ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

Business Analytics

Buzzwords

Περιεχόμενα

- Στατιστική & Έλεγχος Υποθέσεων
- Μηχανική Μάθηση (Machine Learning)
  - Απλή Γραμμική Παλινδρόμηση, Πολλαπλή Παλινδρόμηση (Regression)
  - Λογιστική Παλινδρόμηση (Logistic Regression)
  - K-Κοντινότεροι Γείτονες (k-nearest neighbors)
  - Νευρωνικά Δίκτυα (Neural Networks)
  - Βαθιά Μάθηση (Deep Learning)
- Βάσεις δεδομένων και SQL (Structured Query Language)

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

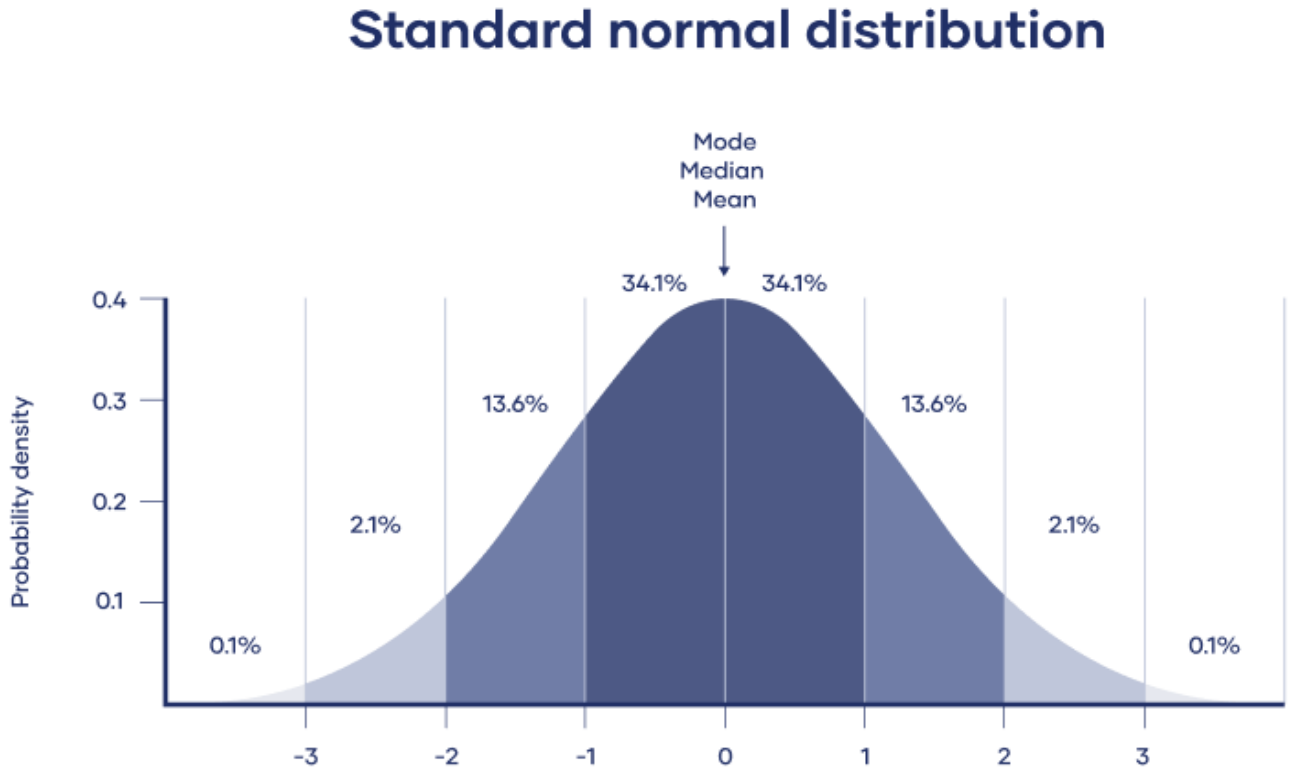
ΣΤΑΤΙΣΤΙΚΗ

ΜΑΘΗΜΑΤΙΚΕΣ ΚΑΙ ΤΕΧΝΙΚΕΣ ΕΝΝΟΙΕΣ ΜΕ ΤΙΣ ΟΠΟΙΕΣ ΚΑΤΑΝΟΟΥΜΕ ΤΑ ΔΕΔΟΜΕΝΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

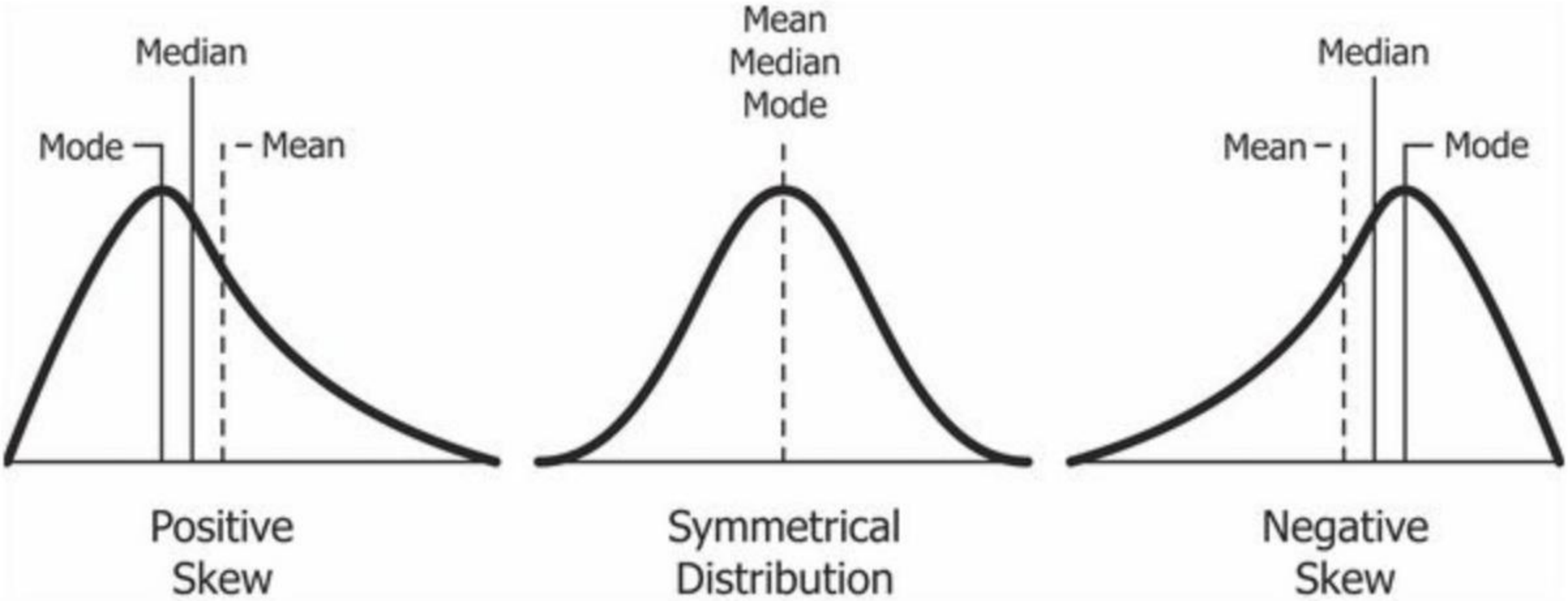
Κεντρικές Τάσεις

- Συχνά χρειαζόμαστε μια έννοια που να εκφράζει το **κέντρο** των δεδομένων μας. Για αυτό πολλές φορές αναφερόμαστε **στον μέσο όρο** του δείγματος μας. Το άθροισμα δηλαδή δια το πλήθος.
- Βέβαια πολλές φορές είναι **ευαίσθητος** στην ύπαρξη **ακραίων σημείων**, για αυτό κάποιες φορές μπορεί να μας δώσει λάθος συμπεράσματα.



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

Ασυμμετρία

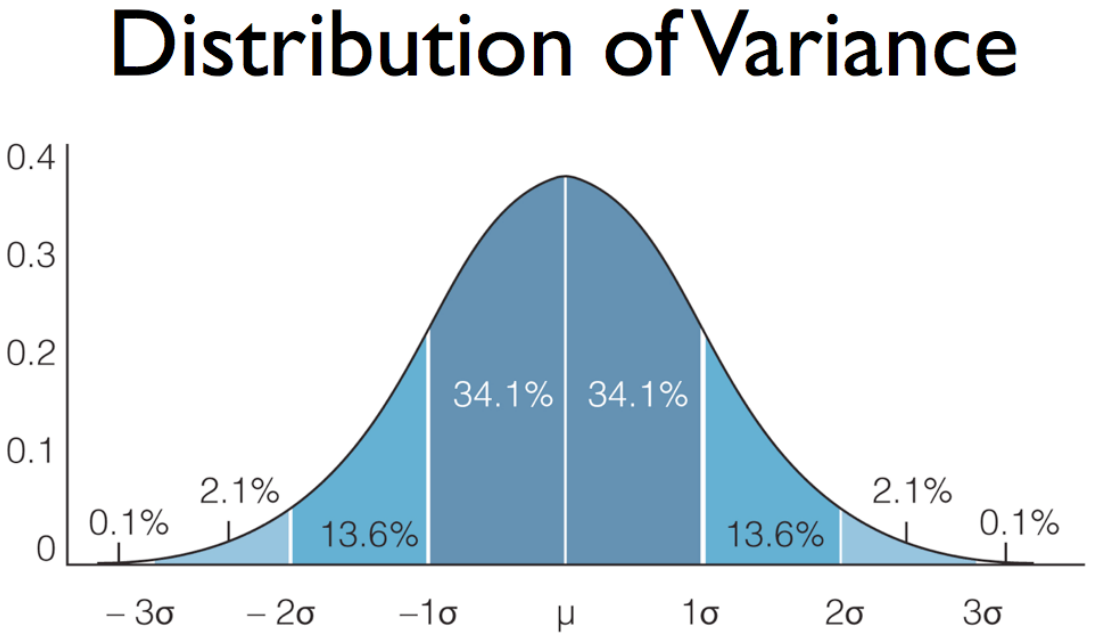


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

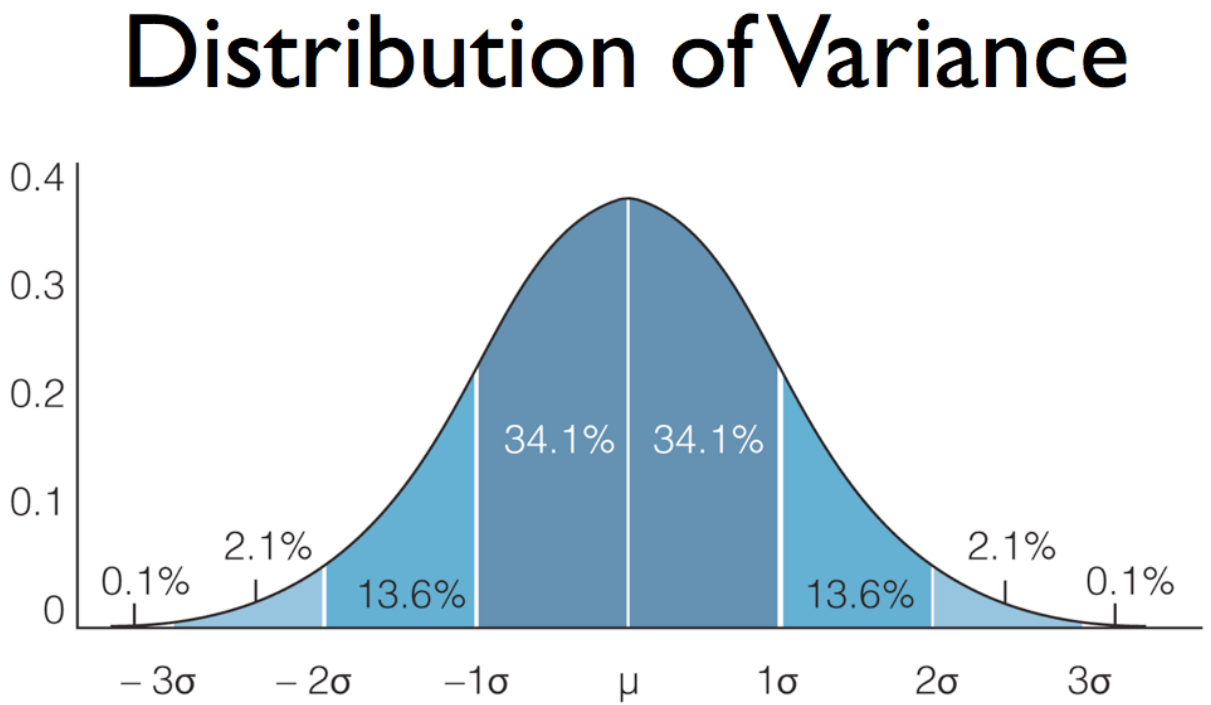
UNIVERSITY OF PIRAEUS

Διασπορά (1/2)

- Ο όρος διασπορά αναφέρεται στο πόσο διασκορπισμένα είναι τα δεδομένα μας. Τα δεδομένα κοντά μηδέν σημαίνουν **καθόλου διασπορά** ενώ μεγάλες τιμές δηλώνουν πως τα δεδομένα είναι **διάσπαρτα**.
- Το range είναι ένας τρόπος έκφρασης της διασποράς αφαιρώντας τον μικρότερο αριθμό από το μεγαλύτερο.
- Ένα άλλο πιο σύνθετο μέτρο διασποράς είναι η **διακύμανση** (var) που στην επόμενη διαφάνεια αποτυπώνουμε πως υπολογίζεται.



