
- Χρειαζόμαστε μια **συνάρτηση κόστους** την **1^η και 2^η παράγωγο** της

If quadratic,
converges in one
step



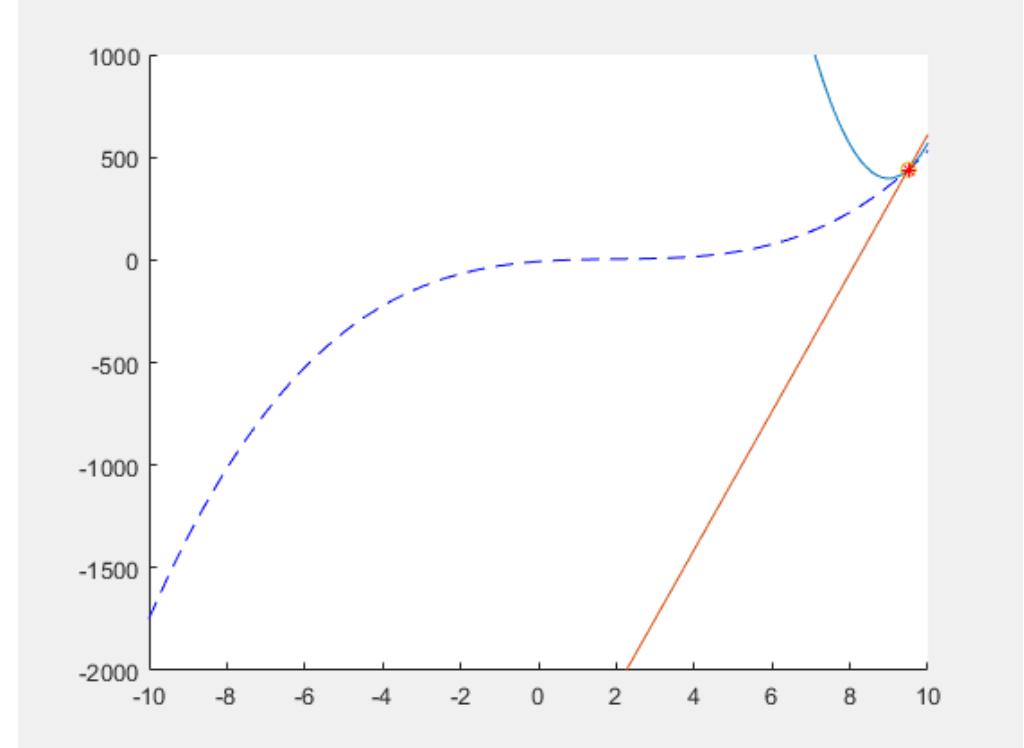
Find the roots of
the derivative

$$x_{n+1} = x_n - \frac{f'(x)}{f''(x)}$$



Newton-Raphson

- Χρησιμοποιεί μια τετραγωνική προσέγγιση της συνάρτησης στο σημείο και ελαχιστοποιεί αυτή
- Η γαλάζια γραμμή είναι η προσέγγιση
- Το ελάχιστο της μας δίνει το επόμενο σημείο
- Το ίδιο ισχύει και για >1 διαστάσεις



Newton-Raphson

- Επαναληπτική διαδικασία
- Second order method
- Χρειαζόμαστε μια **συνάρτηση κόστους την 1^η και 2^η παράγωγο της**

1 dimension

$$x_{n+1} = x_n - \frac{f'(x)}{f''(x)}$$

Higher dimensions

$$x_{n+1} = x_n - (H(x))^{-1} \nabla(x)$$

Inverting the
Hessian is not
easy

Approximation, or
Quasi Newton
Methods

Gradient Descent

- Μέθοδος που βασίζεται μόνο στην πρώτη παράγωγο
- Χρειαζόμαστε μια **συνάρτηση κόστους (objective function)** και την **παράγωγο της**

Για την περίπτωση της ελαχιστοποίησης του αθροίσματος των διαφορών του τετραγώνου

Η συνάρτηση κόστους (objective function)

$$J(w_j) = \frac{1}{2n} \sum_{i=1}^n (f(x_{i,j}) - y_i)^2$$

Η πρώτη παράγωγος

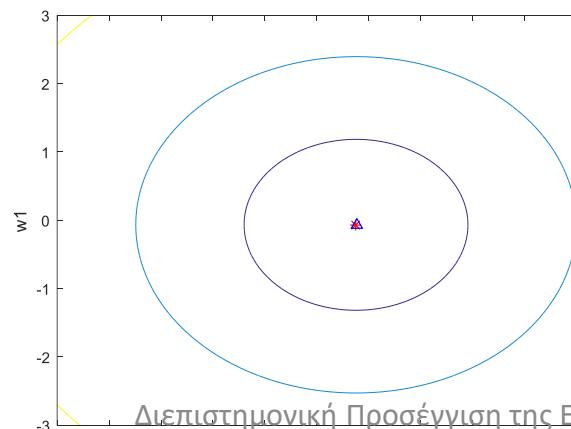
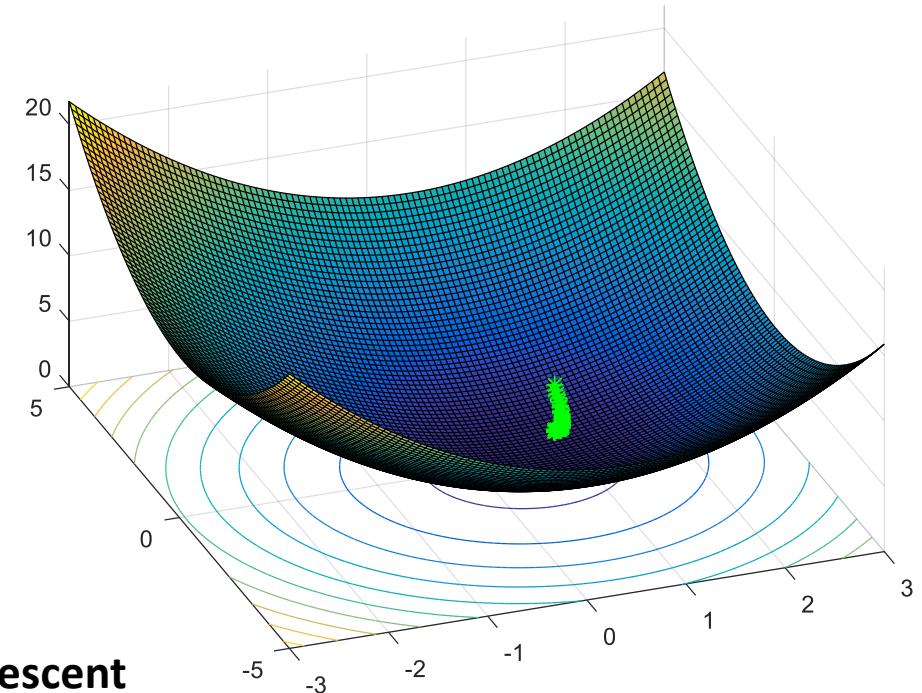
$$g(w_j) = \frac{1}{n} \sum_{i=1}^n (f(x_{i,j}) - y_i)x_{i,j}$$

Gradient Descent

Η συνάρτηση κόστους

$$J(w_j) = \frac{1}{2n} \sum_{i=1}^n (f(x_{i,j}) - y_i)^2$$

Σύγκριση αναλυτικής λύσης με Gradient Descent



Gradient Descent

- First order iterative method

Έστω το παρακάτω πρόβλημα

$$f(x) = w_0 1 + w_1 x_1 + \cdots + w_d x_d = w^T x$$

Η συνάρτηση κόστους

$$J(w) = \frac{1}{2n} \sum_{i=1}^n (y_{i,j} - f(x_{i,j}))^2$$

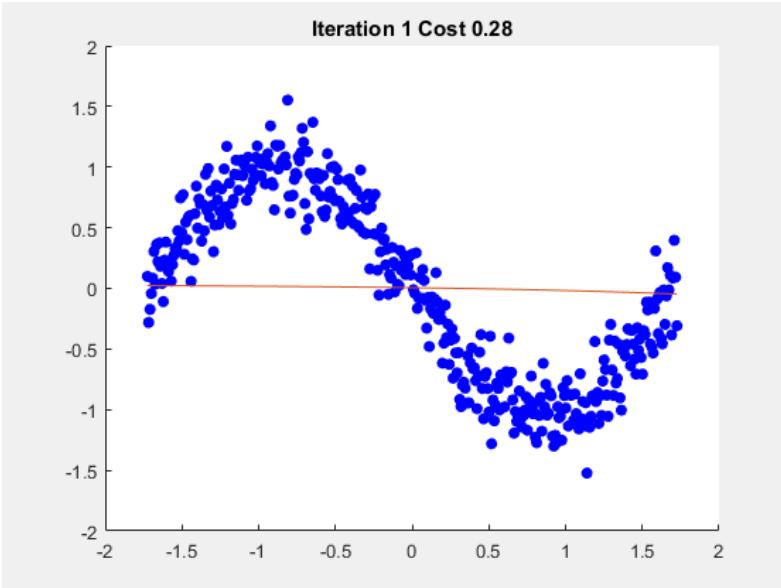
Ενημέρωση Παραμέτρων

$$w_j := w_j - a \frac{1}{n} \sum_{i=1}^n (y_{i,j} - f(x_{i,j})) x_{i,j}$$

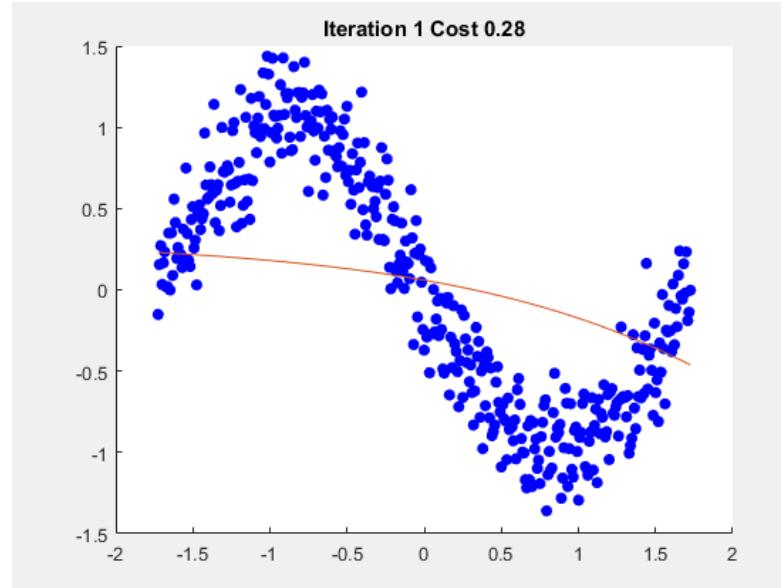
Για όλους τους άξονες j

Gradient Descent

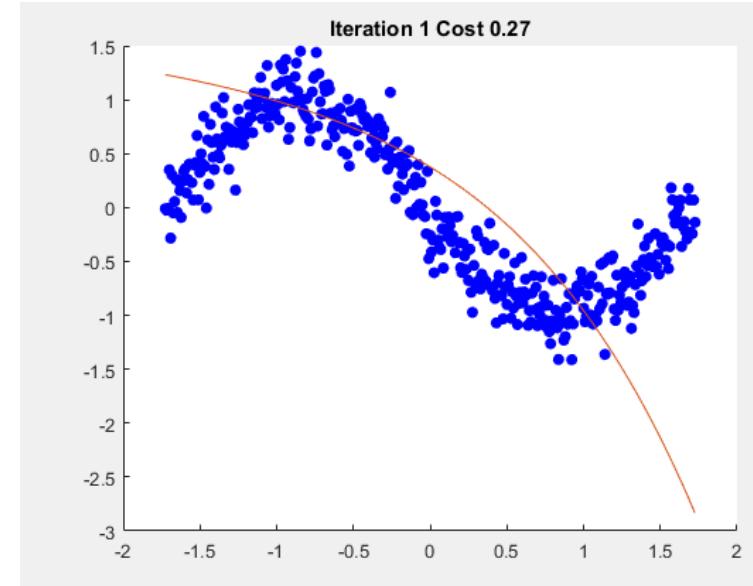
Learning Rate = 0.01



Learning Rate = 0.1



Learning Rate = 1



Online Learning Learning

- Όταν τα δείγματα είναι πάρα πολλά δεν μπορούμε να τα φορτώνουμε όλα στη μνήμη
- Για το σκοπό αυτό προτάθηκε η μέθοδος “Stochastic Gradient Descent”
- Σύμφωνα με αυτή, τα δεδομένα έρχονται σε μικρές ομάδες (≥ 1) και χρησιμοποιούνται για να υπολογίσουμε τα νέα βάρη

