

Μάθημα 1

**ΕΞΟΥΣΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΨΗΦΙΑΚΟ
ΚΑΙ ΔΙΑΔΙΚΤΥΑΚΟ ΠΕΡΙΕΧΟΜΕΝΟ**

Εισαγωγικά Θέματα

- Τα τελευταία χρόνια έχουν δημιουργηθεί τεράστιες αποθήκες δεδομένων σε οργανισμούς, εταιρείες, ινστιτούτα, άτομα και στο διαδίκτυο με εκατοντάδες exabytes δεδομένων

Εισαγωγικά Θέματα

- Τα τελευταία χρόνια έχουν δημιουργηθεί τεράστιες αποθήκες δεδομένων σε οργανισμούς, εταιρείες, ινστιτούτα, άτομα και στο διαδίκτυο με εκατοντάδες exabytes δεδομένων
- Τα δεδομένα αποθηκεύονται σε Βάσεις Δεδομένων (Data Bases) τόσο για την ευκολότερη αποθήκευση όσο και για την πιο αποδοτική επεξεργασία τους

Εισαγωγικά Θέματα

- Τα τελευταία χρόνια έχουν δημιουργηθεί τεράστιες αποθήκες δεδομένων σε οργανισμούς, εταιρείες, ινστιτούτα, άτομα και στο διαδίκτυο με εκατοντάδες exabytes δεδομένων
- Τα δεδομένα αποθηκεύονται σε Βάσεις Δεδομένων (Data Bases) τόσο για την ευκολότερη αποθήκευση όσο και για την πιο αποδοτική επεξεργασία τους
- Εξόρυξη Δεδομένων (Data Mining): Ανάλυση δεδομένων με σκοπό την εξαγωγή γνώσης από αυτά με την χρήση εξειδικευμένων τεχνικών μηχανικής μάθησης

Εισαγωγικά Θέματα

- Τα τελευταία χρόνια έχουν δημιουργηθεί τεράστιες αποθήκες δεδομένων σε οργανισμούς, εταιρείες, ινστιτούτα, άτομα και στο διαδίκτυο με εκατοντάδες exabytes δεδομένων
- Τα δεδομένα αποθηκεύονται σε Βάσεις Δεδομένων (Data Bases) τόσο για την ευκολότερη αποθήκευση όσο και για την πιο αποδοτική επεξεργασία τους
- Εξόρυξη Δεδομένων (Data Mining): Ανάλυση δεδομένων με σκοπό την εξαγωγή γνώσης από αυτά με την χρήση εξειδικευμένων τεχνικών μηχανικής μάθησης
- Πως γίνεται αυτό; Με την αναγνώριση μοτίβων (προτύπων) που υπάρχουν μέσα στα δεδομένα

Εισαγωγικά Θέματα

- Τα τελευταία χρόνια έχουν δημιουργηθεί τεράστιες αποθήκες δεδομένων σε οργανισμούς, εταιρείες, ινστιτούτα, άτομα και στο διαδίκτυο με εκατοντάδες exabytes δεδομένων
- Τα δεδομένα αποθηκεύονται σε Βάσεις Δεδομένων (Data Bases) τόσο για την ευκολότερη αποθήκευση όσο και για την πιο αποδοτική επεξεργασία τους
- Εξόρυξη Δεδομένων (Data Mining): Ανάλυση δεδομένων με σκοπό την εξαγωγή γνώσης από αυτά με την χρήση εξειδικευμένων τεχνικών μηχανικής μάθησης
- Πως γίνεται αυτό; Με την αναγνώριση μοτίβων (προτύπων) που υπάρχουν μέσα στα δεδομένα
- Το 98% της ανθρώπινης μάθησης είναι η αναγνώριση μοτίβων (Ray Kurzweil)

Εισαγωγικά Θέματα

- Τα τελευταία χρόνια έχουν δημιουργηθεί τεράστιες αποθήκες δεδομένων σε οργανισμούς, εταιρείες, ινστιτούτα, άτομα και στο διαδίκτυο με εκατοντάδες exabytes δεδομένων
- Τα δεδομένα αποθηκεύονται σε Βάσεις Δεδομένων (Data Bases) τόσο για την ευκολότερη αποθήκευση όσο και για την πιο αποδοτική επεξεργασία τους
- Εξόρυξη Δεδομένων (Data Mining): Ανάλυση δεδομένων με σκοπό την εξαγωγή γνώσης από αυτά με την χρήση εξειδικευμένων τεχνικών μηχανικής μάθησης
- Πως γίνεται αυτό; Με την αναγνώριση μοτίβων (προτύπων) που υπάρχουν μέσα στα δεδομένα
- Το 98% της ανθρώπινης μάθησης είναι η αναγνώριση μοτίβων (Ray Kurzweil)
- Ο πιο χρήσιμος τρόπος είναι η καταχώρηση των δεδομένων σε μία Βάση Δεδομένων