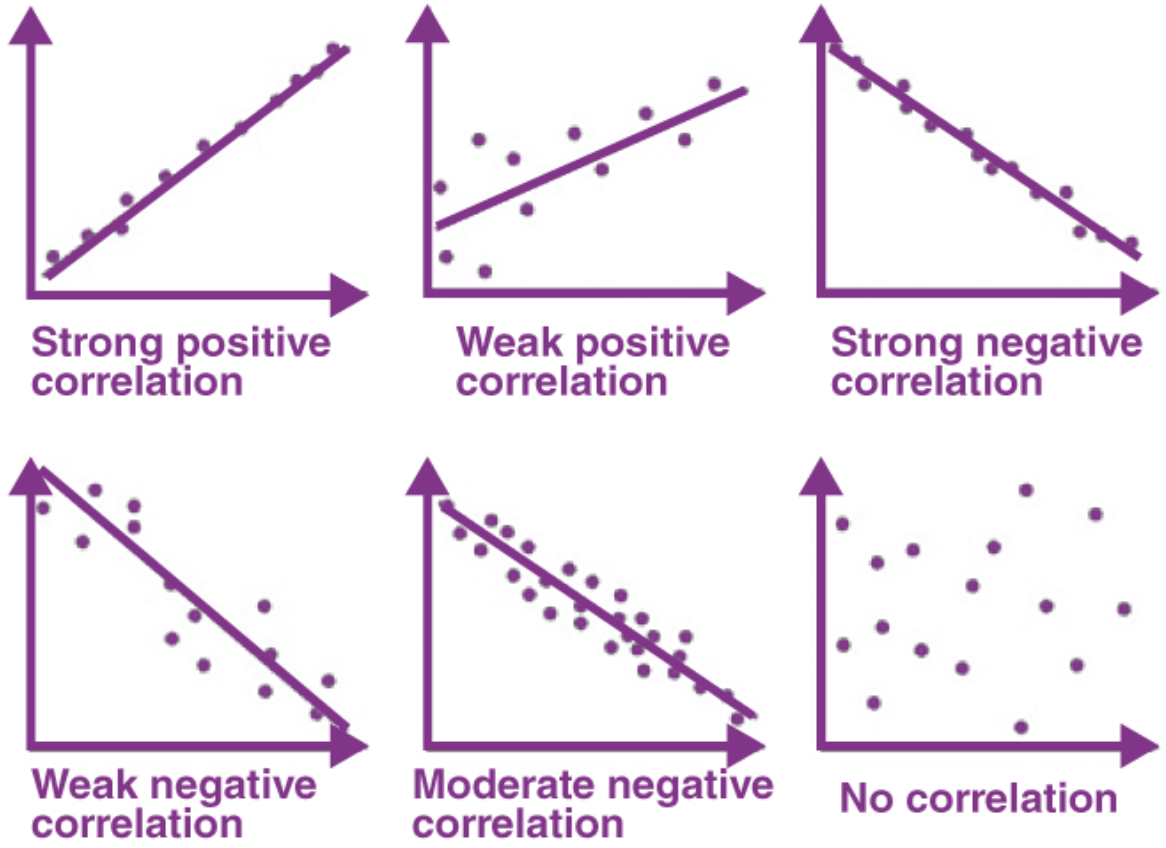
Διασπορά (2/2)


Distribution of Variance
0.4
0.3
0.2
0.1
0
34.1%
34.1%
2.1%
2.1%
13.6%
13.6%
0.1%
0.1%
– 3σ
– 2σ
–1σ
μ
1σ
2σ
3σ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

Συσχέτιση

- Η **συσχέτιση** δεν έχει μονάδες και βρίσκεται **μεταξύ του -1 και του 1.**

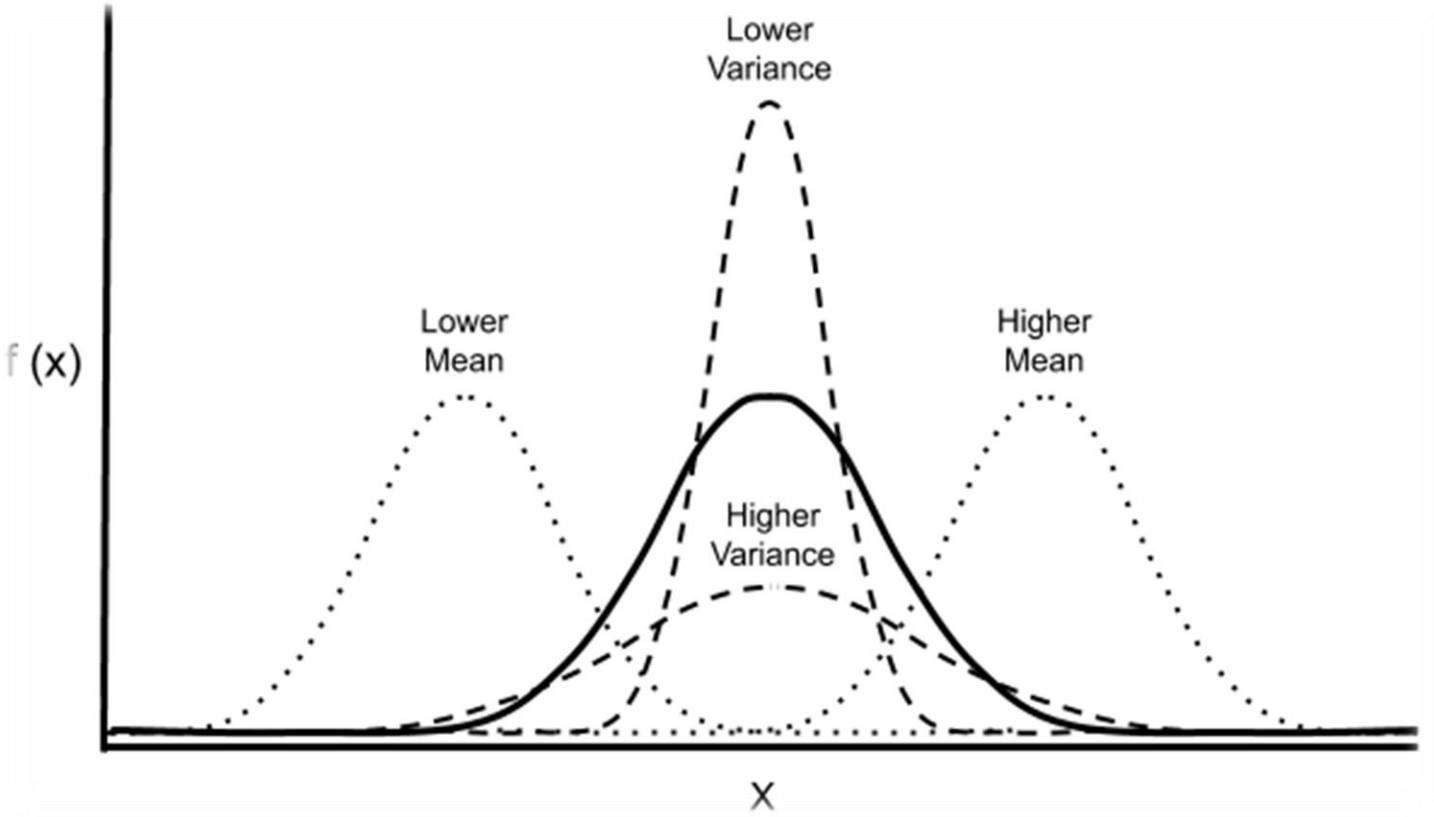


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

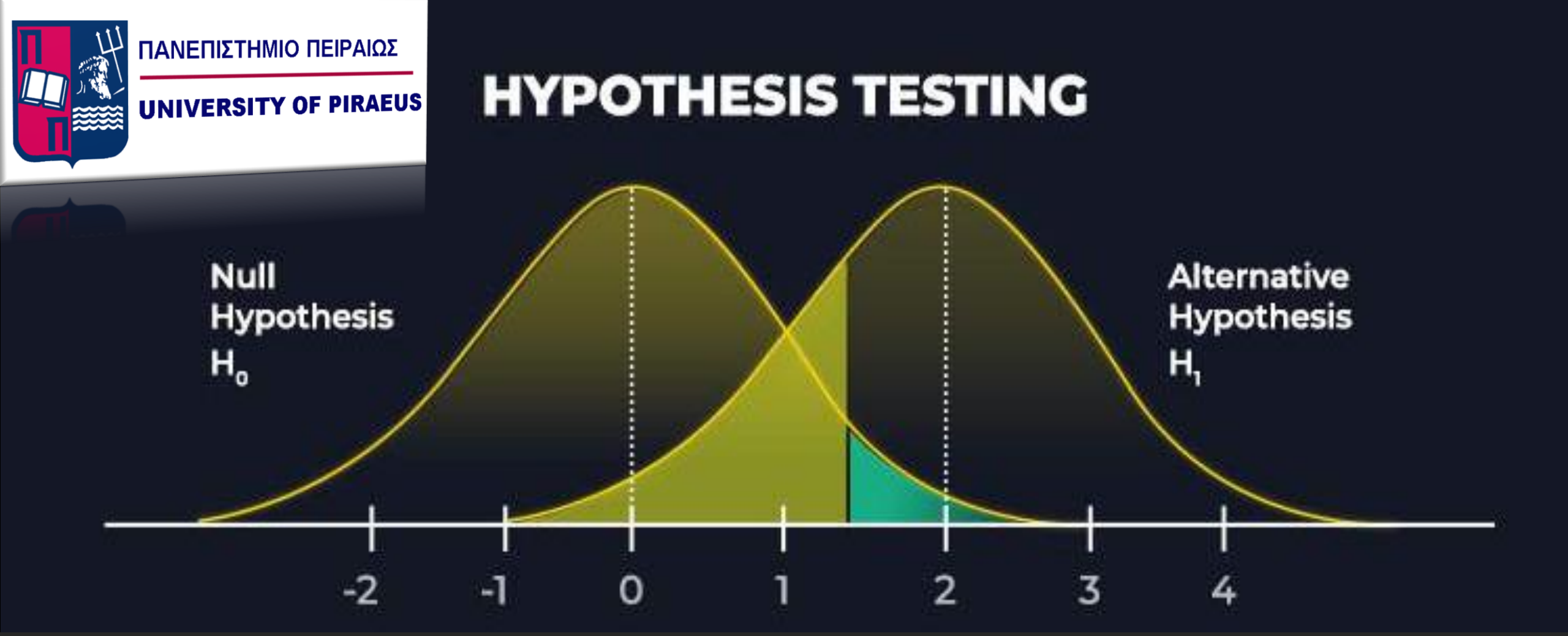
Κανονική Κατανομή

- Η κανονική κατανομή **εξαρτάται** από δύο παραμέτρους το **μ (mean)** και **την τυπική απόκλιση (σ)**.
- Όταν μ=0 και σ=1 η κατανομή λέγεται τυπική.
- Η κανονική κατανομή είναι **συμμετρική** ως προς τον **μέσο όρο**, δείχνοντας πως τα δεδομένα που βρίσκονται κοντά στο μέσο όρο εμφανίζονται πιο συχνά.



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS



ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ

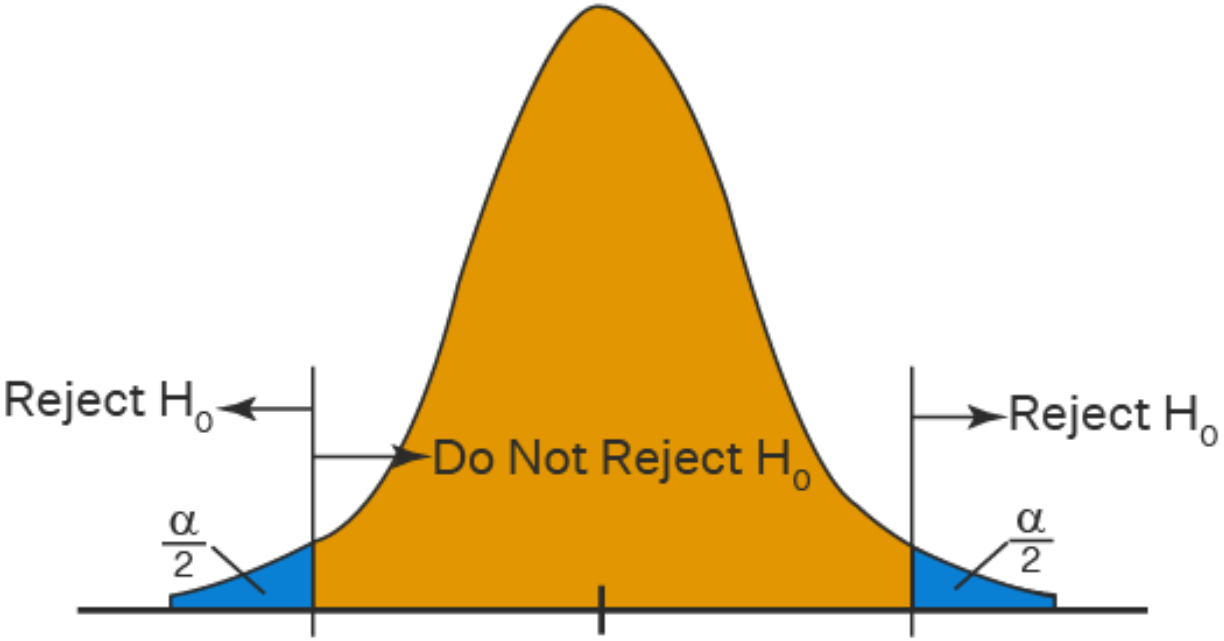
ΧΡΗΣΙΜΟΠΟΙΩΝΤΑΣ ΔΕΔΟΜΕΝΑ ΓΙΑ ΤΗΝ ΕΞΑΓΩΓΗ ΣΥΜΠΕΡΑΣΜΑΤΩΝ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Έλεγχος Υποθέσεων (1/2)