(stochastic) Gradient Descent

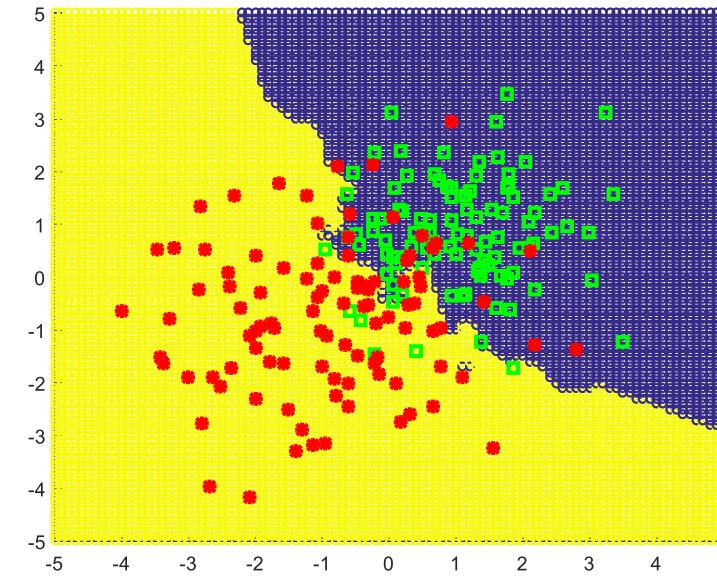
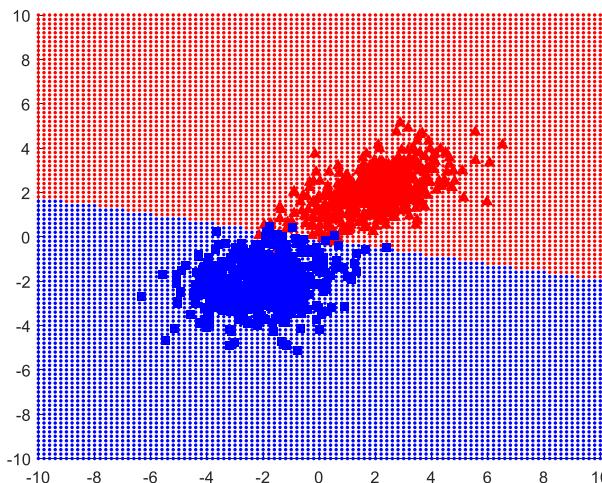
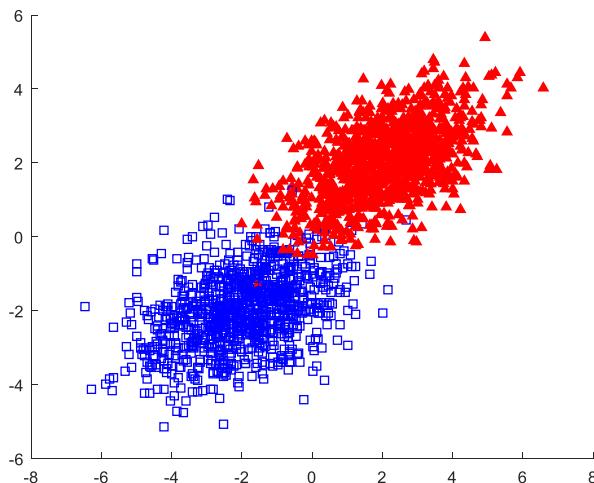
- Ρυθμός εκμάθησης
 - Αποτελεί μια πολύ κρίσιμη παράμετρο
 - Χρειάζεται να υπολογιστεί εμπειρικά
 - Μπορεί να εγκλωβιστούμε σε τοπικά ελάχιστα
- Μικρές προσθήκες βελτιώνουν σημαντικά τα αποτελέσματα της μεθόδου
 - Εισαγωγή momentum, Nesterov Accelerated Gradient
 - ADAM
 - AdaGrad
 - RMSprop

Ταξινόμηση

Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας,
της Μηχανικής και των Μαθηματικών – STEM στην
Εκπαίδευση 2017

Προβλήματα Ταξινόμησης

- Παρουσίαση του προβλήματος
- Να εντοπίσουμε την ή τις περιοχές που διαχωρίζουν δύο ή περισσότερους πληθυσμούς



Προβλήματα Ταξινόμησης

- **Παρουσίαση του προβλήματος**
- Εδώ θα ασχοληθούμε με προβλήματα όπου η έξοδος του συστήματος θα αποτελεί categorical variables
- Τα δεδομένα παρέχονται σε μορφή $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, όπου $y \in \{0, \dots, K\}$ για K μεταβλητές και $x \in \mathbb{R}^d$ για d -διάστατα δεδομένα
- Ταξινόμηση σε δύο βασικές κατηγορίες
 - **Discriminant functions:** Logistic Regression, Generative Discriminant Analysis
 - Modeling Decision boundary (Perceptron, SVM)



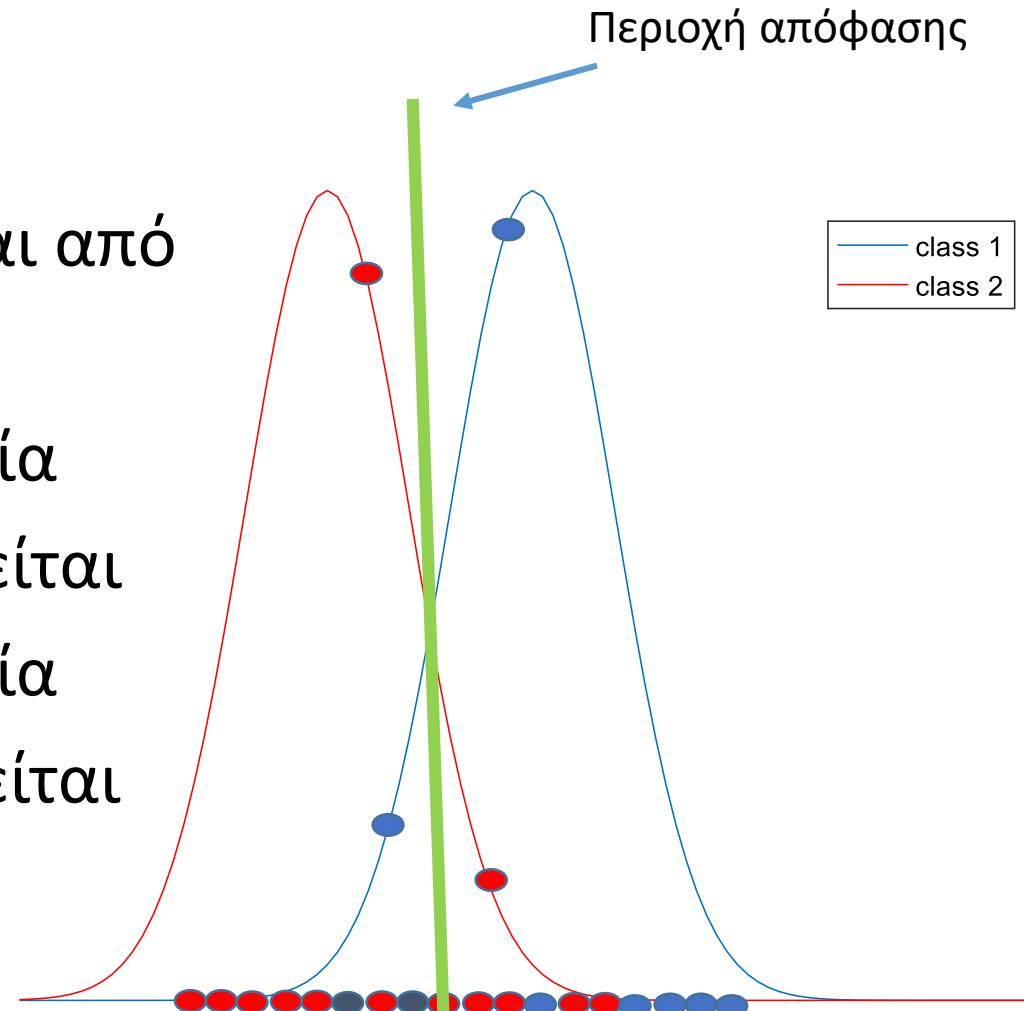
Προβλήματα Ταξινόμησης

- Δυο βασικές προσεγγίσεις
- Θεωρούμε πως τα δεδομένα ακολουθούν κάποια κατανομή (π.χ. Κανονική κατανομή)
 - Naive Bayes classification
 - Gaussian discriminant analysis
 - Logistic Regression
- Περιπτώσεις που δεν κάνουμε κάποια θεώρηση για την κατανομή από την οποία παρήχθησαν τα δεδομένα
 - Perceptron
 - Multilayer Linear Networks
 - Support Vector Machines



Bayes Classifier

- Θεωρούμε ότι τα δεδομένα προέρχονται από κανονική κατανομή
- Κάθε δείγμα ταξινομείται στην κατηγορία για την οποία η πιθανότητα μεγιστοποιείται
- Κάθε δείγμα ταξινομείται στην κατηγορία για την οποία η πιθανότητα μεγιστοποιείται



k-NN classifier

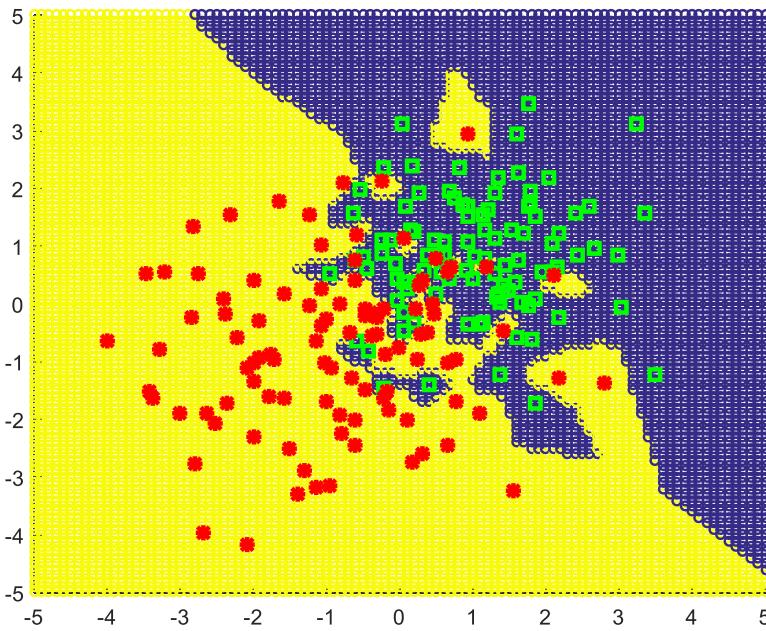
- Ταξινομητής k - εγγύτερων γειτόνων
- Κάθε δείγμα ταξινομείται στην κατηγορία εκείνη που ανήκουν οι k πλησιέστεροι γείτονες του
- Τοπική μέθοδος- Ικανοποιητικά αποτελέσματα για πολλά δείγματα και μικρό αριθμό διαστάσεων
- Δεν λειτουργεί καλά όταν η διάσταση των χαρακτηριστικών μεγαλώνει