Εισαγωγικά Θέματα

Η Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων (Knowledge Discovery in Databases-KDD) είναι ακριβώς το αντικείμενο της Εξόρυξης Δεδομένων και αναφέρεται στην ανακάλυψη χρήσιμης πληροφορίας από τα δεδομένα, η οποία μπορεί από τον ειδικό να μετουσιωθεί σε γνώση

Εισαγωγικά Θέματα

- Τι είναι Πληροφορία και τι η Γνώση?

Εισαγωγικά Θέματα

- Τι είναι Πληροφορία και τι η Γνώση?

Παράδειγμα

- Ένας επιδημιολόγος συλλέγει δεδομένα ανά μήνα για τα περιστατικά γρίπης

Εισαγωγικά Θέματα

- Τι είναι Πληροφορία και τι η Γνώση?

Παράδειγμα

- Ένας επιδημιολόγος συλλέγει δεδομένα ανά μήνα για τα περιστατικά γρίπης
- Αφού τα καταχωρήσει σε βάση δεδομένων (π.χ. με την βοήθεια του excel) παρατηρεί ότι τα περισσότερα περιστατικά συμβαίνουν του χειμερινούς μήνες. Αυτό το γεγονός είναι **πληροφορία** και τίποτα περισσότερο

Εισαγωγικά Θέματα

- Τι είναι Πληροφορία και τι η Γνώση?

Παράδειγμα

- Ένας επιδημιολόγος συλλέγει δεδομένα ανά μήνα για τα περιστατικά γρίπης
- Αφού τα καταχωρήσει σε βάση δεδομένων (π.χ. με την βοήθεια του excel) παρατηρεί ότι τα περισσότερα περιστατικά συμβαίνουν του χειμερινούς μήνες. Αυτό το γεγονός είναι **πληροφορία** και τίποτα περισσότερο
- Ο ειδικός προσπαθώντας να εξηγήσει το γεγονός αυτό φτάνει στο συμπέρασμα ότι η υγρασία που υπάρχει στο περιβάλλον τους χειμερινούς μήνες βοηθά στην επέκταση της ζωής του ιού της γρίπης με αποτέλεσμα αυτός να βρίσκει εύκολα ξενιστή και να διαδίδεται πιο αποδοτικά

Εισαγωγικά Θέματα

- Τι είναι Πληροφορία και τι η Γνώση?

Παράδειγμα

- Ένας επιδημιολόγος συλλέγει δεδομένα ανά μήνα για τα περιστατικά γρίπης
- Αφού τα καταχωρήσει σε βάση δεδομένων (π.χ. με την βοήθεια του excel) παρατηρεί ότι τα περισσότερα περιστατικά συμβαίνουν του χειμερινούς μήνες. Αυτό το γεγονός είναι **πληροφορία** και τίποτα περισσότερο
- Ο ειδικός προσπαθώντας να εξηγήσει το γεγονός αυτό φτάνει στο συμπέρασμα ότι η υγρασία που υπάρχει στο περιβάλλον τους χειμερινούς μήνες βοηθά στην επέκταση της ζωής του ιού της γρίπης με αποτέλεσμα αυτός να βρίσκει εύκολα ξενιστή και να διαδίδεται πιο αποδοτικά
- Το τελευταίο είναι η **γνώση** που προέκυψε από την πληροφορία.

Εισαγωγικά Θέματα

Τι είναι τα Δεδομένα και τι τα Πρότυπα;

Εισαγωγικά Θέματα

Τι είναι τα Δεδομένα και τι τα Πρότυπα;

- Δεδομένα: Περιγράφουν οντότητες ή/και συσχετίσεις του πραγματικού κόσμου, π.χ. ένα σύνολο εγγραφών που αναφέρονται στις τραπεζικές συναλλαγές μέσα σε ένα χρονικό διάστημα

Εισαγωγικά Θέματα

Τι είναι τα Δεδομένα και τι τα Πρότυπα;

- Δεδομένα: Περιγράφουν οντότητες ή/και συσχετίσεις του πραγματικού κόσμου, π.χ. ένα σύνολο εγγραφών που αναφέρονται στις τραπεζικές συναλλαγές μέσα σε ένα χρονικό διάστημα
- Πρότυπα: Ένα υποσύνολο των δεδομένων, το οποίο έχει μία συγκεκριμένη δομή. Το εξαγόμενο πρότυπο πρέπει να είναι έγκυρο με κάποιο σημαντικό βαθμό βεβαιότητας εκφρασμένο σε % ποσοστό.