- Στο κλασσικό σενάριο έχουμε μια **μηδενική υπόθεση την Ho και μια εναλλακτική την H1.** Χρησιμοποιούμε την στατιστική για να αποφασίσουμε εάν απορρίπτουμε την Ho ως λανθασμένη ή όχι. Για παράδειγμα μπορούμε να προσδιορίσουμε εάν ο μέσος όρος είναι πιθανό να είναι αληθής.
- Αν ο δειγματικός μέσος είναι **κοντά** στον δηλωμένο μέσο του πληθυσμού, **η μηδενική υπόθεση δεν απορρίπτεται**.
- Αν ο δειγματικός μέσος **είναι μακριά** από τον δηλωμένο μέσο του πληθυσμού, **η μηδενική υπόθεση απορρίπτεται.**

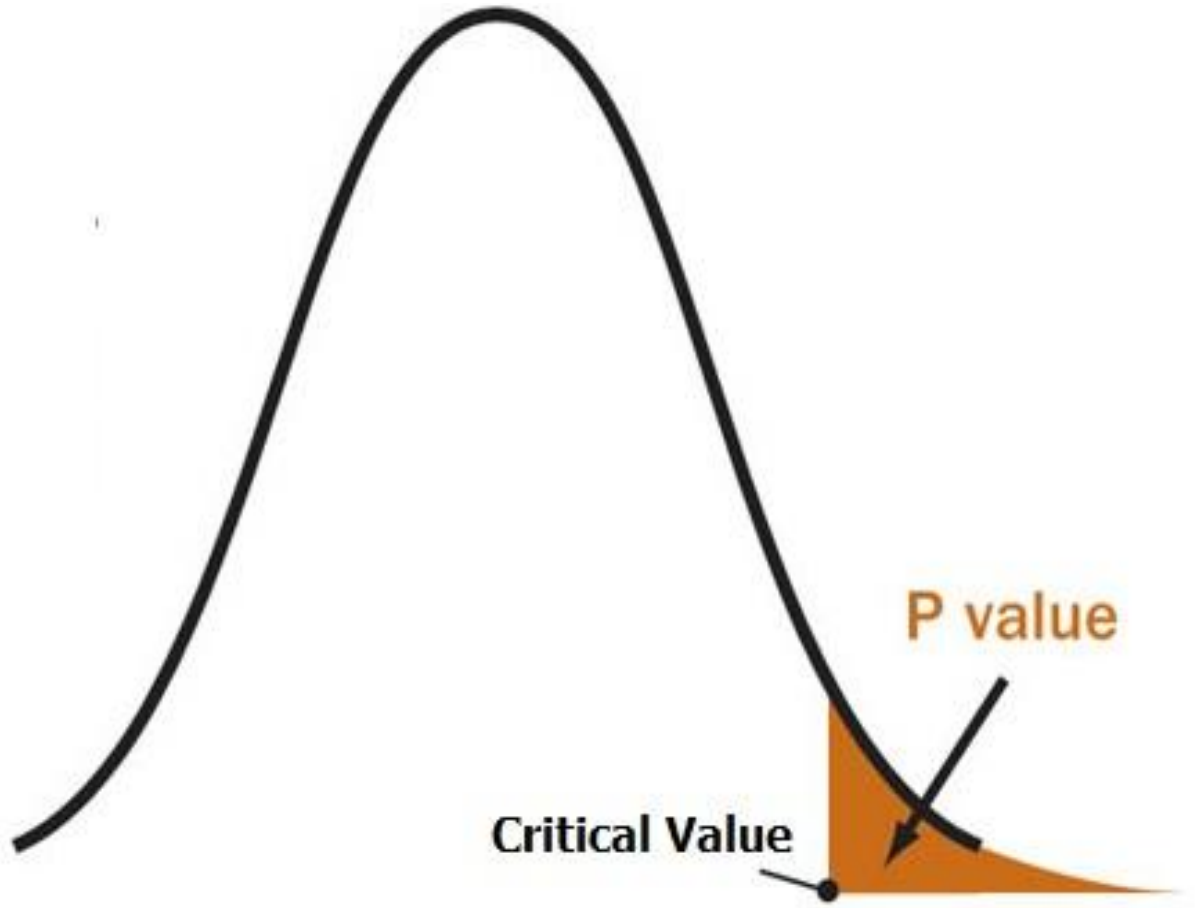


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Έλεγχος Υποθέσεων (2/2)

- Ένας **εναλλακτικός τρόπος** για να δεχτούμε ή να απορρίψουμε υποθέσεις είναι η **τιμές p**. Αντί δηλαδή να επιλέγουμε όρια βασισμένοι σε κάποια «βάση» πιθανοτήτων, **υπολογίζουμε την πιθανότητα, υποθέτοντας πως η Ho είναι αληθής**.
- Έστω δείκτης σπουδαιότητας α=5%, εάν p-value > α τότε δεν απορρίπτουμε την μηδενική υπόθεση.
- Όσο **μικρότερη είναι η τιμή p**, τόσο **ισχυρότερη** είναι η απόδειξη ότι πρέπει να **απορρίψουμε την μηδενική υπόθεση.**



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

ΧΡΉΣΗ ΠΡΟΥΠΑΡΧΟΝΤΩΝ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗΝ ΑΝΑΠΤΥΞΗ ΝΕΩΝ ΜΟΝΤΕΛΩΝ ΠΡΟΒΛΕΨΗΣ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Μηχανική Μάθηση (1/3)

- Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν κατασκευάζοντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασιζόμενες στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως το αποτέλεσμα

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Μηχανική Μάθηση (2/3)

- **Υπερπροσαρμογή**

Δημιουργία ενός μοντέλου που αποδίδει καλά στα δεδομένα μας έχοντας μέτρια απόδοση σε οποιαδήποτε νέα.

- **Υποπροσαρμογή**

Δημιουργία ενός μοντέλου που δεν αποδίδει καλά, με αποτέλεσμα να μην έχει νόημα η ανάλυση του.

- Κάποιες φορές **πολύπλοκα μοντέλα** οδηγούν σε **υπερπροσαρμογή**, επομένως ο απλούστερος τρόπος για να μην είναι το μοντέλο μας πολύπλοκο **είναι να διαχωρίσουμε τα δεδομένα μας.**

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Μηχανική Μάθηση (3/3)

- Ένας άλλος τρόπος για την αντιμετώπιση της υπερπροσαρμογής είναι ο συμβιβασμός της Μεροληψίας και της διακύμανσης.
- **Υψηλή μεροληψία και χαμηλή διακύμανση αντιστοιχούν σε υποπροσαρμογή.**
- **Χαμηλή μεροληψία και υψηλή διακύμανση αντιστοιχούν σε υπερπροσαρμογή.**
- Η υψηλή μεροληψία αντιμετωπίζεται προσθέτοντας περισσότερα χαρακτηριστικά.
- Η υψηλή διακύμανση αντιμετωπίζεται αφαιρώντας χαρακτηριστικά.

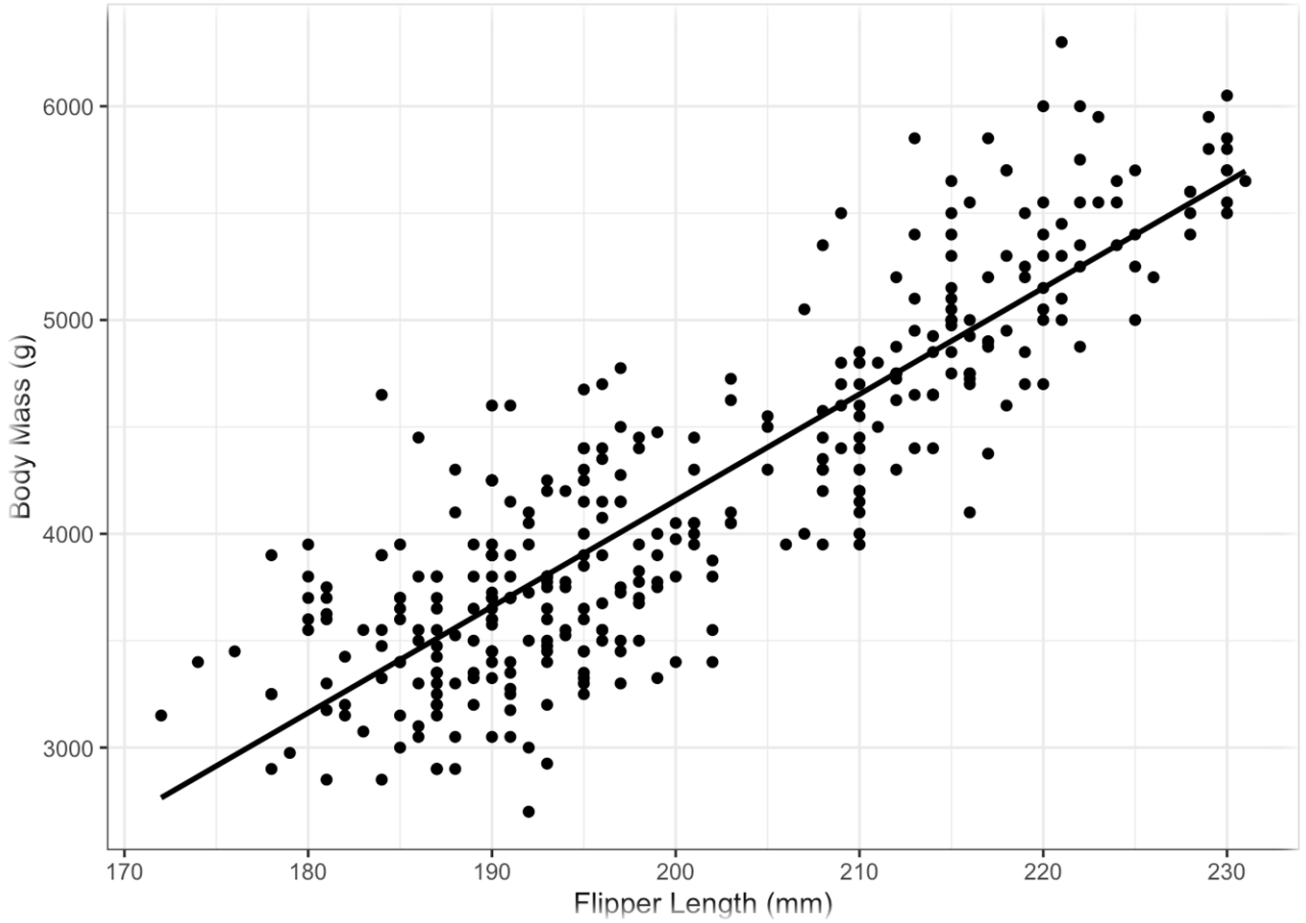
Όσο περισσότερα δείγματα έχουμε τόσο δυσκολότερο είναι να έχουμε υπερπροσαρμογή, χωρίς αυτό να σημαίνει απαραίτητα πως θα βοηθήσουν στην μεροληψία του δείγματος.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

$$Y = A + BX + E$$



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

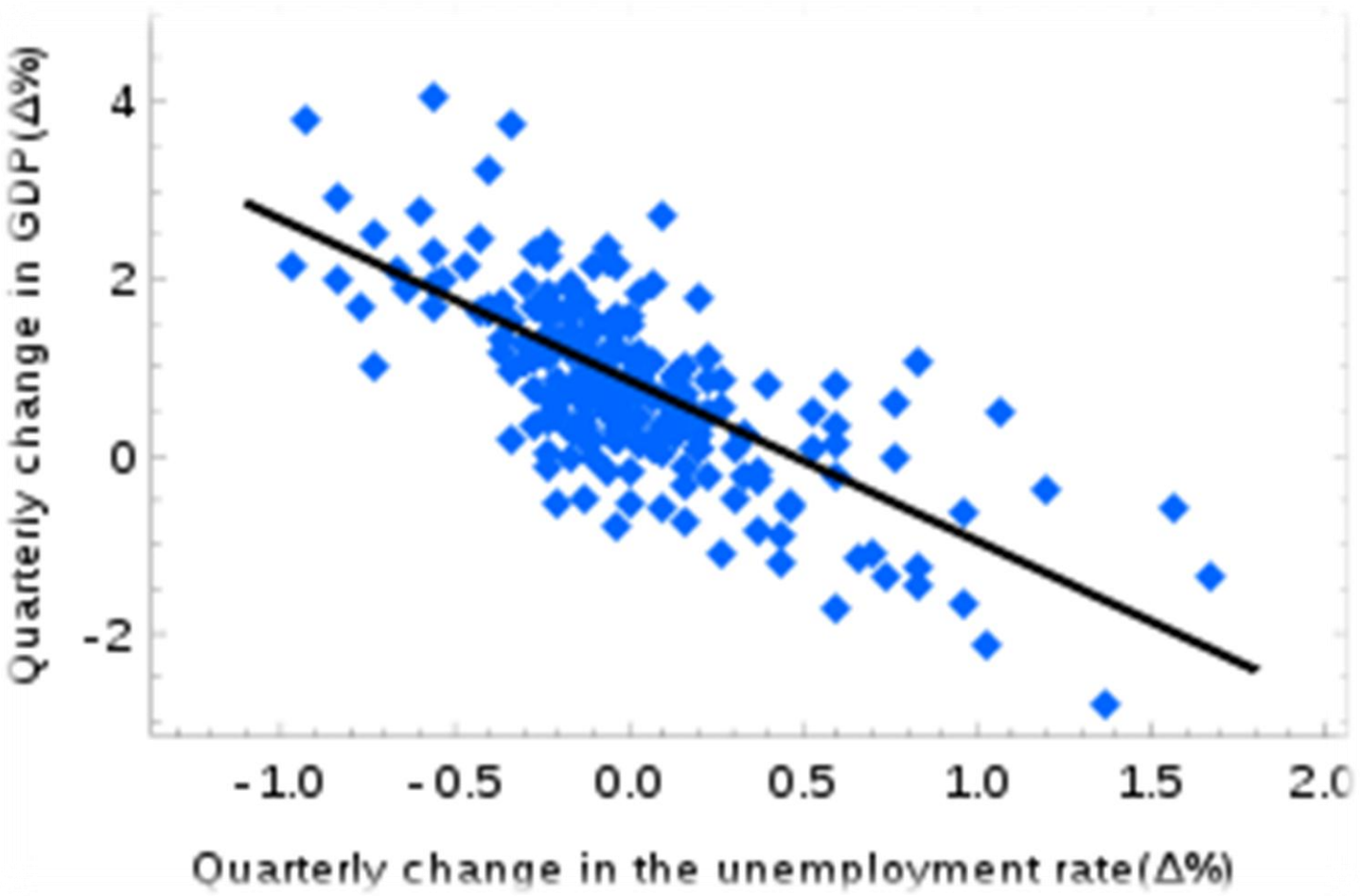
UNIVERSITY OF PIRAEUS

Απλή Γραμμική Παλινδρόμηση(1/2)

- Στην παραπάνω εξίσωση ο **όρος Ε είναι ο όρος σφάλματος**. Για κάθε ζεύγος Α και Β χρειάζεται να υπολογίσουμε το σφάλμα. Βέβαια αυτό που θα θέλαμε να γνωρίζαμε είναι το συνολικό σφάλμα σε ολόκληρο το σύνολο δεδομένων.