k-NN classifier

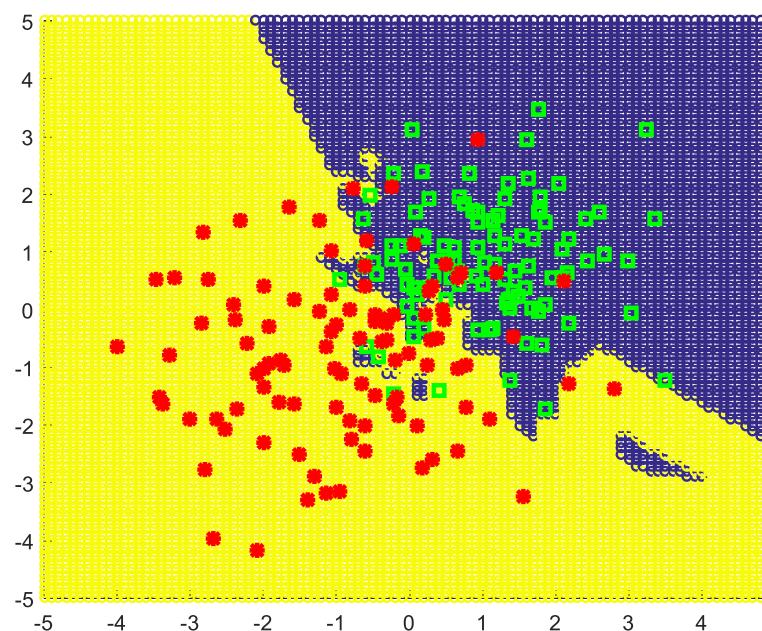
- Παραδείγματα

Αριθμός γειτόνων

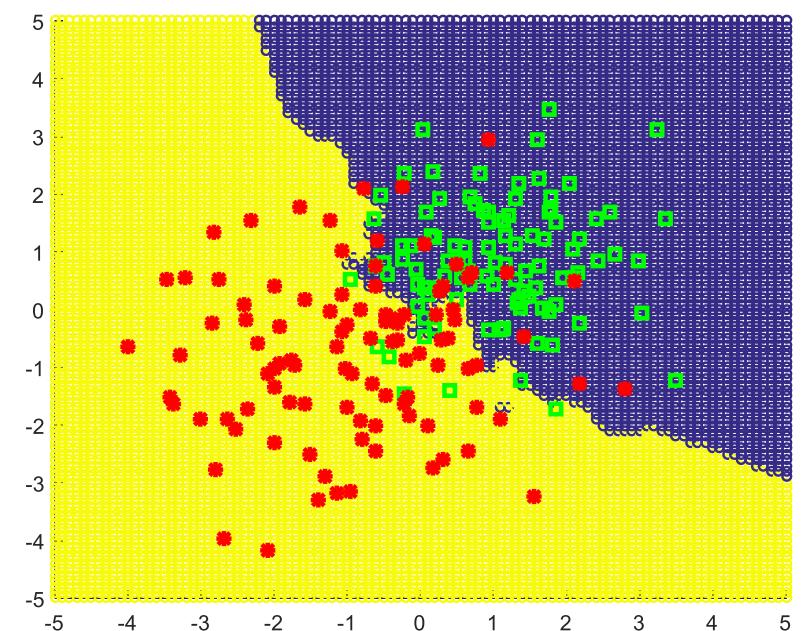
K=1



K=3



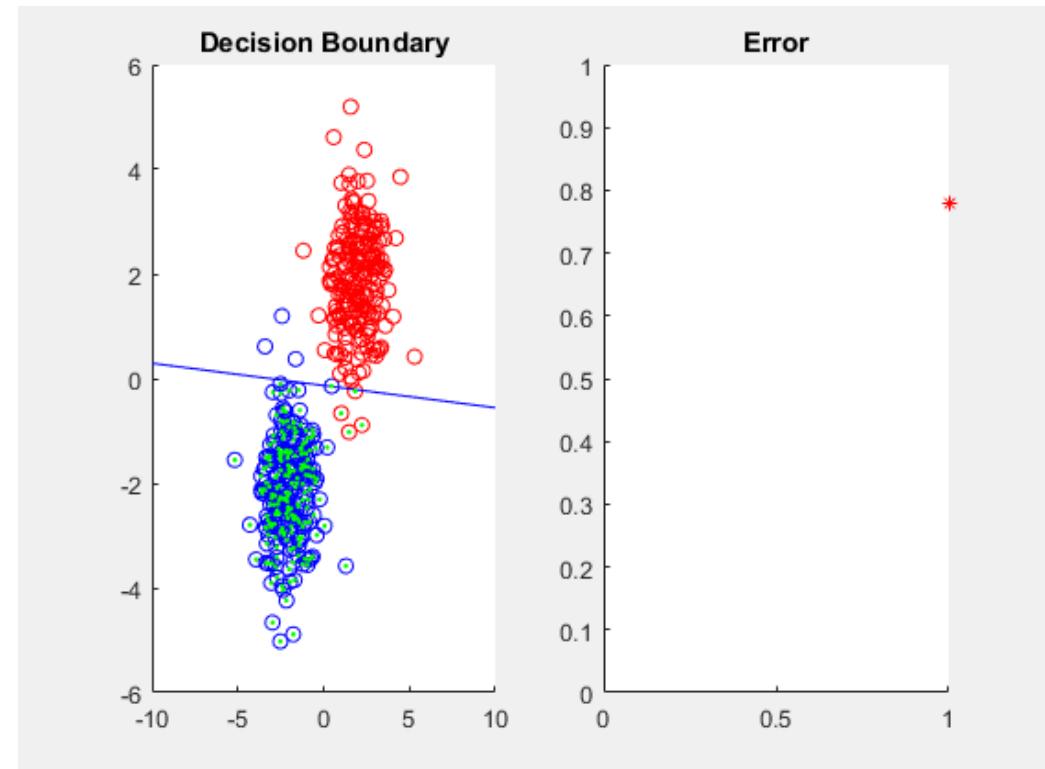
K=5



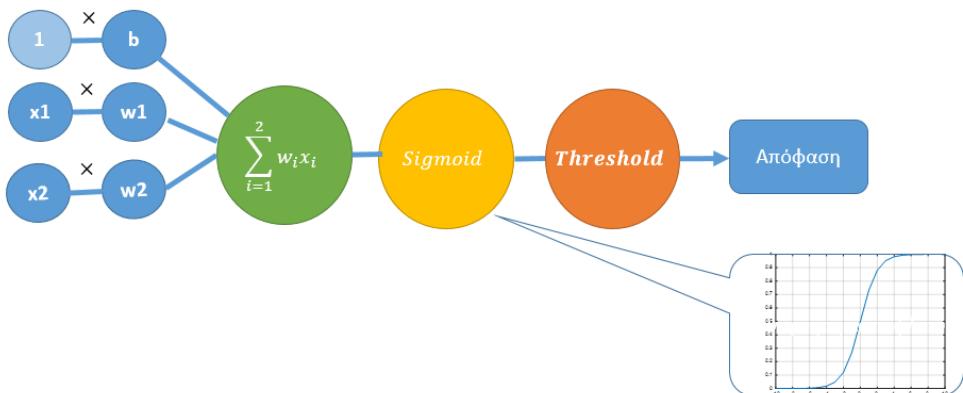
Perceptron

- Η έξοδος (ανάθεση ενός δείγματος x) δίνεται από την παρακάτω έκφραση

$$y = \text{sing}\left(\sum_{i=1}^d w_i x_i\right)$$



- Κατά την εκπαίδευση κάθε εσφαλμένη ταξινόμηση τροποποιεί τα βάρη



- Online Learning rule (perceptron rule)
- Stochastic gradient descent
- Ρυθμός εκμάθησης
- Απαιτεί γραμμικά διαχωρίσιμα δεδομένα

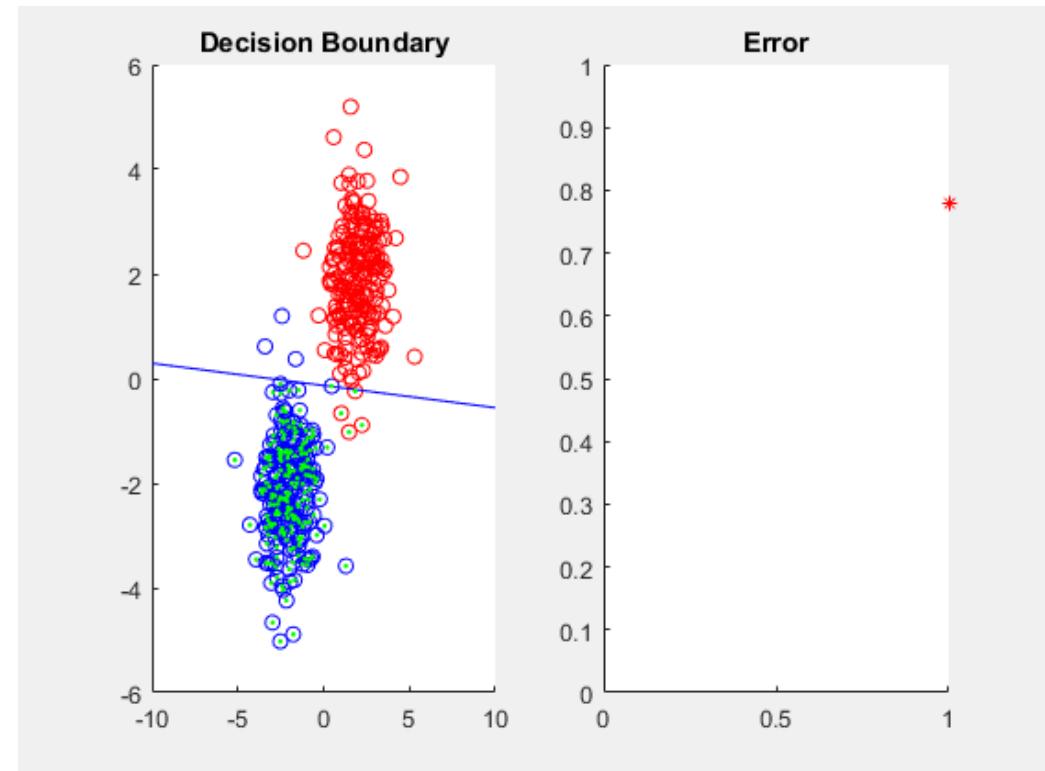


Perceptron

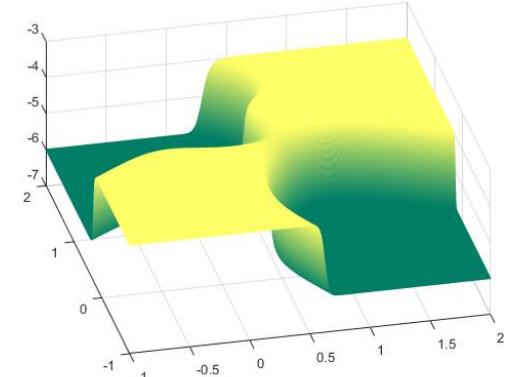
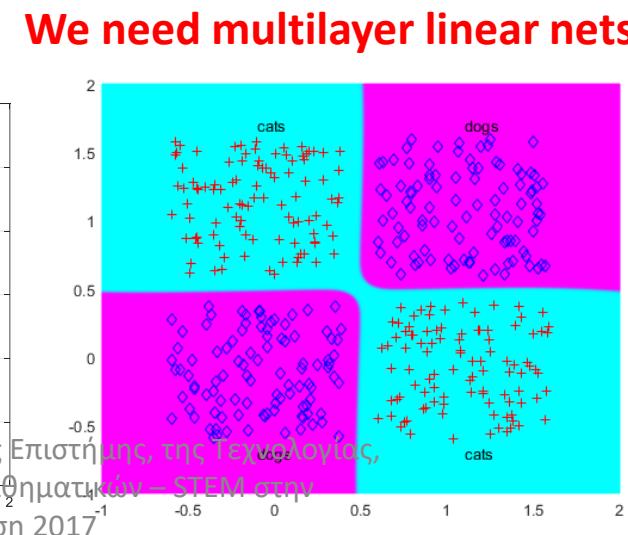
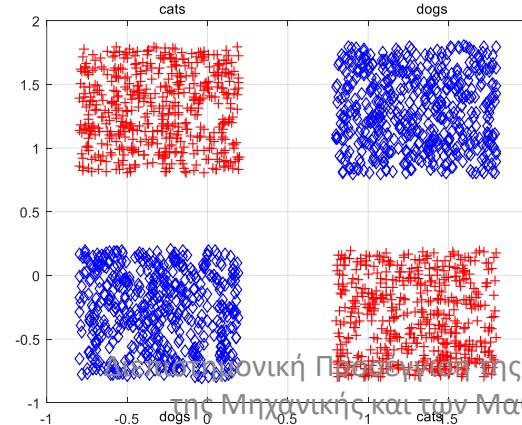
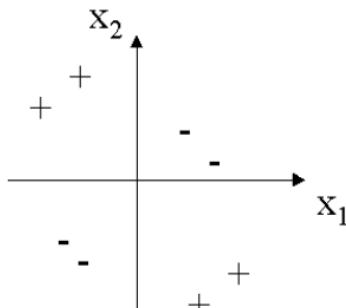
- Η έξοδος (ανάθεση ενός δείγματος x) δίνεται από την παρακάτω έκφραση

$$y = \text{sign}(\sum_{i=1}^d w_i x_i)$$

- Online Learning rule (delta rule)
- Stochastic gradient descent
- Ρυθμός εκμάθησης
- Απαιτεί γραμμικά διαχωρίσιμα δεδομένα