Εισαγωγικά Θέματα

Παράδειγμα

- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο

Εισαγωγικά Θέματα

Παράδειγμα

- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου

Εισαγωγικά Θέματα

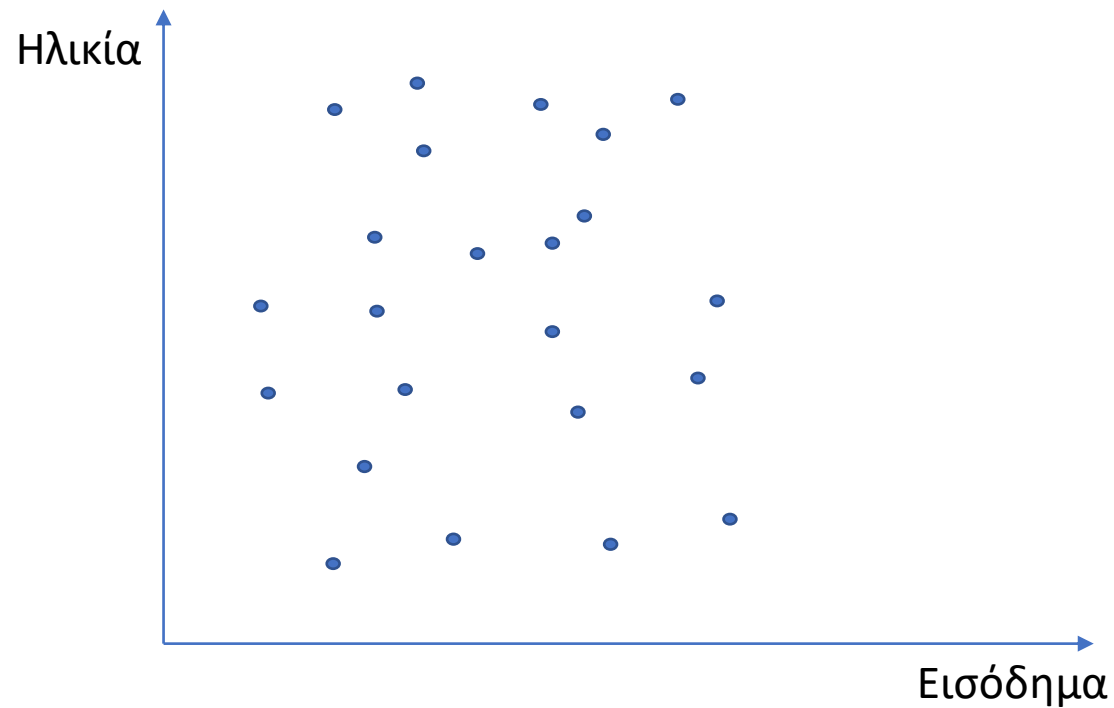
Παράδειγμα

- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου
- Συλλέγει δεδομένα που περιέχουν την ηλικία και το εισόδημα.