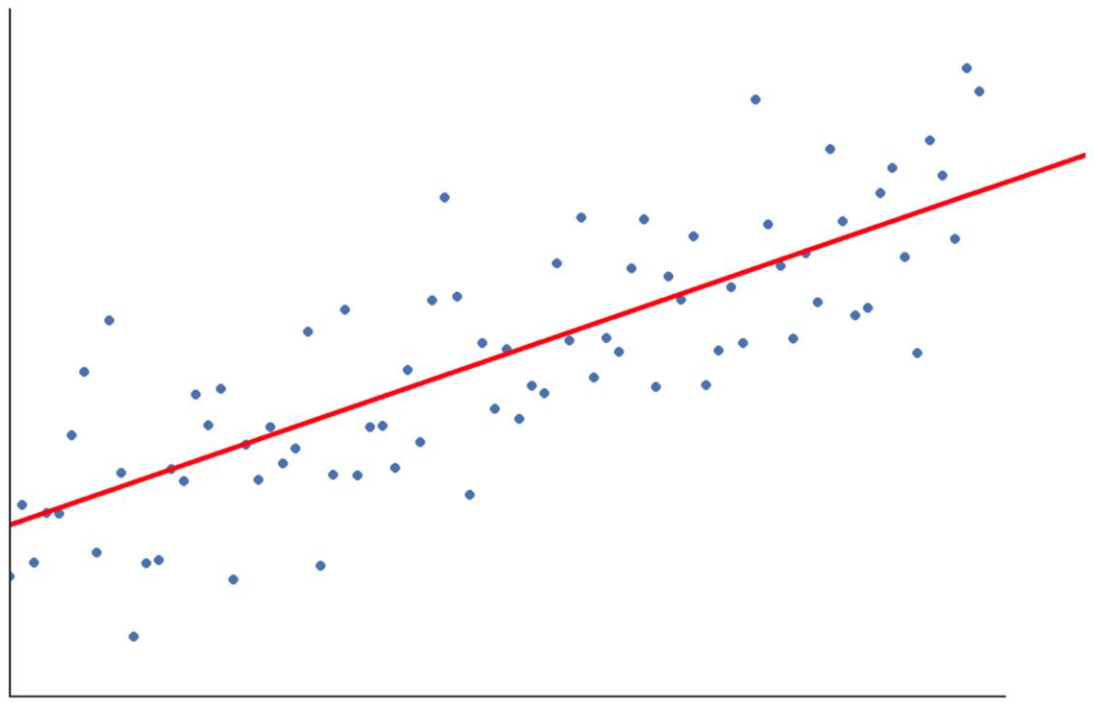
- Στην ουσία η απλή γραμμική παλινδρόμηση είναι **η μοντελοποίηση μιας εξαρτημένης μεταβλητής και μιας ανεξάρτητης μεταβλητής.**

- Στο παράδειγμα της πολλαπλής παλινδρόμησης όμως έχουμε ένα μεγαλύτερο πλήθος ανεξάρτητων μεταβλητών.



Απλή Γραμμική Παλινδρόμηση (2/2)

- **Η Μέθοδος Ελαχίστων Τετραγώνων μας δίνει τη δυνατότητα να επιλέξουμε συγκεκριμένες τιμές για Α και Β ελαχιστοποιώντας το σφάλμα**.
- Βέβαια δεν αρκεί μόνο αυτό για να βεβαιωθούμε πως έχουμε επιλέξει το σωστό για το μοντέλο μας.
- Ο συντελεστής **R sqrt είναι ένας συντελεστής που μετρά το ποσοστό της συνολικής απόκλισης της εξαρτημένης μεταβλητής.**

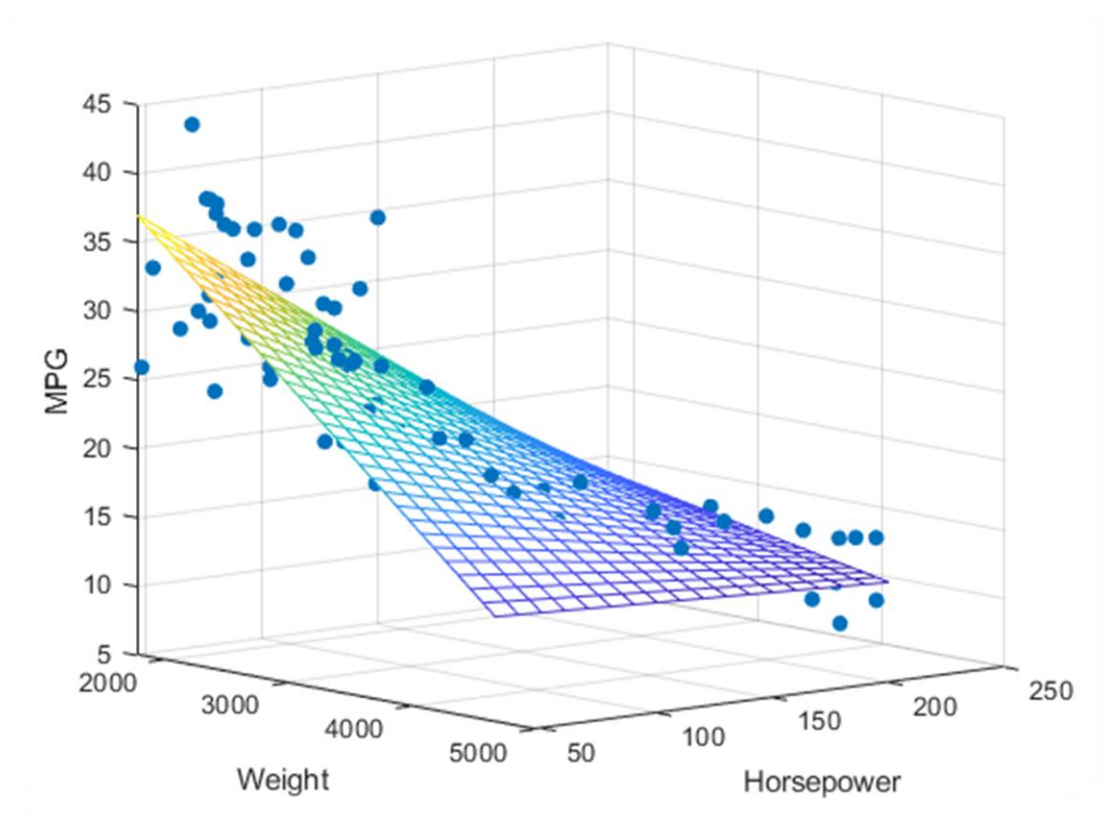


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

ΠΟΛΛΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

$$Y = A + B_1X_1 + B_2X_2 + B_3X_3 + E$$



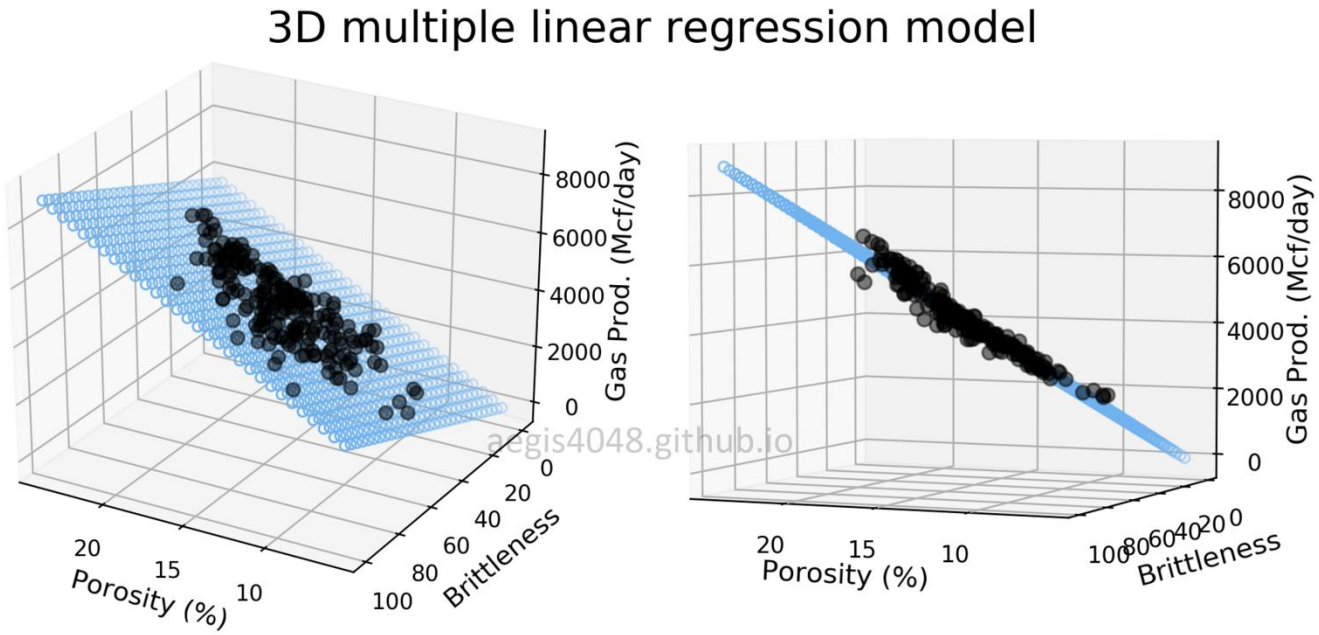
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Πολλαπλή Παλινδρόμηση(1/2)

- Ένα γραμμικό μοντέλο με περισσότερες ανεξάρτητες μεταβλητές. Όπως κάναμε και στο απλό γραμμικό μοντέλο θα επιλέξουμε “beta” ώστε να ελαχιστοποιήσουμε το άθροισμα των τετραγωνικών σφαλμάτων με την διαφορά πως **σε αυτήν την περίπτωση το beta θα δέχεται ένα διάνυσμα αυθαίρετου μήκους.**

- Γενικά η **προσθήκη νέων μεταβλητών** σε μια παλινδρόμηση θα **αυξάνουν συνεχώς το R sqrt**. Για αυτό το λόγο θα χρειαστεί να εξετάζουμε και τα **τυπικά σφάλματα**.

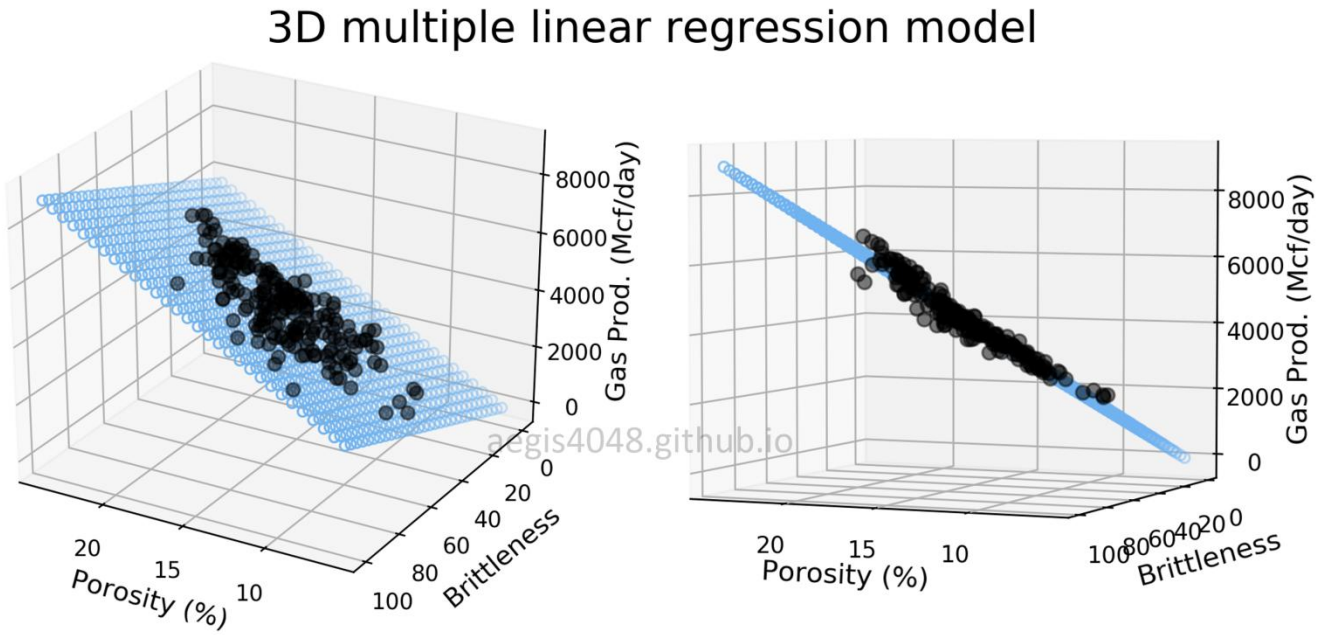


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Πολλαπλή Παλινδρόμηση(2/2)

- Όσο **περισσότερες μεταβλητές** χρησιμοποιούμε τόσο πιο πιθανό είναι να **«υπερπροσαρμόσουμε»** το μοντέλο μας.
- Δεύτερον όσο περισσότερους **μη μηδενικούς** συντελεστές έχουμε στο μοντέλο μας τόσο πιο **δύσκολο** είναι να βγάλουμε νόημα μέσα από αυτούς.