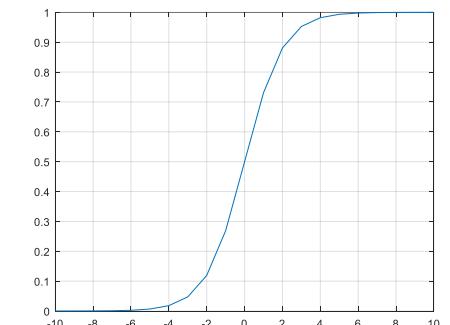
XOR Problem



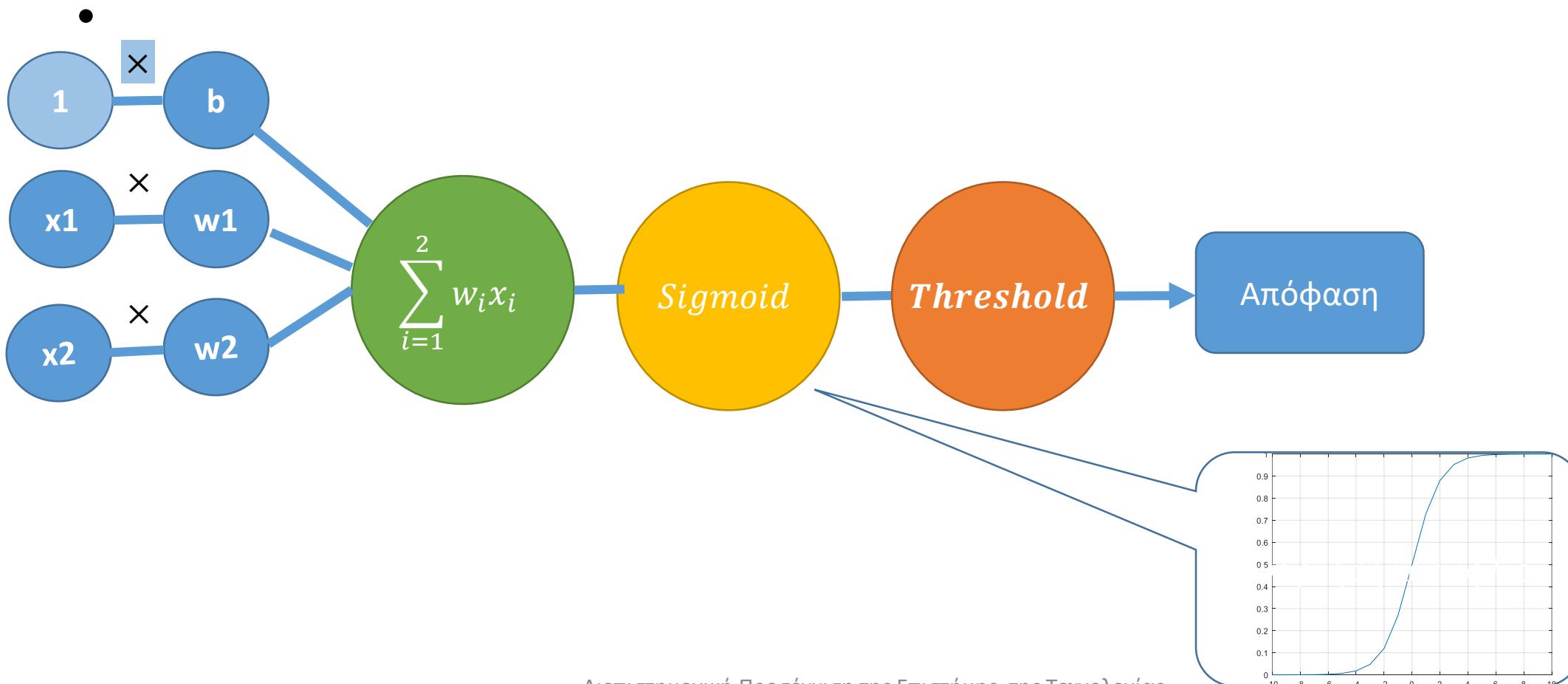
Logistic Regression- Binary Classification

- Εφαρμογή μιας μη γραμμικότητας στην **έξοδο του μοντέλου γραμμικής παλινδρόμησης** ακολουθούμενη από μια συνάρτηση κατωφλίου
- Η συνάρτηση αυτή απεικονίζει τα δεδομένα (οριζόντιος άξονας) στο εύρος $[0,1]$ (κάθετος άξονας)
- Αναθέτει πιθανότητες στα δεδομένα μας

$$f(w^T x) = \frac{1}{1 + e^{-w^T x}} = \frac{e^{w^T x}}{1 + e^{w^T x}}$$



Logistic Regression- Binary Classification



Logistic Regression

- Εφαρμογή μιας συνάρτησης κατωφλίου στην έξοδο του μοντέλου γραμμικής παλινδρόμησης- μη γραμμικότητα
- Πλέον έχουμε ένα πρόβλημα **μη-γραμμικό** δεν υπάρχει αναλυτική λύση

$$f(w^T x) = h_w(x) = \frac{1}{1 + e^{-w^T x}}, \text{όπου } x \in \mathbb{R}^{d+1}$$

- Έχουμε μια μη γραμμική σχέση, και δύο καταστάσεις εξόδου.
- Πως λύνουμε αυτό το πρόβλημα;
 - **Υπάρχουν δύο βασικές προσεγγίσεις**

Logistic Regression- Ελαχιστοποίηση κόστους

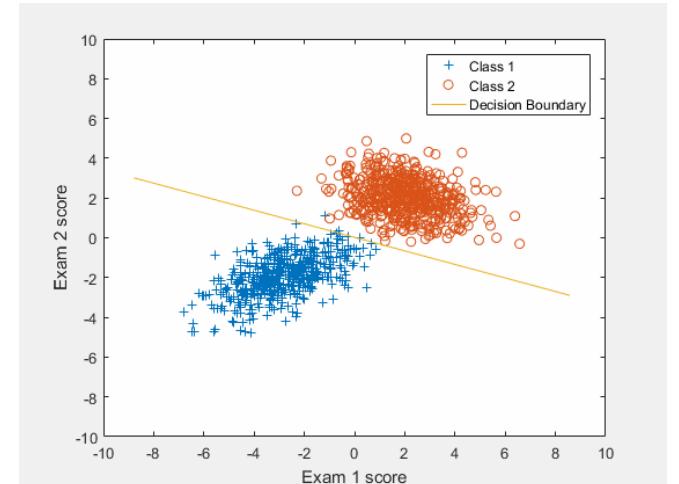
- Ελαχιστοποίηση συνάρτησης κόστους με Newton-Raphson
- Θέλουμε τη **συνάρτηση κόστους** καθώς και την 1^η και 2^η **παράγωγο της (Hessian)**

$$\text{Cost}(w) = \frac{1}{n} \left(\sum_{i=1}^n -y_i (\log(h_w(x_i))) - (1 - y_i) (\log 1 - (h_w(x_i))) \right)$$

$$\bullet \frac{\partial \text{Cost}(w)}{\partial w_j} = \frac{1}{n} \sum_{i=1}^n (h_w(x_i) - y) x_i$$

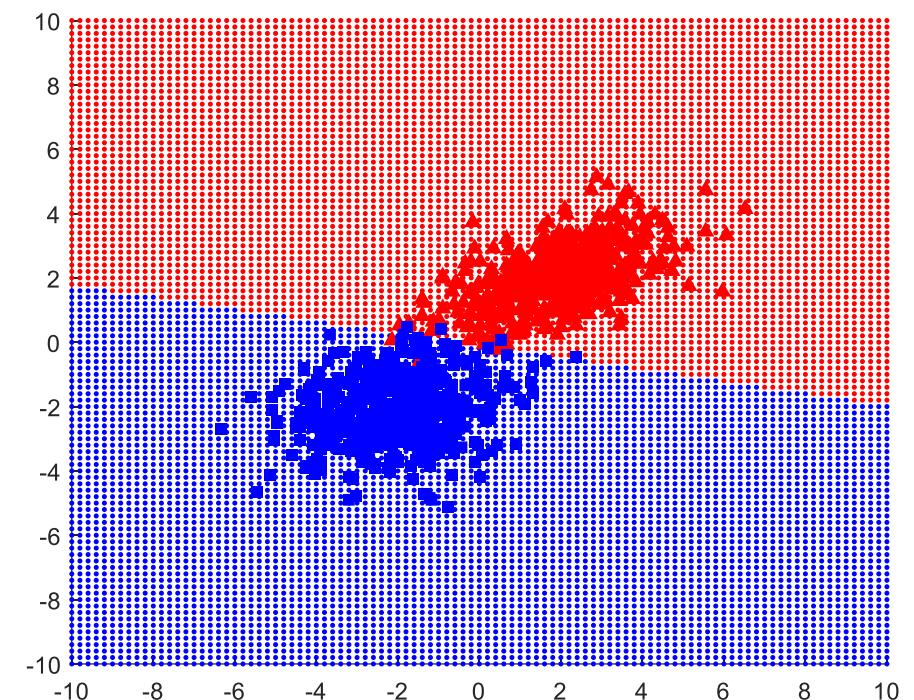
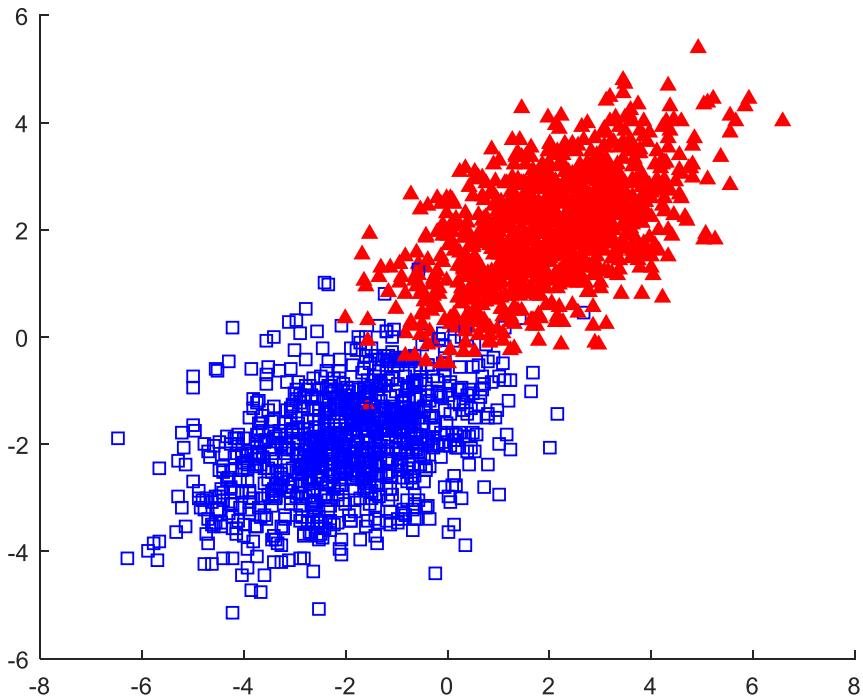
$$\bullet H_{ij} = \frac{\partial^2 \text{loss}(w)}{\partial w_i \partial w_j}$$

Ενημέρωση Παραμέτρων
 $w_j := w_j - H^{-1} \nabla_w \text{loss}(w)$



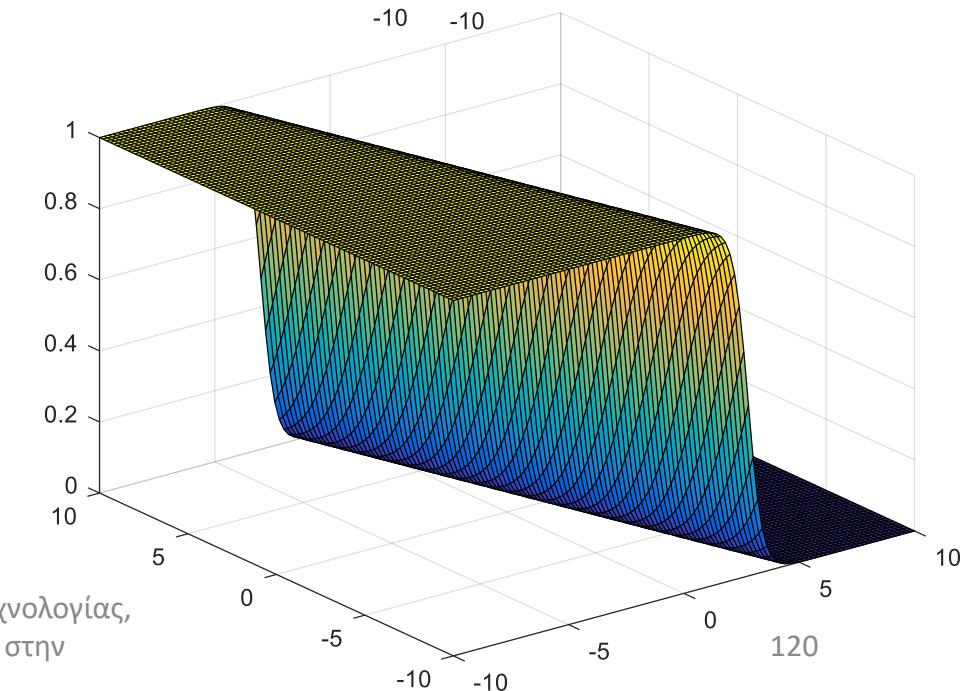
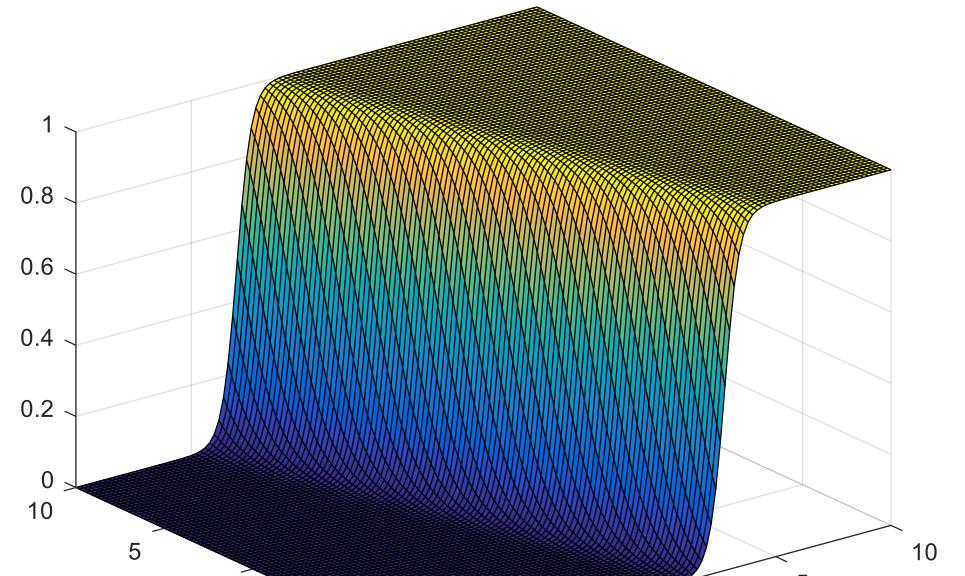
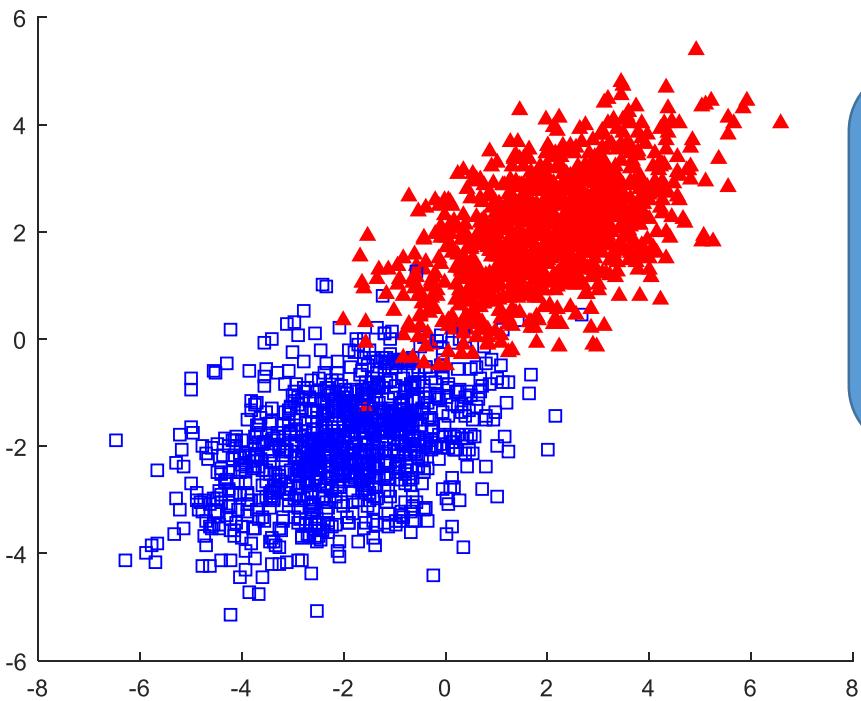
Logistic Regression

- Παράδειγμα με δύο κλάσεις στον \mathbb{R}^2



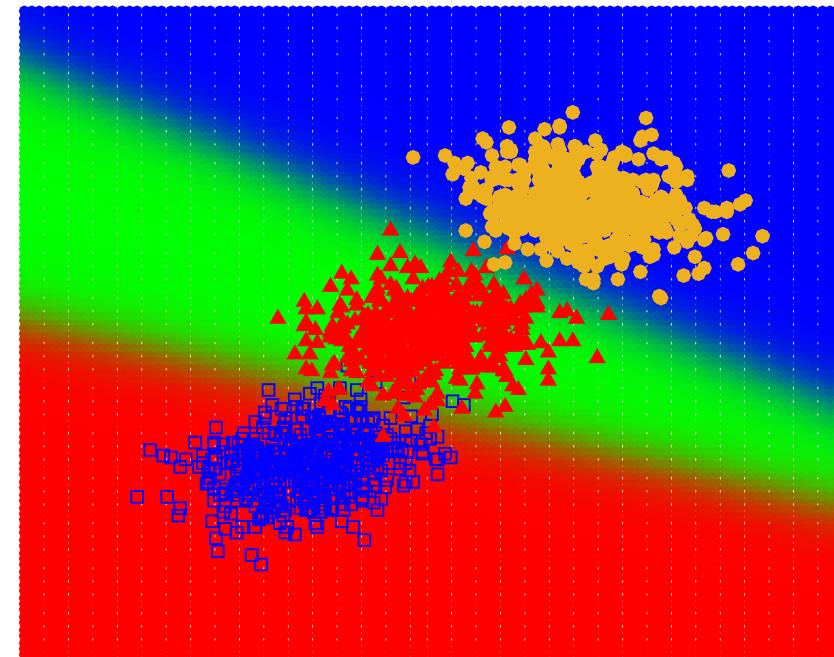
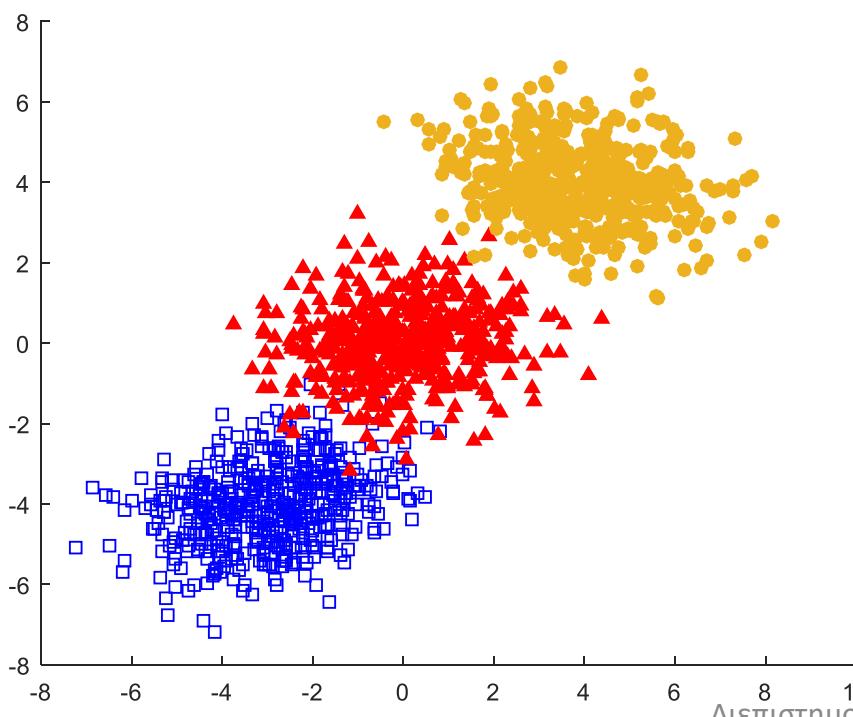
Logistic Regression

- Παράδειγμα με δύο κλάσεις στον \mathbb{R}^2



Multinomial Logistic Regression

- Γενίκευση της μεθόδου Logistic Regression (LR)
- Η μέθοδος SoftMax regression → Γενίκευση της LR για $K > 2$



Multinomial Logistic Regression

- Η μέθοδος SoftMax regression → Γενίκευση της LR για K>2
- Αντικατάσταση της Sigmoid με τη Softmax
- One-hot Vectors

$$s_{w_i}(x) = \frac{e^{w_i^T x}}{\sum_{i=1}^K e^{w_i^T x}}, i = 0, \dots, K$$

From Logistic Regression to Gaussian Discriminant

- Θέλουμε οι παράμετροι μας να μεγιστοποιούν την παρακάτω πιθανότητα

$$p(y|x; w) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y - w^t x)^2}{2\sigma^2}\right)$$



- Γιατί να μην μοντελοποιήσουμε απευθείας τις κλάσεις ως Gaussian κατανομές και στη συνέχεια να αναθέτουμε κάθε δείγμα στην κατηγορία με την μεγαλύτερη πιθανότητα;

Linear Discriminant Analysis

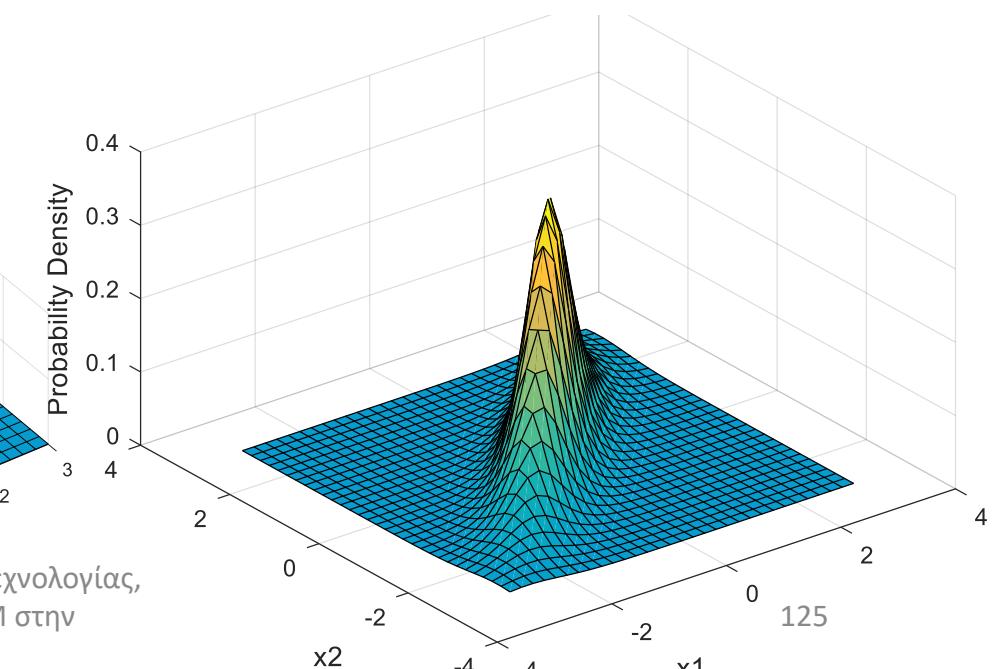
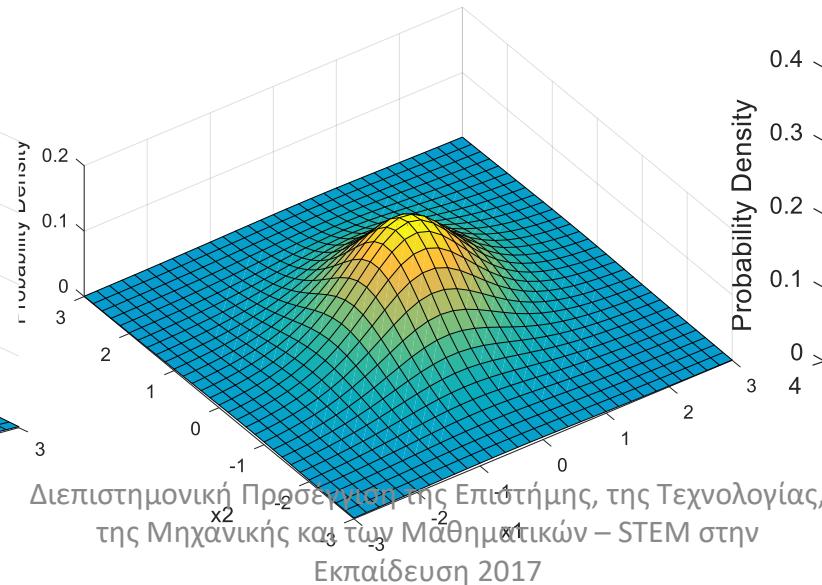
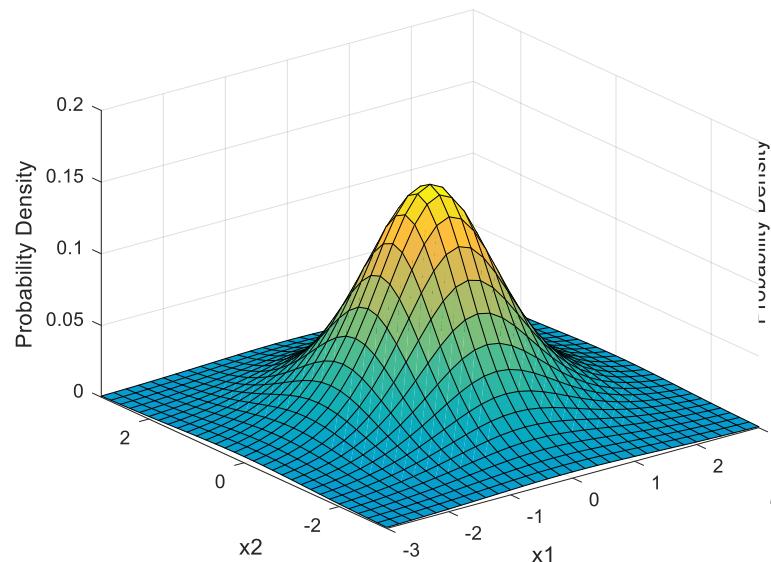
- **Generative Learning Method**
 - Δηλαδή, μοντελοποιούμε **απευθείας** τη διαδικασία από την οποία παρήχθησαν τα δεδομένα
- Εκτίμηση παραμέτρων με την μέθοδο Maximum Likelihood Estimation
- Απευθείας υπολογισμός της μέσης τιμής και της διακύμανσης
- Σε περισσότερες από μια διαστάσεις έχουμε πίνακα Συνδιακύμανσης (Covariance Matrix)
- Σφαιρικές κατανομές → Γραμμικές περιοχές απόφασης