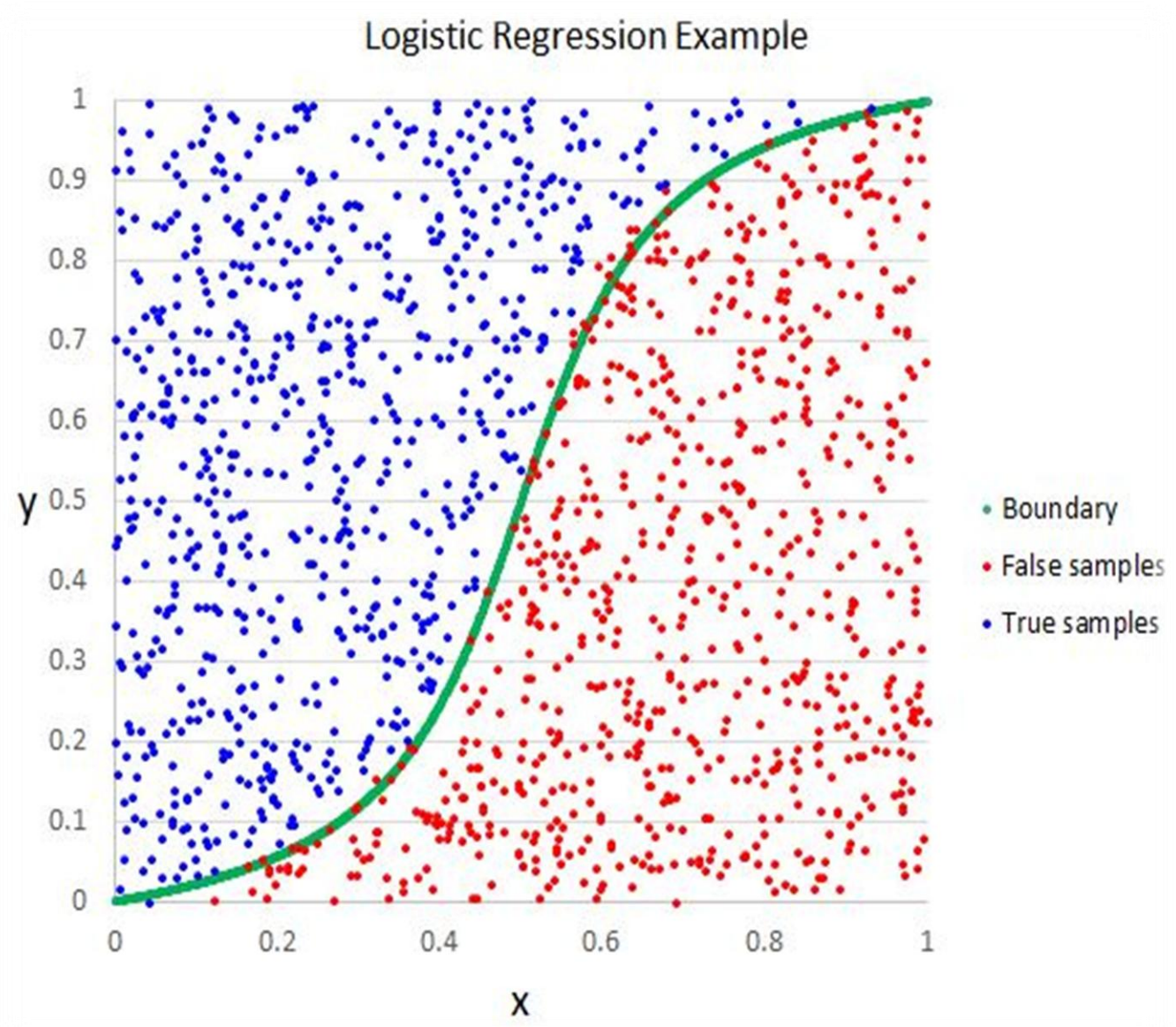


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

LOGISTIC REGRESSION



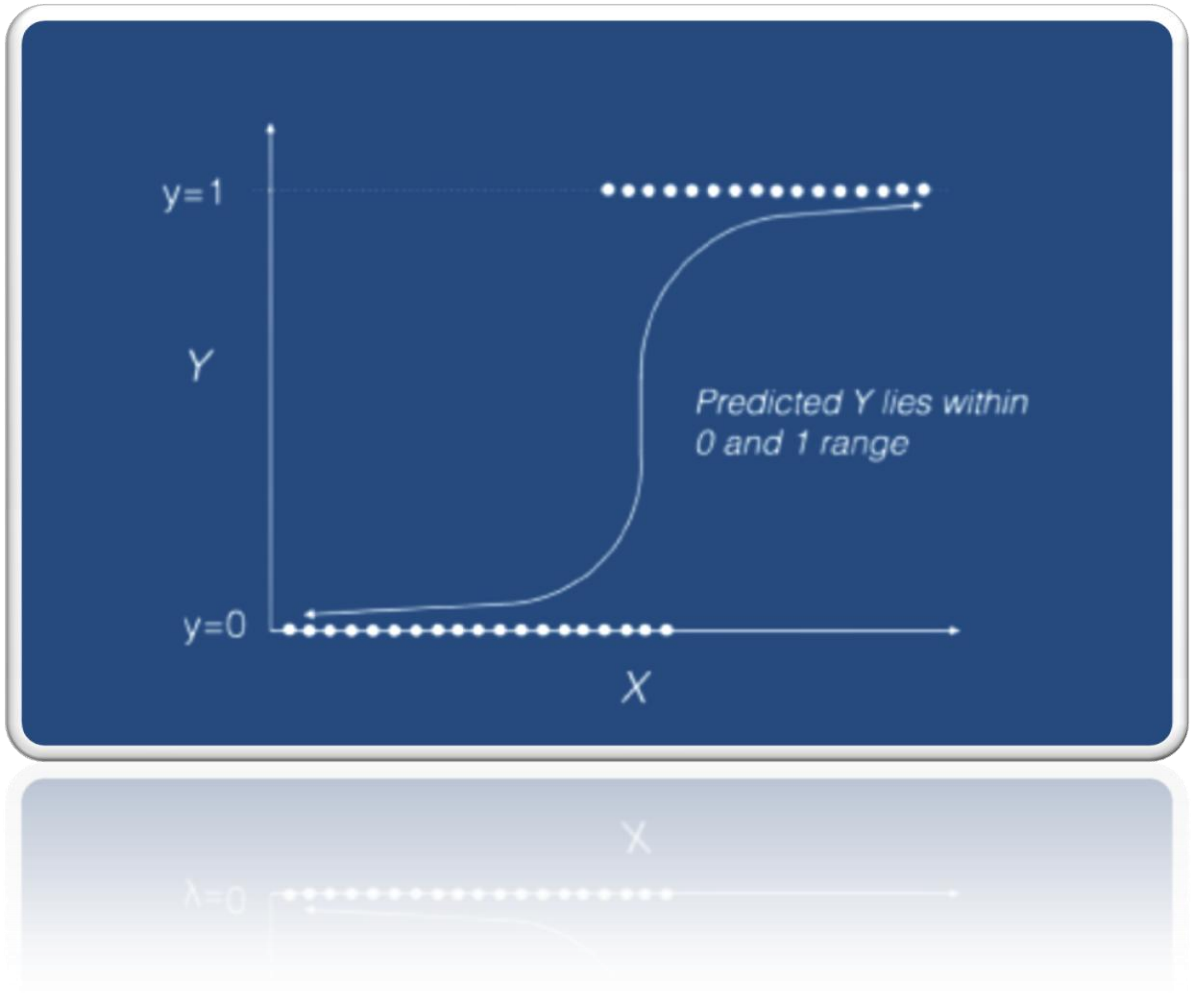
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Λογιστική Παλινδρόμηση(1/2)

- Η λογιστική παλινδρόμηση είναι μια S-shaped συνάρτηση. **Όσο η είσοδος της παίρνει υψηλότερες θετικές τιμές τόσο πιο πολύ πλησιάζει στο 1.**
- Η συνάρτηση είναι της μορφής :

  Yi = f(Xiβ) + εi , όπου f λογιστική συνάρτηση.

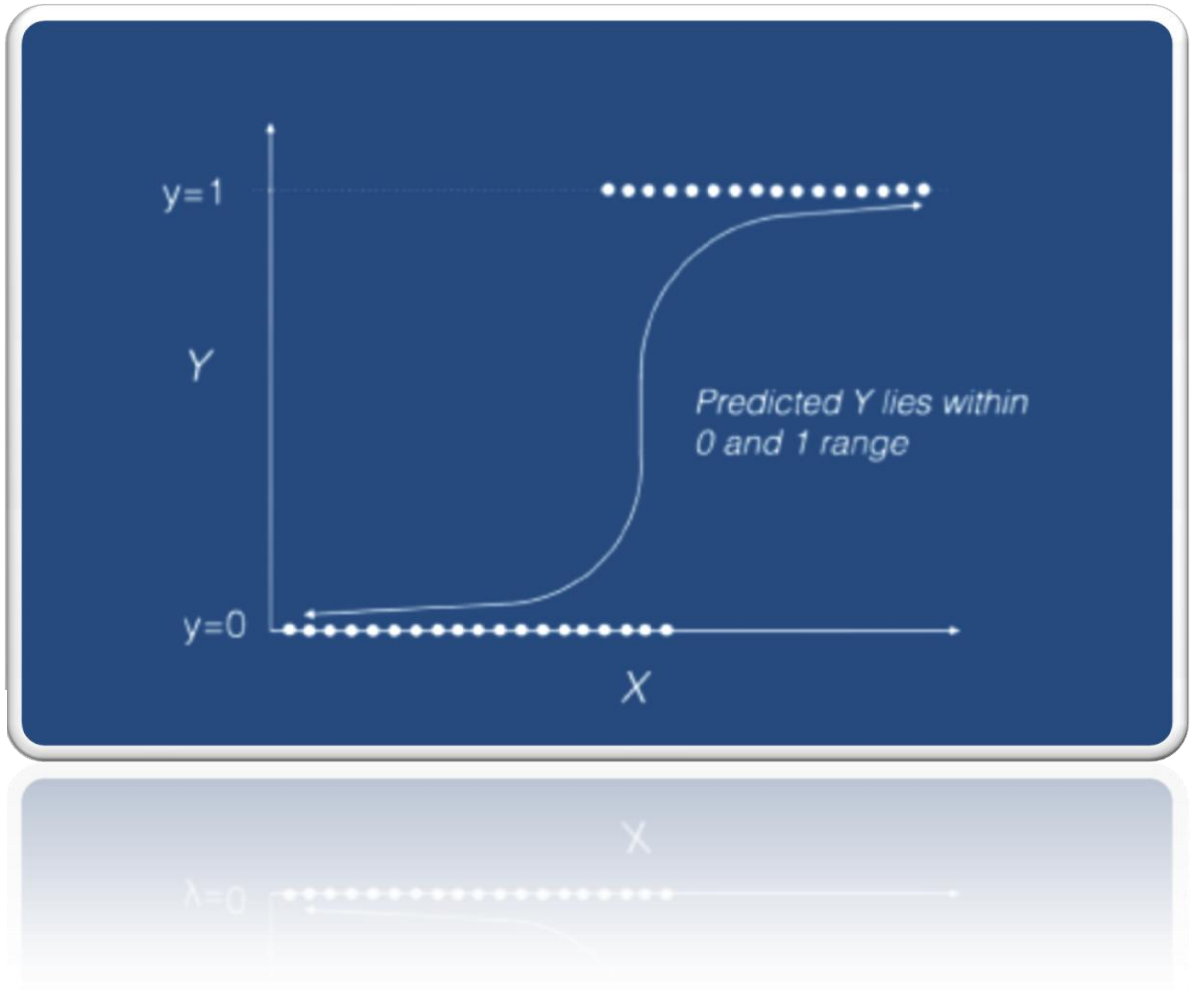


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Λογιστική Παλινδρόμηση(2/2)

- Μια **βασική διαφορά** είναι πως σε αντίθεση με την απλή γραμμική παλινδρόμηση που χρησιμοποιεί την ελαχιστοποίηση τετραγώνων, **η λογιστική παλινδρόμηση εφαρμόζει την τεχνική της μέγιστης πιθανοφάνειας.**