Linear Discriminant Analysis

- Κάποια βασικά στοιχεία
- Multivariate Normal distribution

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \right).$$

- Examples of 2-D Gaussians:



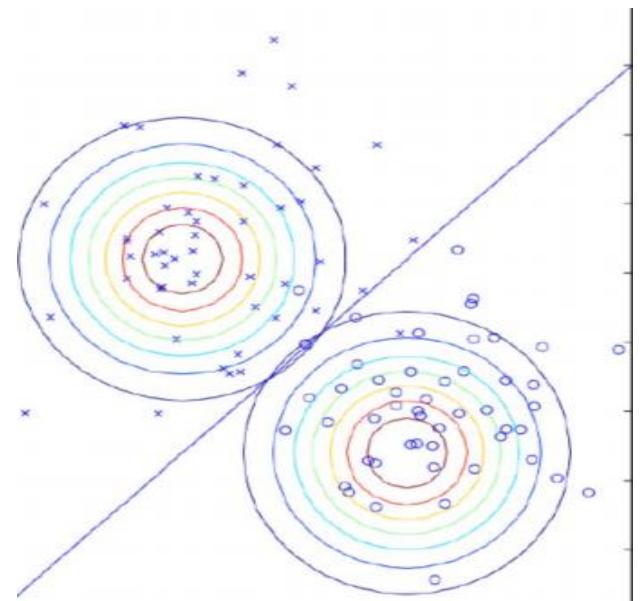
Linear Discriminant Analysis

- Ταξινόμηση δείγματος σε μια κατηγορία:

Για ένα σύνολο κατηγοριών $y = \{1, \dots, K\}$, το δείγμα $x \sim N(\mu, \Sigma)$ θα ανατεθεί στην κατηγορία εκείνη για την οποία μεγιστοποιείται η posterior probability $P(y|x)$.

$$\operatorname{argmax}_y P(y|x) = \operatorname{argmax}_y \frac{P(x|y)P(y)}{P(x)}$$

If assuming all classes are independent \rightarrow Naïve Bayes Classifier



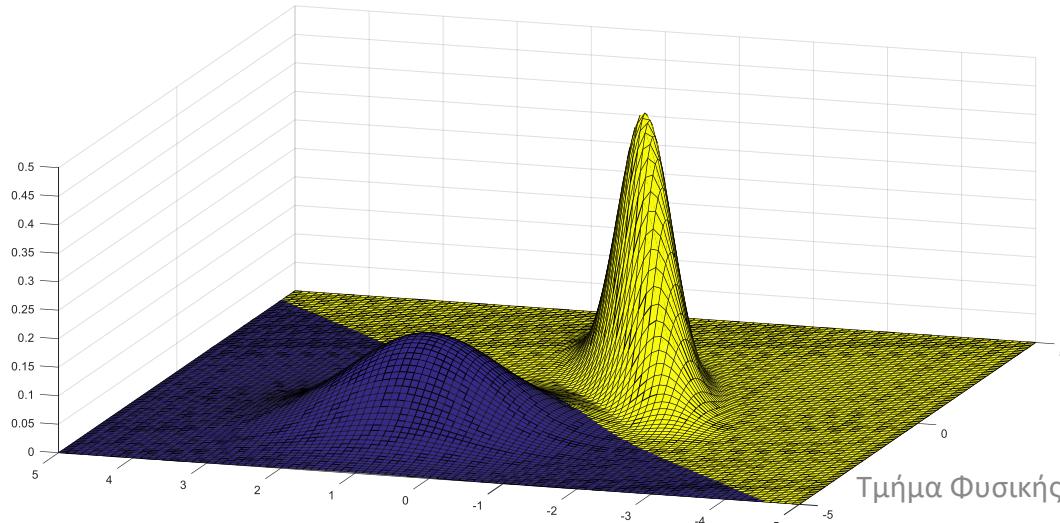
Quadratic Discriminant Analysis

- Τι σημαίνει όμως Linear και τι Quadratic Discriminant Analysis;
- Διαφορετικές υποθέσεις για τον πίνακα Σ

Linear

Υπόθεση: Όλες οι κατηγορίες έχουν ένα κοινό πίνακα Σ

- Παράγονται από την ίδια κατανομή με διαφορετική μεση τιμή
- Περιοχή απόφασης → γραμμική (μεσοκάθετος)

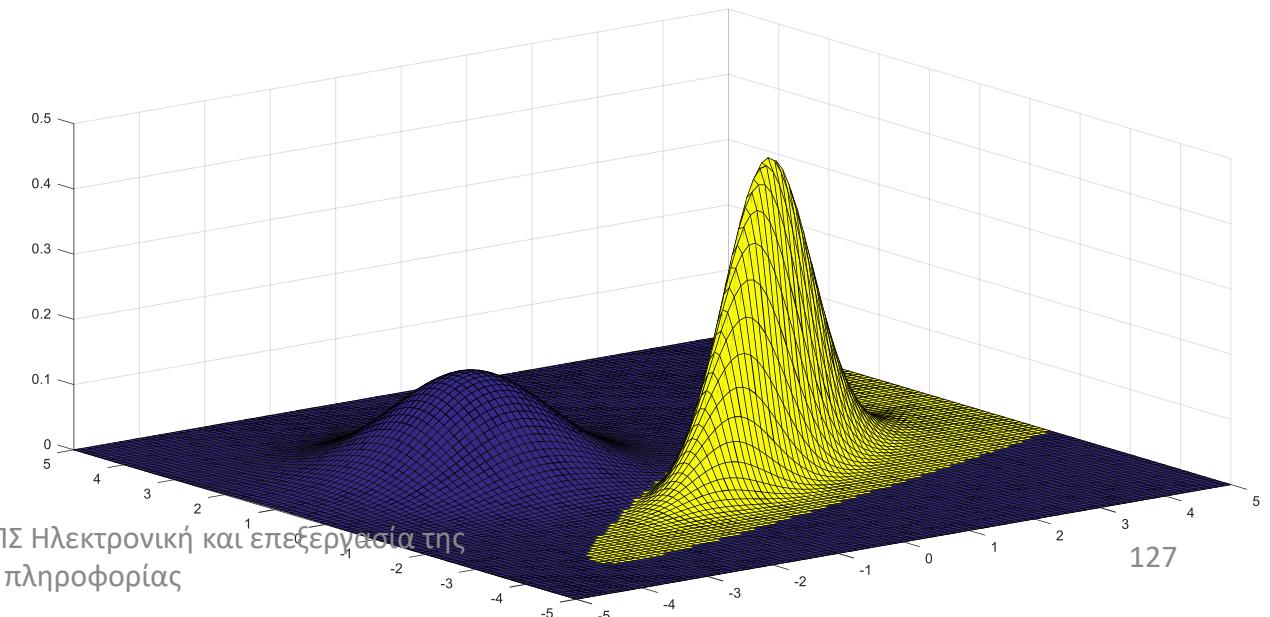


Τμήμα Φυσικής- ΔΜΠΣ Ηλεκτρονική και επεξεργασία της πληροφορίας

Quadratic

Υπόθεση: Κάθε κατηγορία έχει ένα δικό της Σ

- Περιοχή απόφασης μη γραμμική!



Ταξινόμηση- Kernel Methods- Support Vector Machines Regression and Classification

Kernels

- Εκφράζουν ομοιότητες (συγκρίσεις) μεταξύ των χαρακτηριστικών
- Έστω $X = \{x_1, x_2, x_3\}$ τρία διανύσματα στον \mathbb{R}^2
- $x_1 = [0.1 \ 0.2]$, $x_2 = [1.2 \ 0.8]$ and $x_3 = [-0.2 \ 0.3]$
- $k: X \times X \rightarrow \mathbb{R}$

0.0500	0.2800	0.0400
0.2800	2.0800	0.0000
0.0400	0.0000	0.1300
- Συνεπώς εάν σχεδιάσουμε αλγορίθμους που εργάζονται στον χώρο αυτό μπορούμε να εργαστούμε με δεδομένα οποιουδήποτε τύπου (κείμενο, διανύσματα, πίνακες, Γράφους κ.α.)



Kernels

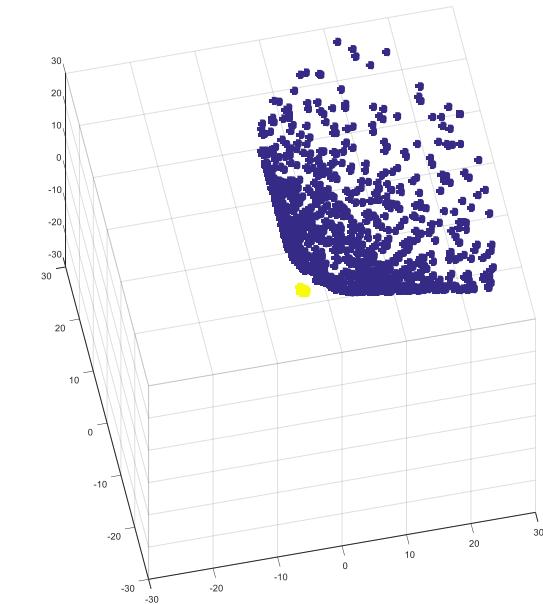
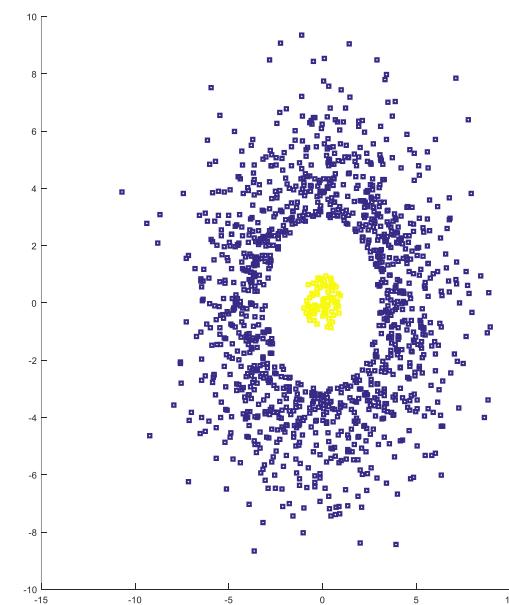
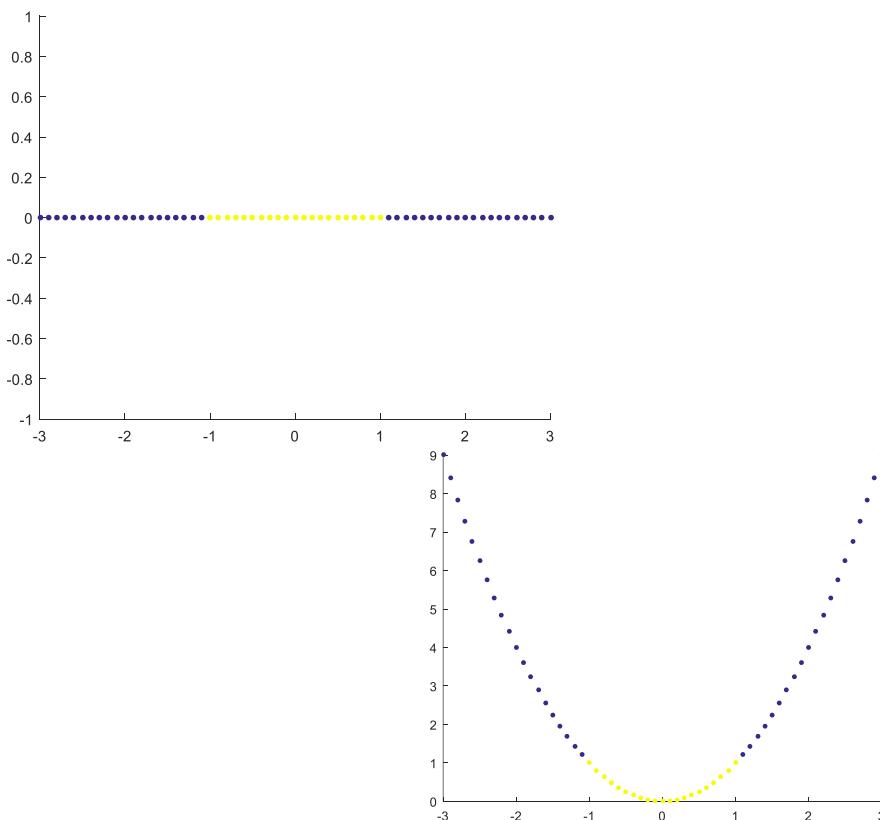
- Δημιουργούνται υπολογίζοντας το εσωτερικό γινόμενο μιας συνάρτησης (γραμμικής ή μη) των χαρακτηριστικών του αρχικού χώρου.
- Απεικονίζει τα διανύσματα στον \mathbb{R}
- Ορίζουμε λοιπόν μια συνάρτηση ως Kernel:

$$k: X \times X \rightarrow \mathbb{R}$$
$$k(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$$

$$\langle \varphi(x_i), \varphi(x_j) \rangle = \sum_{k=1}^d \varphi(x_{i,k}), \varphi(x_{j,k})$$