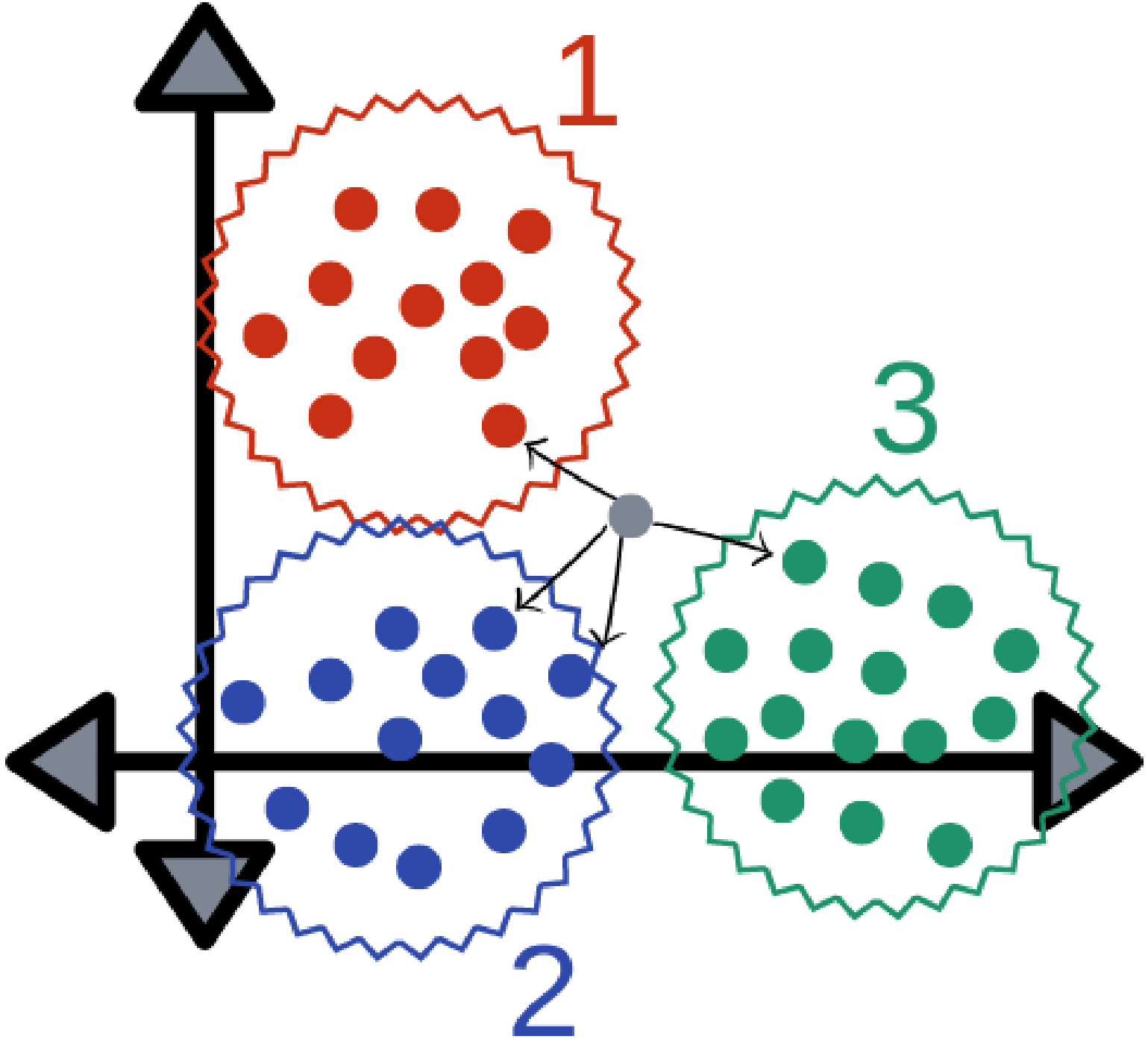
Αν $X_1, \ldots, X_n$ ανεξάρτητες, τότε $L(\underline{\vartheta}|\underline{x}) = \prod_{i=1}^{n} f(x_i; \underline{\vartheta})$.

Ερμηνεία: Για δεδομένο δείγμα παρατηρήσεων $\underline{x}$, η συνάρτηση πιθανοφάνειας $L(\underline{\vartheta}|\underline{x})$ εκφράζει το πόσο πιθανό είναι να έχουμε παρατηρήσει τα δεδομένα $\underline{x}$ που παρατηρήσαμε για τις διάφορες τιμές της άγνωστης παραμέτρου $\underline{\vartheta}$.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

Κ-ΠΛΗΣΙΕΣΤΕΡΟΙ ΓΕΙΤΟΝΕΣ

K-NEAREST NEIGHBORS (KNN)

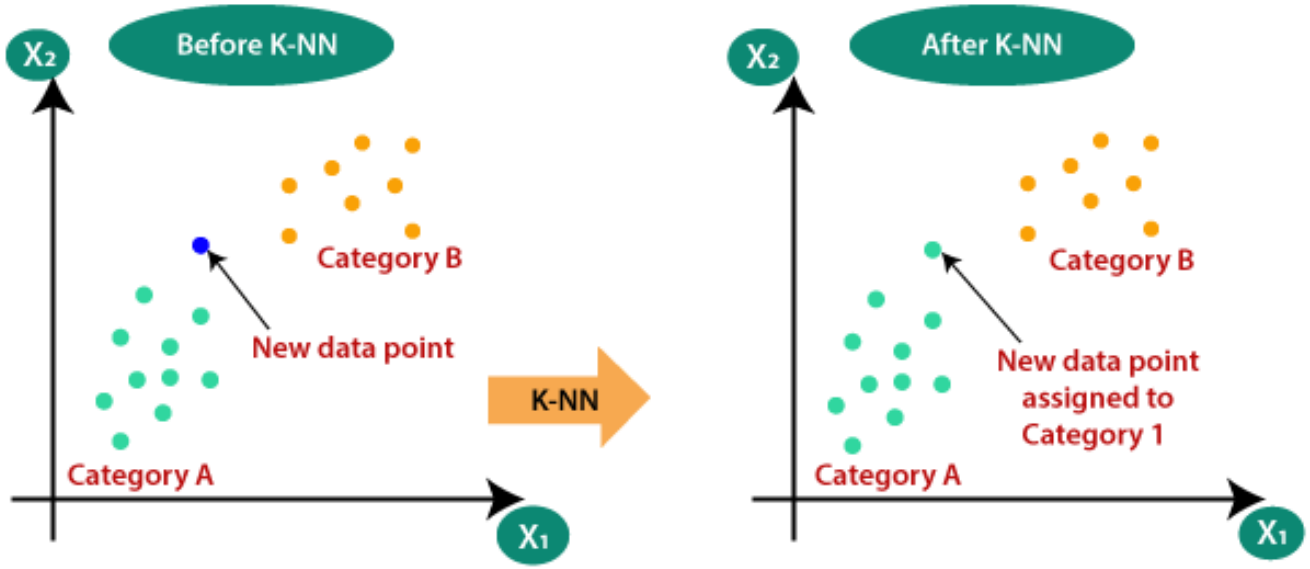


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

k-nearest neighbors (KNN)

- Λειτουργεί βρίσκοντας **αποστάσεις** μεταξύ ενός ερωτήματος μέσω όλων των παραδειγμάτων στα δεδομένα, επιλέγοντας αριθμούς K πιο κοντά στο ερώτημα. Έπειτα τα συλλέγει σε μία λίστα βρίσκοντας την πιο **συχνή παρατήρηση** ή υπολογίζοντας τον **μέσο όρο**.
- Ένας σύνηθες και εύκολος **τρόπος** είναι να λάβουμε υπόψη **το άθροισμα των τετραγωνικών σφαλμάτων κάθε σημείου σαν συνάρτηση του K.**
- Βέβαια το μοντέλο αυτό συναντά προβλήματα σε υψηλότερες διαστάσεις. **Καθώς το πλήθος των διαστάσεων αυξάνεται, αυξάνεται και η μέση απόσταση τους.**

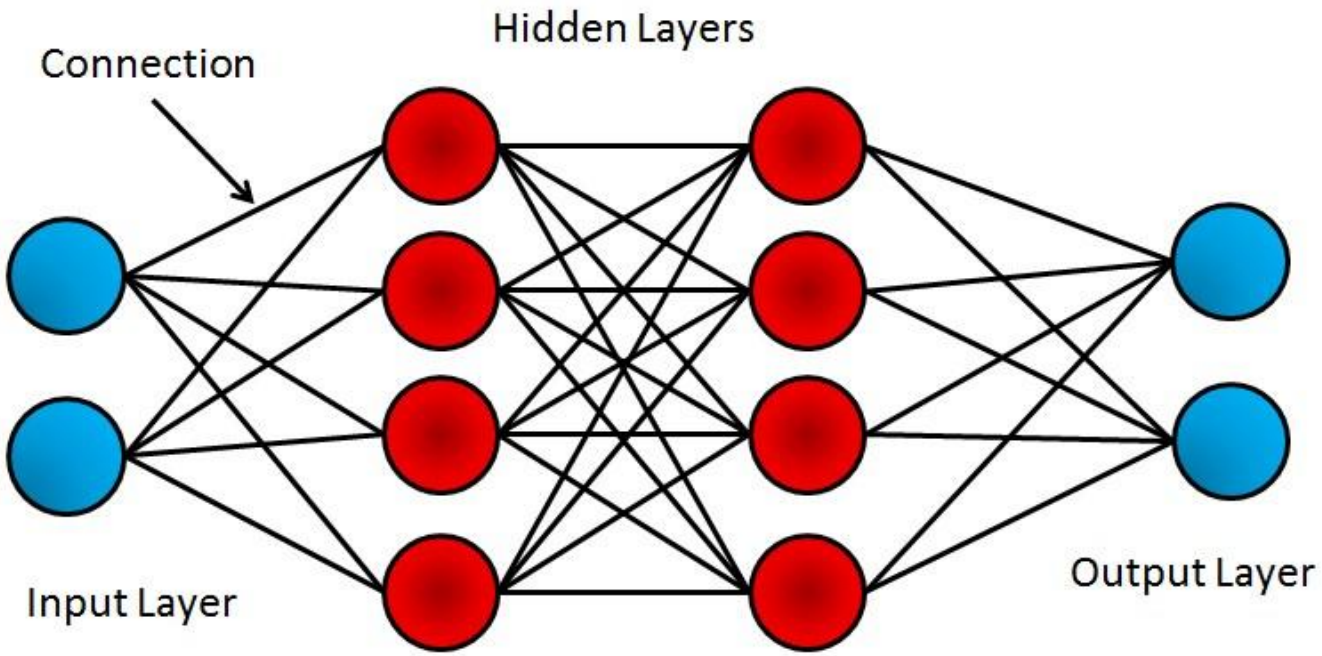


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

ΕΞΕΤΑΖΟΝΤΑΣ ΤΙΣ ΕΞΟΔΟΥΣ ΤΩΝ ΑΛΛΩΝ ΝΕΥΡΩΝΩΝ ΠΟΥ ΤΟΝ ΤΡΟΦΟΔΟΤΟΥΝ

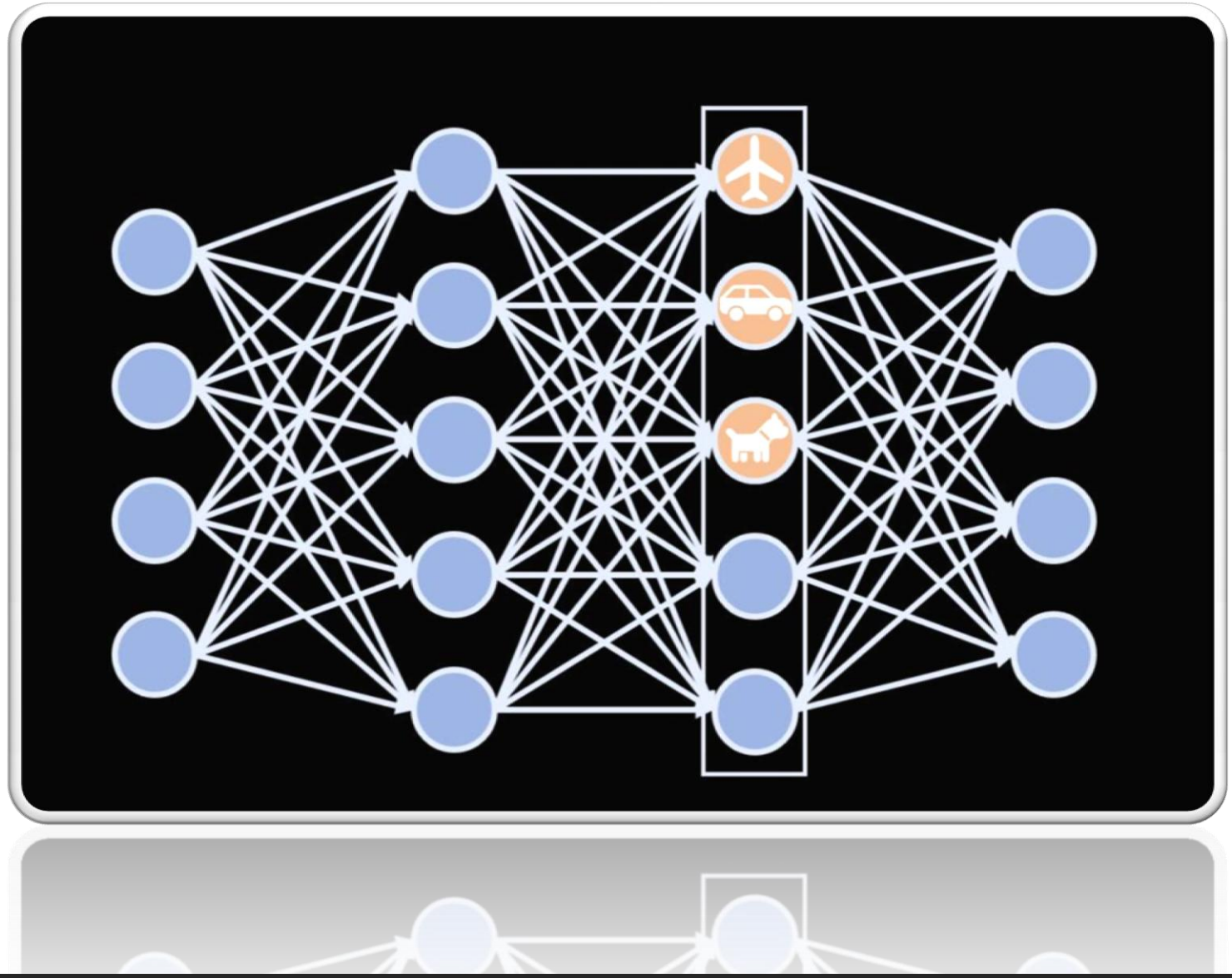


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Νευρωνικά Δίκτυα

- Τα νευρωνικά δίκτυα αποτελούνται από κόμβους που συνδέονται με άλλους κόμβους.
- **Κάθε σύνδεση** μεταξύ των κόμβων αποτελείτε από αριθμούς που **εκφράζουν την κλήση μιας ευθείας** η οποία εκφράζει μια απλή παλινδρόμηση πάνω στα δεδομένα μας.
- Οι κόμβοι που βρίσκονται ανάμεσα στους κόμβους εισόδου και εξόδου ονομάζονται **Hidden layers**. Και είναι ιδιαίτερα βασικά διότι στην αρχή κατασκευής ενός νευρωνικού δικτύου είναι απαραίτητος ο καθορισμός της ποσότητας των Hidden layers.



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

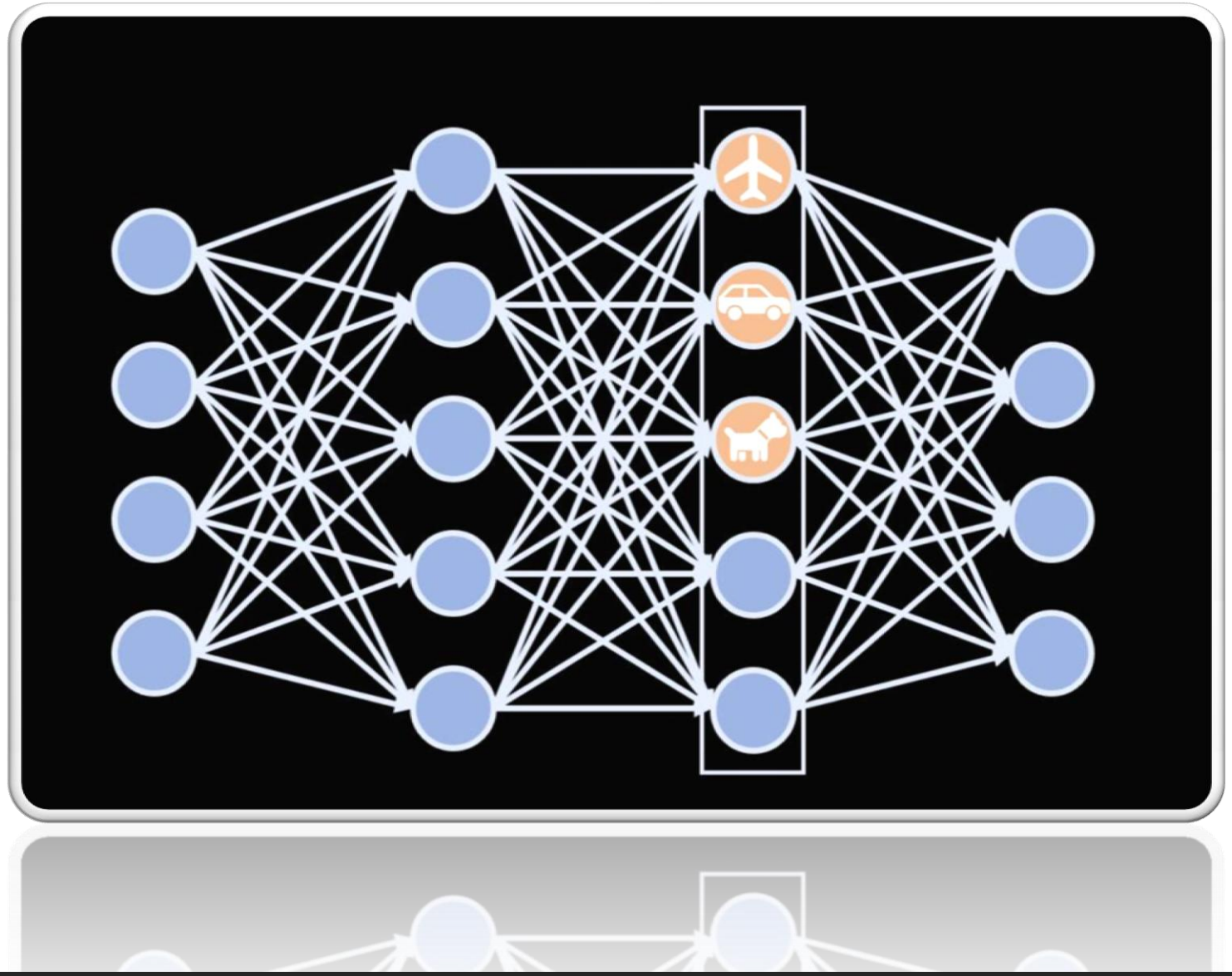
UNIVERSITY OF PIRAEUS

Νευρωνικά Δίκτυα

- **Η συνάρτηση** που χρησιμοποιείται για τα σύνθετα νευρωνικά δίκτυα είναι μια **sigmoid function**.Είναι υπεύθυνη για να δέχεται εντολές εισόδου ορίζοντας την έξοδο του κάθε κόμβου.

- Η κατασκευή ενός νευρωνικού δικτύου πραγματοποιείται μέσα από δεδομένα «εκπαίδευσης». Η τυπική προσέγγιση είναι ένας αλγόριθμος που ονομάζεται **Backpropagation.**

- Η τεχνική **Backpropagation** επιτρέπει στο δίκτυο να **αναγνωρίζει ποιοι κόμβοι έχουν την μεγαλύτερη επιρροή στην έξοδο, επιτρέποντας να ενισχύσουμε η να αποδυναμώσουμε τα connections μεταξύ των κόμβων με σκοπό να καταλήξουμε στο επιθυμητό αποτέλεσμα.**



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

ΒΑΘΙΑ ΜΑΘΗΣΗ
DEEP LEARNING

ΤΑ ΒΑΘΥΤΕΡΑ
ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

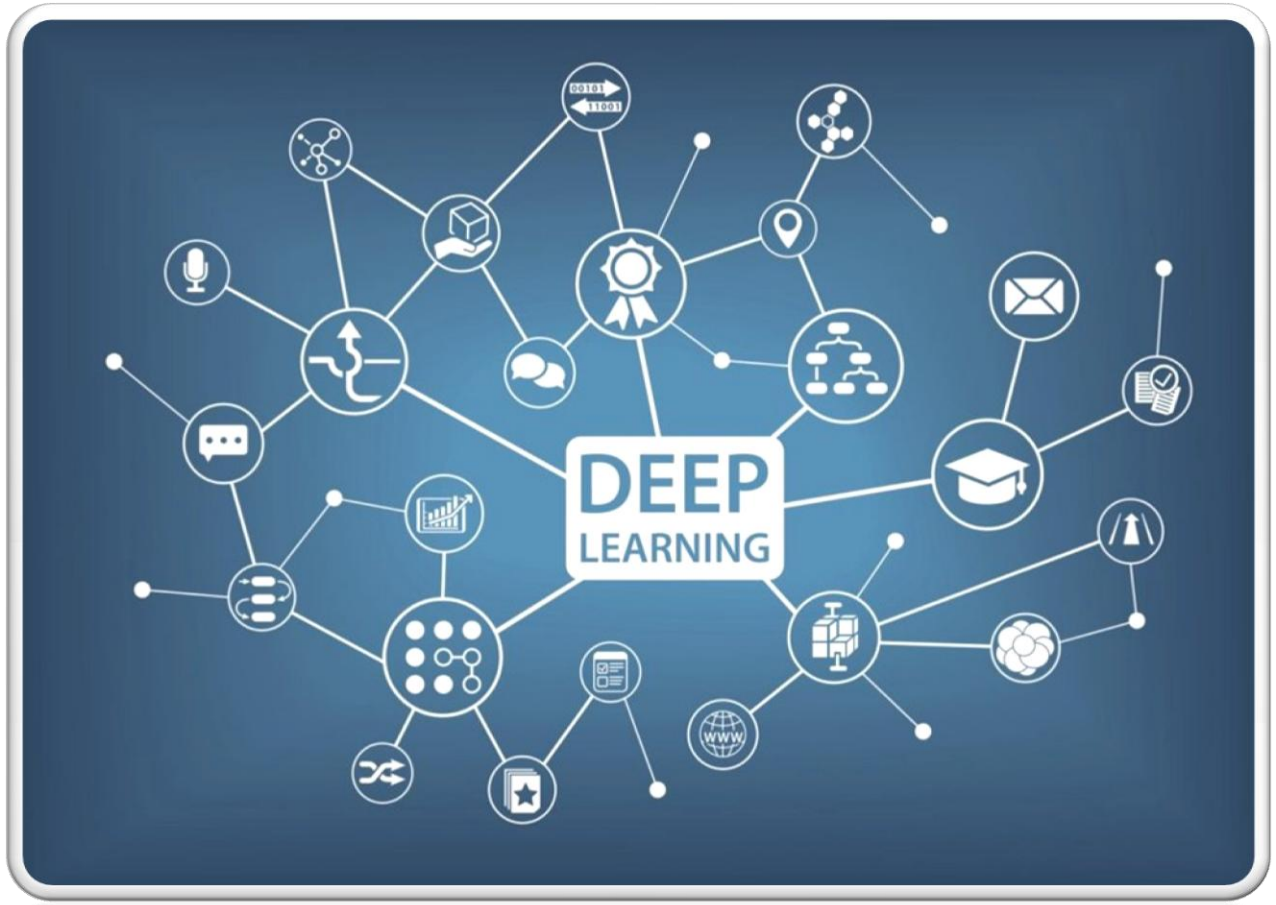
UNIVERSITY OF PIRAEUS

Βαθιά Μάθηση(1/2)

Η Βαθιά Μάθηση (Deep Learning) είναι μία τεχνική μηχανικής μάθησης (υπάρχουν πολλές τεχνικές που απαρτίζουν την μηχανική μάθηση) στην οποία αρκετά «επίπεδα» απλών μονάδων επεξεργασίας συνδέονται σε ένα δίκτυο, το ένα πίσω από το άλλο, με αποτέλεσμα η είσοδος στο σύστημα να διέρχεται διαδοχικά μέσα από κάθε ένα από αυτά.

Η συγκεκριμένη αρχιτεκτονική επινοήθηκε με πρότυπο την επεξεργασία των οπτικών πληροφοριών στον εγκέφαλο, δηλ. των πληροφοριών που εισέρχονται μέσω των ματιών και συλλαμβάνονται από τον αμφιβληστροειδή χιτώνα, προχωρούν μέσα στο οπτικό νεύρο και φτάνουν στον εγκέφαλο.

Αυτό το βάθος επιτρέπει στο δίκτυο να μαθαίνει πιο σύνθετες δομές

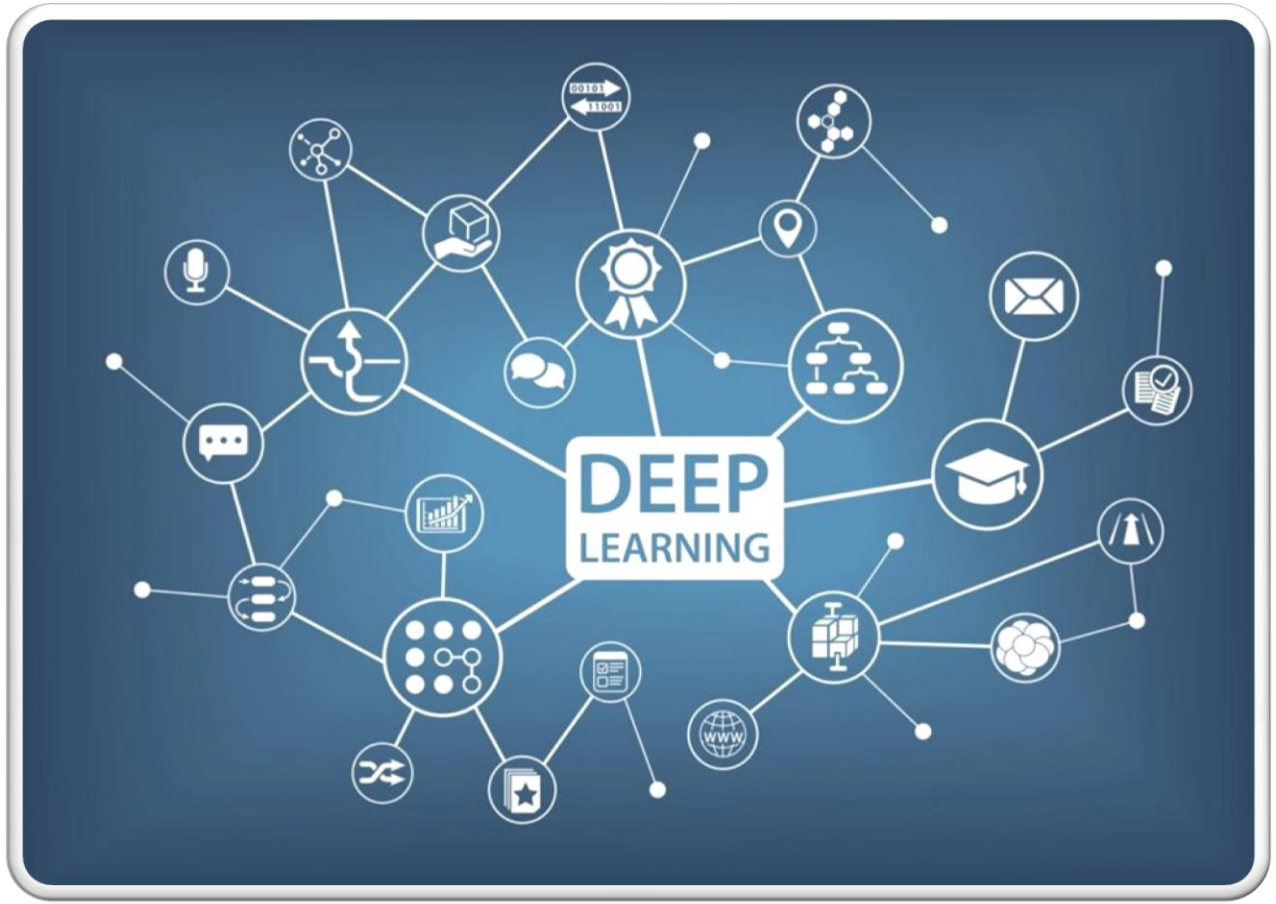


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Βαθιά Μάθηση(2/2)

- Όπως τα περισσότερα μοντέλα μηχανικής μάθησης, τα νευρωνικά δίκτυα **είναι επιρρεπή στην υπερπροσαρμογή**.
- Η βελτίωση τους πραγματοποιείται μέσω της «**Κανονικοποίησης**».
- Ένας σύνηθες τρόπος κανονικοποίησης είναι **μέθοδος dropout.** Κατά την εκπαίδευση **κάθε νευρώνας απενεργοποιείται με τυχαίο τρόπο με κάποια συγκεκριμένη πιθανότητα**.
- Έτσι μαθαίνει ο κάθε νευρώνας να μην εξαρτάται από οποιονδήποτε μεμονωμένο νευρώνα γεγονός που «**αντιμετωπίζει**» την υπερπροσαρμογή.