Ως μια απεικόνιση στο Hilbert Space F_k

Επιλογή παραμέτρων μοντέλου



Kernel Trick

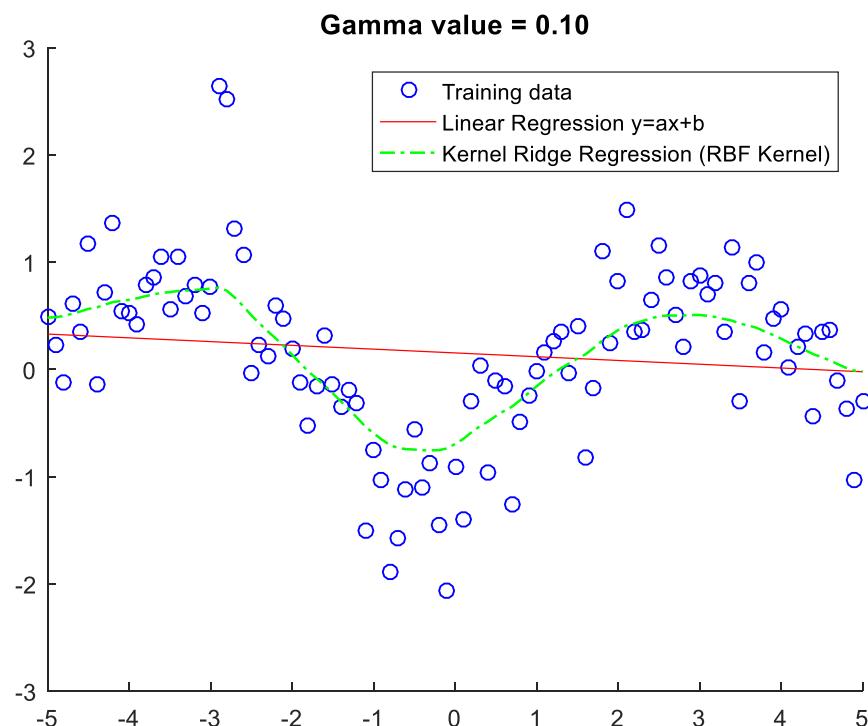
Στο προηγούμενο μάθημα είδαμε ότι μπορούμε να απεικονίσουμε τα δεδομένα μας σε ένα χώρο μεγαλύτερης διάστασης

Σε αυτό το χώρο τα δεδομένα μας είναι γραμμικά διαχωρίσιμα

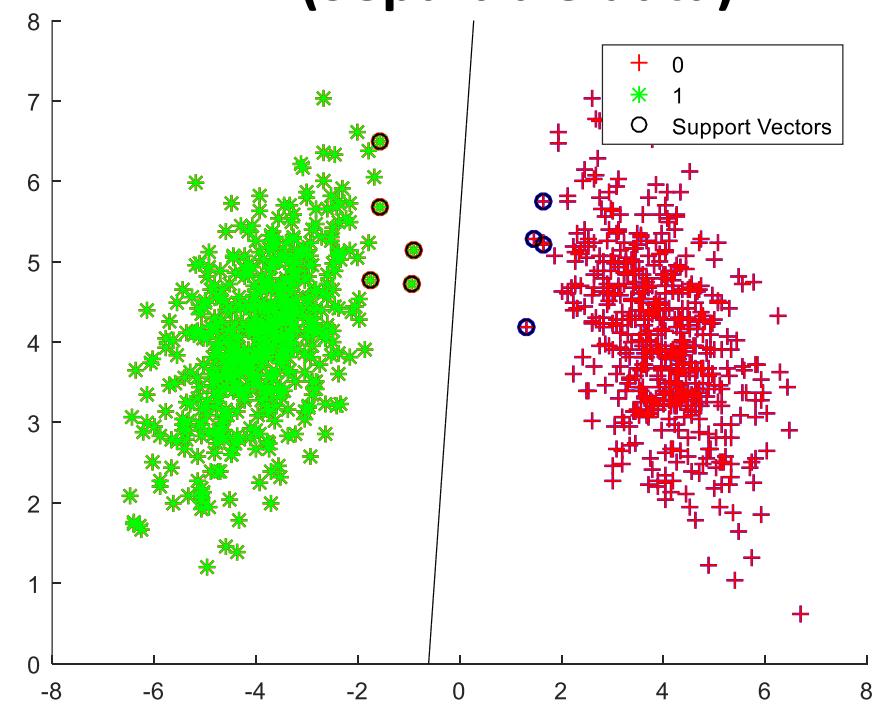
Το Kernel Trick μας επιτρέπει να αποφύγουμε την απεικόνιση σε ένα χώρο μεγαλύτερης διάστασης καθώς τα δεδομένα ανεξάρτητα αυτής αναπαρίστανται ως εσωτερικά γινόμενα

Support Vector Regression and Classification

Kernel Ridge Regression with RBF



Linear Support Vector Classification (Separable data)



Kernel Functions- Radial Basis Functions

$$K_{RBF}(x, x_j) = \exp(-\gamma \|x - x_j\|^2)$$

$$\begin{aligned} \exp(x - x_j)^2 &= \exp(-x^2) \exp(-x_j^2) \exp(2xx_j) = \\ &= \exp(-x^2) \exp(-x_j^2) \sum_{k=0}^{\infty} \frac{2^k (x^k)(x_j^k)}{k!} \end{aligned}$$

Polynomial Kernel of infinite degree

$$\exp(x) = \sum_{k=0}^{\infty} \frac{1}{k!} x^k$$

Επιλογή μοντέλου

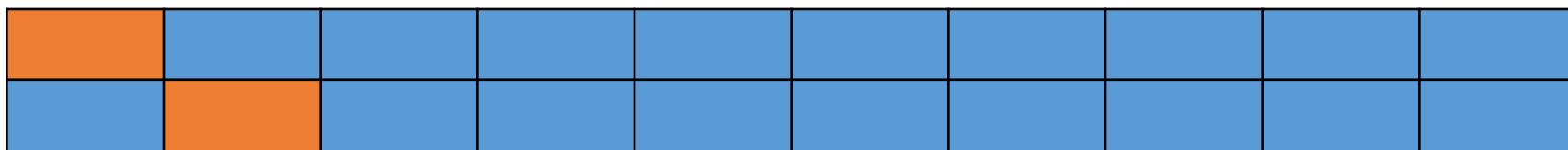
Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας,
της Μηχανικής και των Μαθηματικών – STEM στην
Εκπαίδευση 2017

Επιλογή παραμέτρων μοντέλου

- k-Fold cross validation
- Holdout method
- Leave one sample out method

Διαδικασία k-fold validation

- Χωρίζουμε τα δεδομένα μας σε **K μέρη**.
- Σε κάθε μια από τις K επαναλήψεις χρησιμοποιούμε τα K-1 σύνολα ως δεδομένα εκπαίδευσης και το σύνολο που μένει ως δεδομένα development. Εδώ δίνεται ένα παράδειγμα για **K=10**

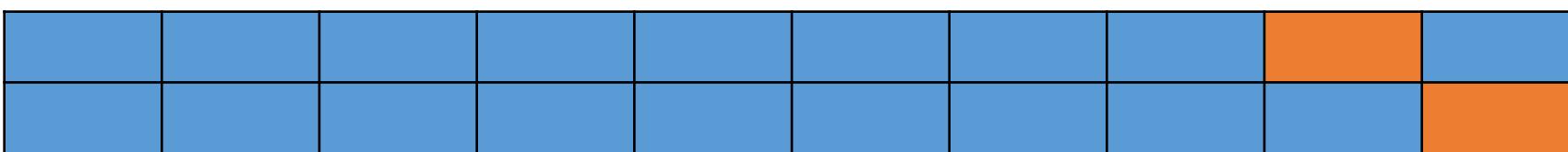


• • •

Dev Error₁

Dev Error₂

$$\text{Error} = \frac{1}{10} \sum_{i=1}^{10} \text{Dev Error}_i$$



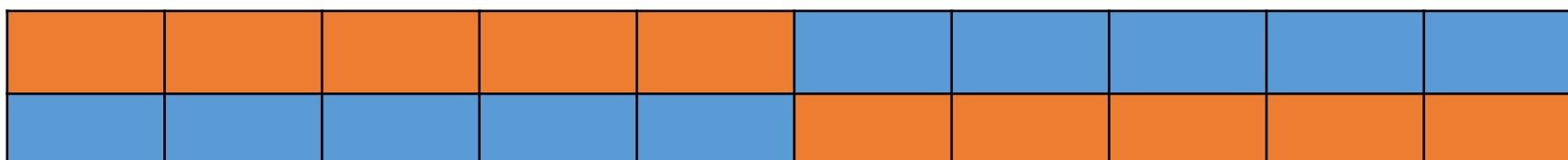
Dev Error₉

Dev Error₁₀

Διαδικασία Holdout

- Ειδική περίπτωση του K-fold.
- Χωρίζουμε τα δεδομένα μας σε **2 μέρη**.
- Χρησιμοποιούμε το ένα από τα δύο κάθε φορά για εκπαίδευση και το άλλο για δοκιμή (train and dev set)
- **Χρησιμοποιούμε την μέση τιμή των δύο δοκιμών.**

$$\text{Error} = \frac{1}{2} \sum_{i=1}^2 \text{Dev Error}_i$$



Dev Error₁

Dev Error₂

Πρόσφατες εξελίξεις και προκλήσεις στο χώρο της μηχανικής εκμάθησης

Βασικό πρόβλημα

- **Κλασική προσέγγιση:** Τα χαρακτηριστικά εξάγονται από ανθρώπους
- Πολλές φορές δεν είναι ικανοποιητικά- δεν διαχωρίζουν τους πληθυσμούς
- **Μοντέρνα προσέγγιση:** Προτάθηκε λοιπόν η εισαγωγή μοντέλων τα οποία θα μαθαίνουν και τα χαρακτηριστικά και το μοντέλο μέσα από μια ενιαία διαδικασία εκμάθησης

Deep Neural Networks- επόμενο μάθημα

- Ένα δίκτυο, εκπαιδεύεται από ένα μεγάλο πλήθος δειγμάτων
 - Μαθαίνει μια απεικόνιση υψηλού επιπέδου. Εξάγει χαρακτηριστικά σύνθετα αλλά ικανά να δώσουν πολύπλοκες και εύρωστες αναπαραστάσεις
 - Ταυτόχρονα μαθαίνει και το μοντέλο παλινδρόμησης/ταξινόμησης

Deep Neural Networks- επόμενο μάθημα

- Μια νέα εποχή στον προγραμματισμό
 - Παραδοσιακά ο προγραμματισμός αποτελεί προγράμματα που εκτελούν διαδικασίες ακολουθώντας ένα σύνολο εντολών
 - Το σύνολο εντολών παράγει ένα εκτελέσιμο
- Με τα δίκτυα νευρώνων, ο προγραμματιστής δίνει στο σύστημα δεδομένα και το σύστημα εκτιμά τα βάρη