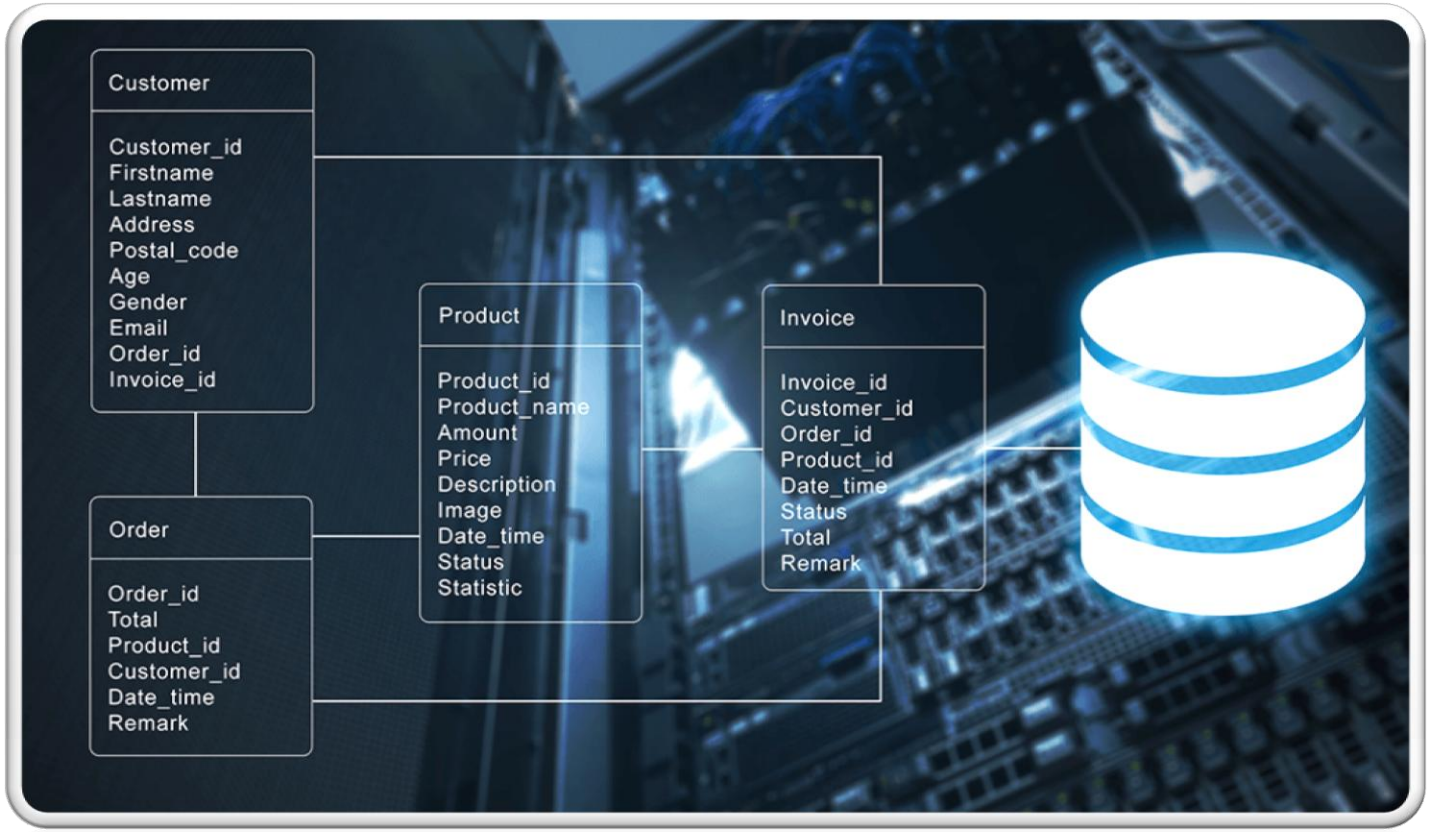


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Βάσεις Δεδομένων και SQL - Structured Query Language

ΣΥΣΤΗΜΑΤΑ ΣΧΕΔΙΑΣΜΕΝΑ ΓΙΑ ΤΗΝ ΑΠΟΘΗΚΕΥΣΗ ΚΑΙ ΑΝΑΖΗΤΗΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

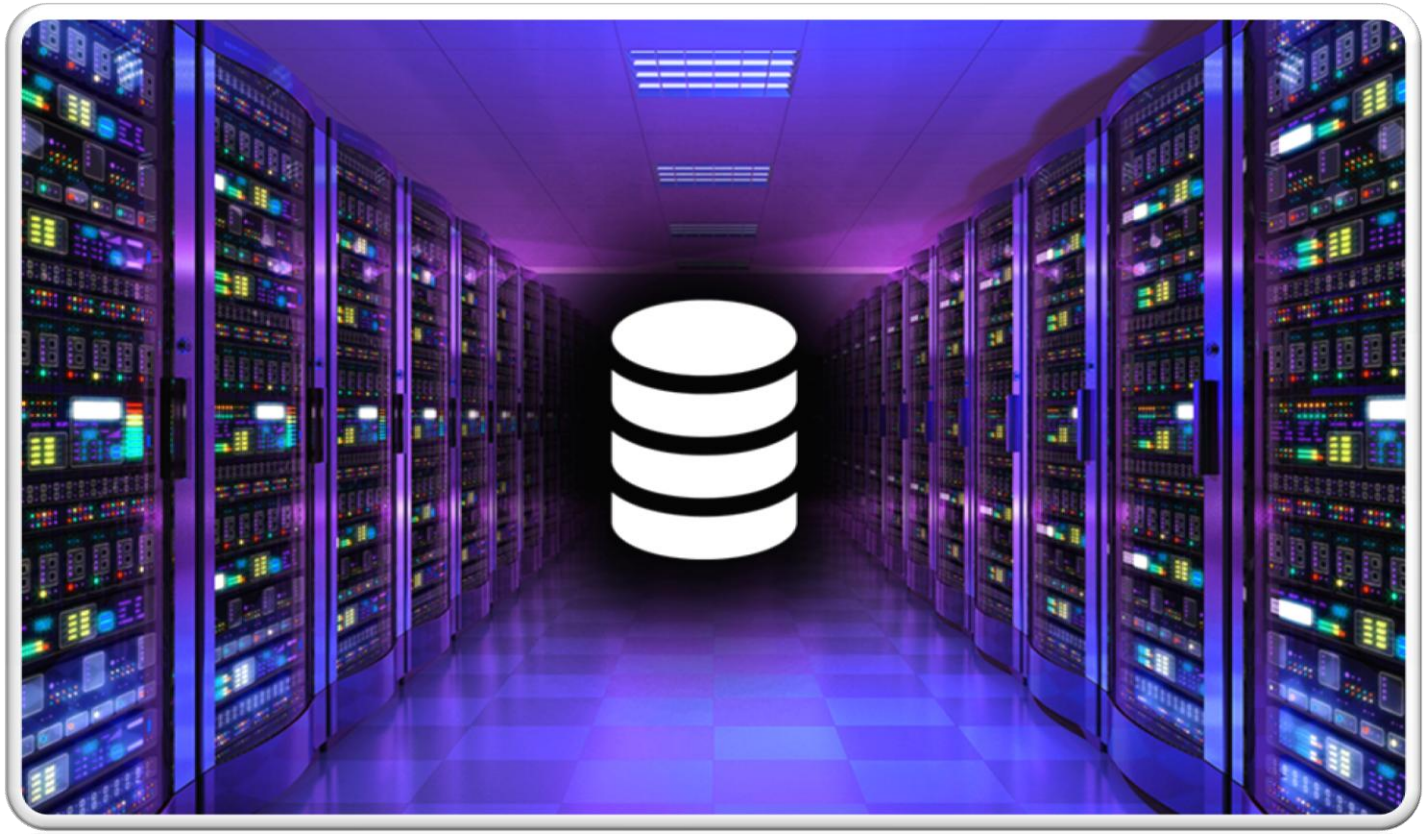


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Βάσεις Δεδομένων και SQL (1/4)

- **Ο κύριος όγκος των βάσεων δεδομένων είναι relational** (σχεσιακές), όπως η PostgreSQL, MySQL και ο SQL server, οι οποίες αποθηκεύουν **τα δεδομένα σε πίνακες.**



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

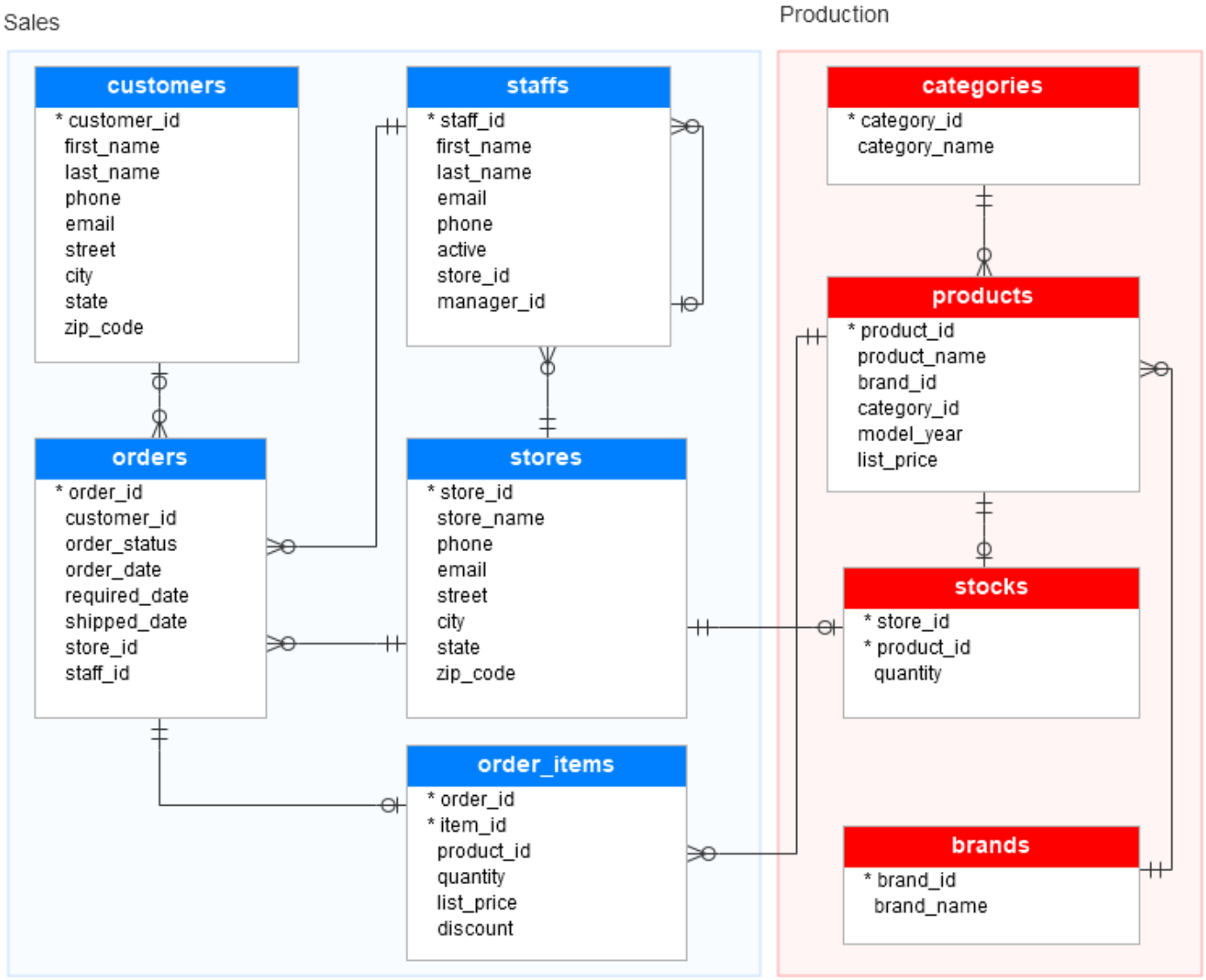
Βάσεις Δεδομένων και SQL (2/4)

Employee table

Empno (PK)	Ename	Job	Deptno (FK)
101	A	Salesman	10
102	B	Manager	10
103	c	Manager	20

Department table

Deptno (PK)	dname	loc
10	Sales	Chicago
20	Sales	Chicago
30	Finance	New York



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Βάσεις Δεδομένων και SQL (3/4)

Basic SQL query structure

```
SELECT Attributes
  FROM relations
 WHERE condition
```

For example:

```
SELECT sid,sname
  FROM students
 WHERE sid=1122
```

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Βάσεις Δεδομένων και SQL (4/4)

```
-- get  employees who joined company in 2000
SELECT
    first_name
FROM
    employees
WHERE
    YEAR(hire_date) = 2000
```

SELECT clause

FROM clause

WHERE clause

Comment

Predicate