Deep Neural Networks- επόμενο μάθημα



<https://github.com/jcjohnson/fast-neural-style>

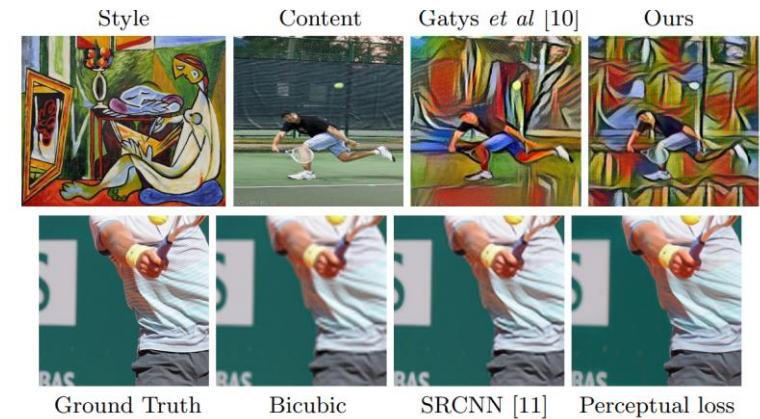
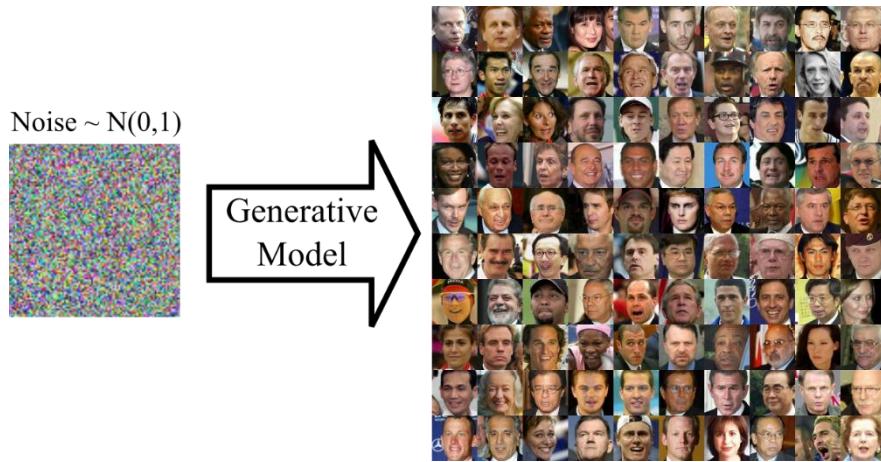
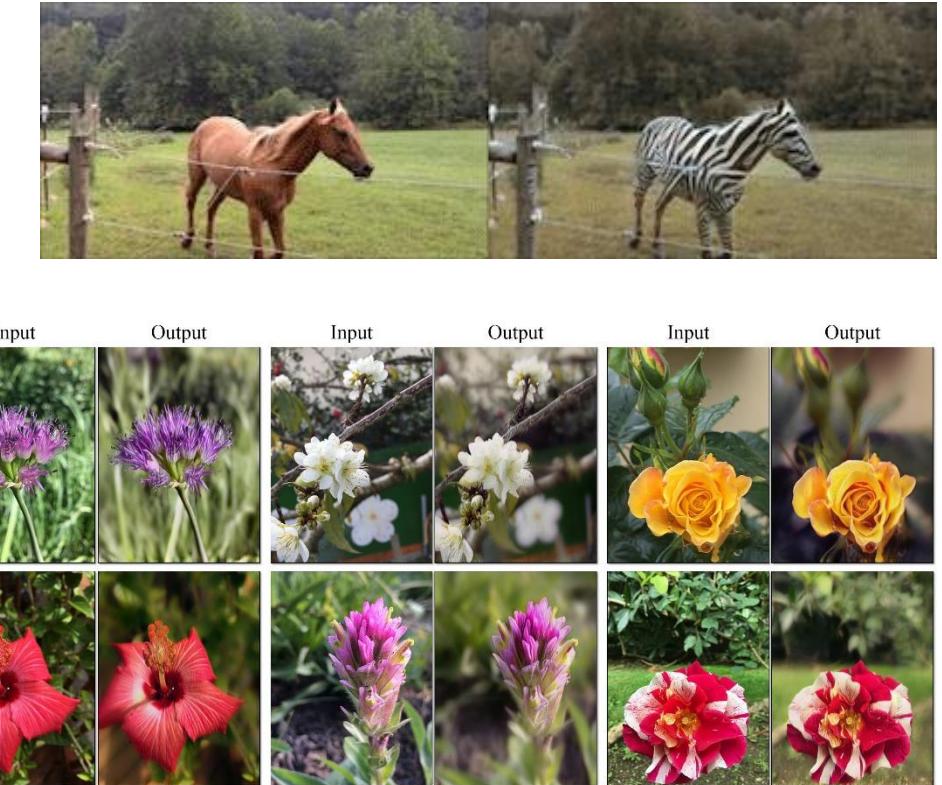


Fig. 1. Example results for style transfer (top) and $\times 4$ super-resolution (bottom). For style transfer, we achieve similar results as Gatys *et al* [10] but are three orders of magnitude faster. For super-resolution our method trained with a perceptual loss is able to better reconstruct fine details compared to methods trained with per-pixel loss.

Deep Neural Networks- επόμενο μάθημα



Generative Models

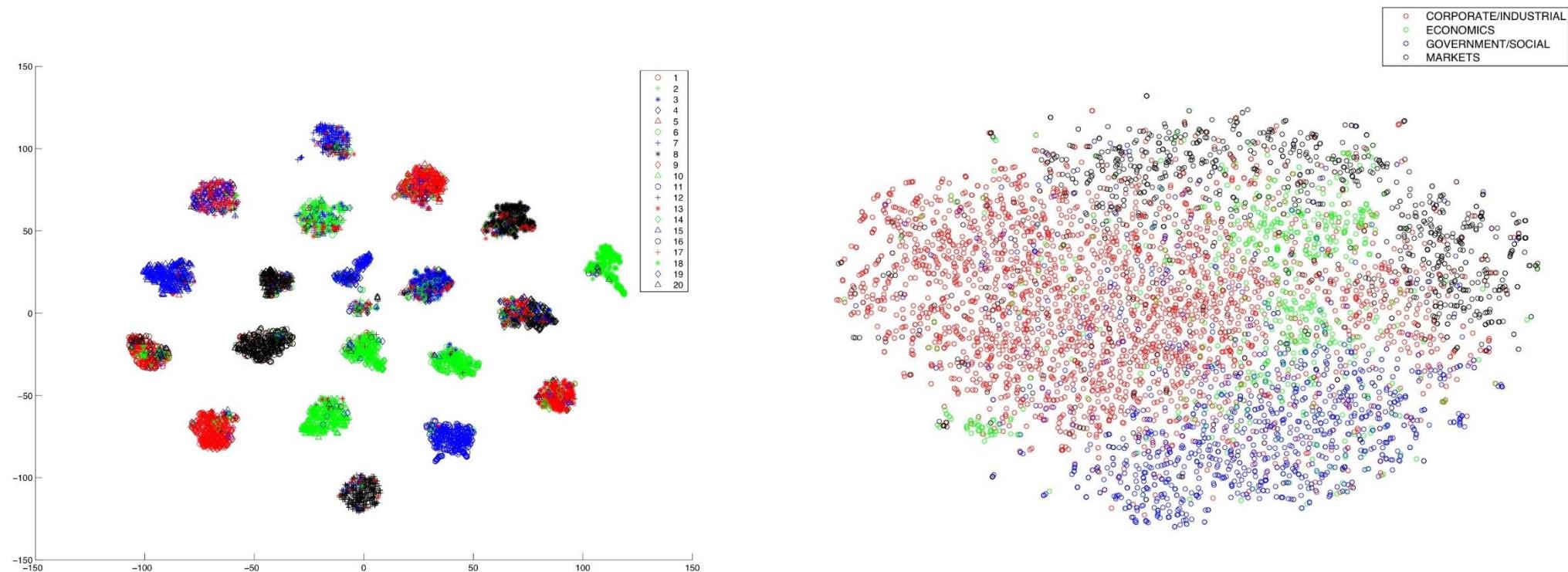


- Generating cartoon characters, art, etc.

Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A Efros, [Unpaired image-to-image translation using cycle-consistent adversarial networks](#), 2017/3/30, Proceedings of the IEEE International Conference on Computer Vision

Deep Neural Networks- επόμενο μάθημα

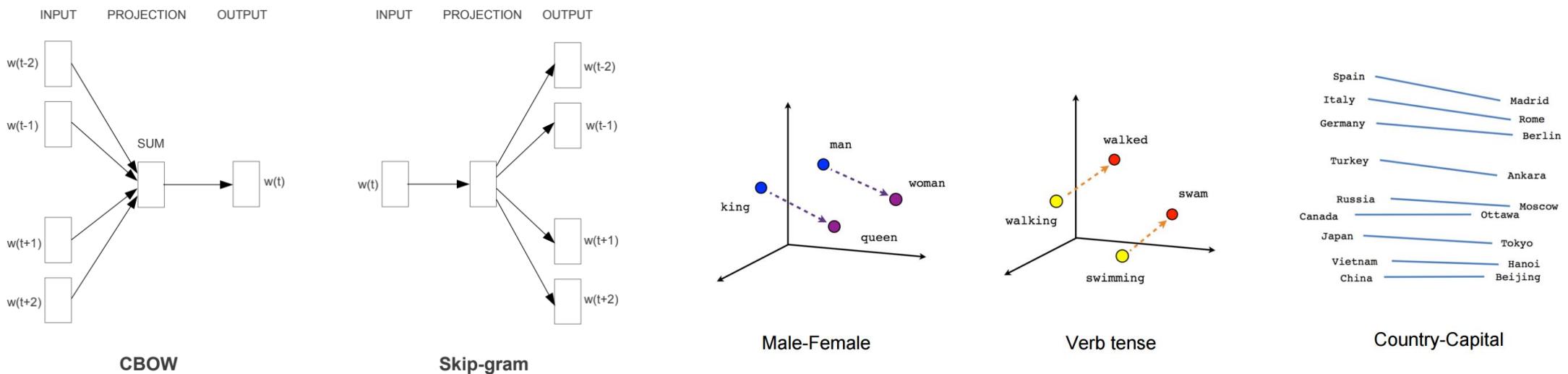
Απεικόνιση δειγμάτων στο επίπεδο



<https://lvdmaaten.github.io/tsne/>

Deep Neural Networks- επόμενο μάθημα

Word2Vec (Continuous Bag of Word , Skip-Gram)

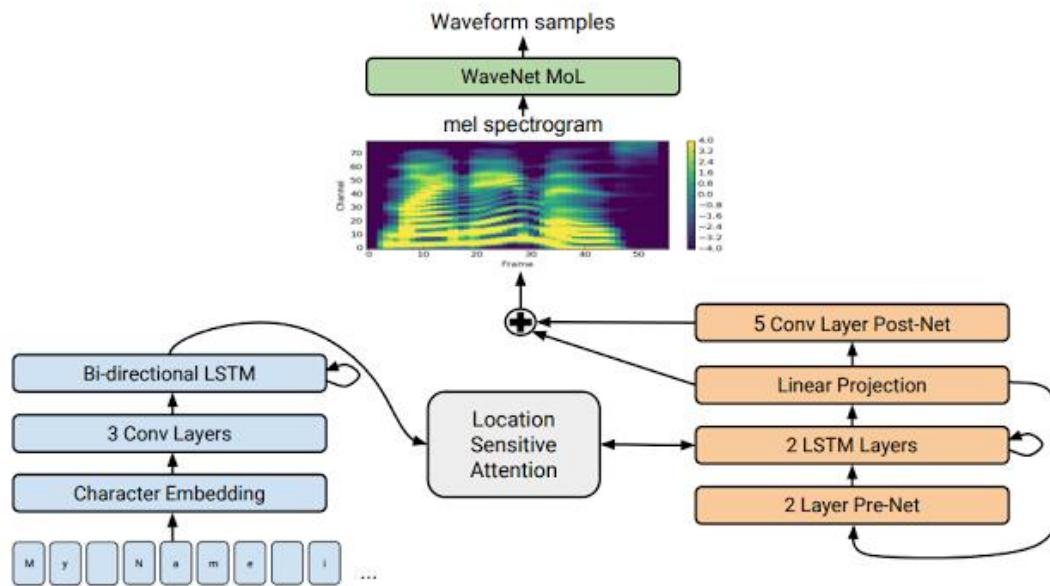


<https://www.tensorflow.org/tutorials/word2vec>

Mikolov et al. <https://www.aclweb.org/anthology/N13-1090>

Deep Neural Networks- επόμενο μάθημα

Generating Speech from text



<https://research.googleblog.com/2017/12/tacotron-2-generating-human-like-speech.html>

Επόμενο μάθημα

- Deep Learning for Modern Computer Vision
- Deep Learning for Natural Language Processing

Ευχαριστώ πολύ

Διεπιστημονική Προσέγγιση της Επιστήμης, της Τεχνολογίας,
της Μηχανικής και των Μαθηματικών – STEM στην
Εκπαίδευση 2017