Εισαγωγικά Θέματα

Παράδειγμα

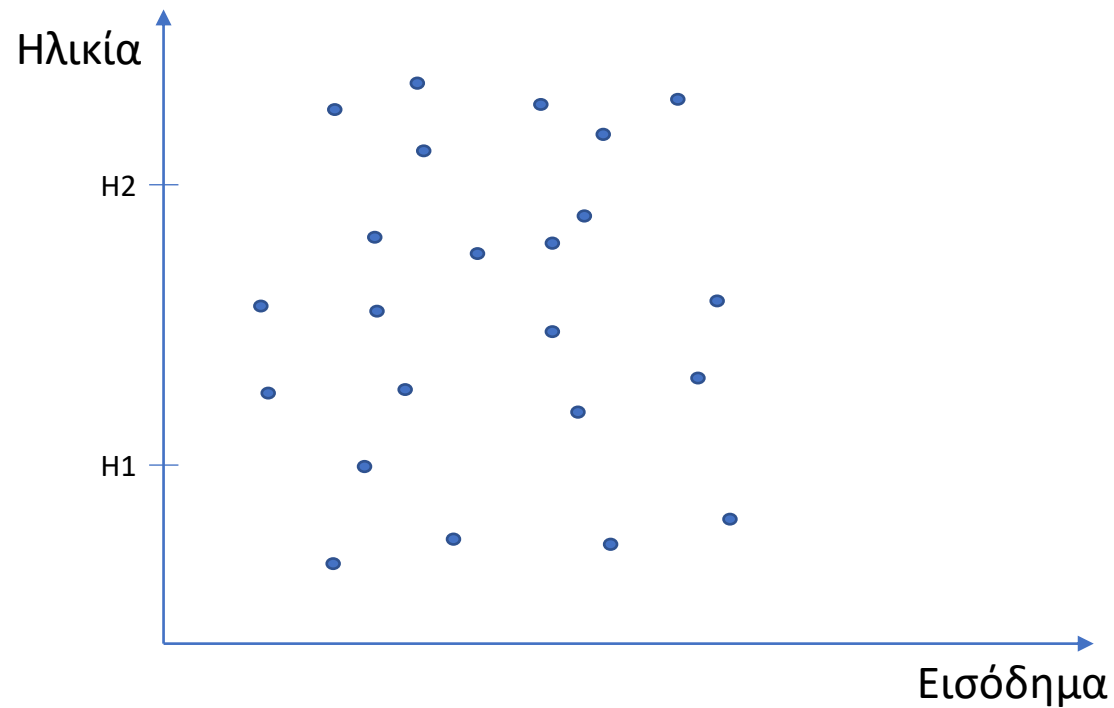
- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου
- Συλλέγει δεδομένα που περιέχουν την ηλικία και το εισόδημα.



Εισαγωγικά Θέματα

Παράδειγμα

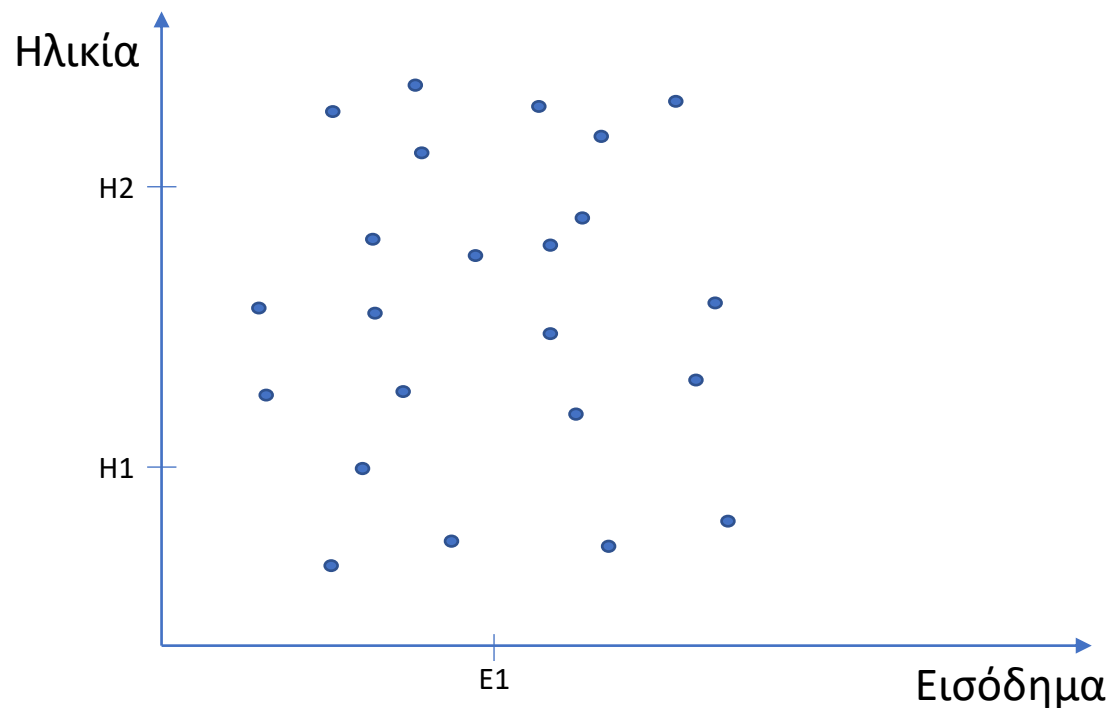
- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου
- Συλλέγει δεδομένα που περιέχουν την ηλικία και το εισόδημα.



Εισαγωγικά Θέματα

Παράδειγμα

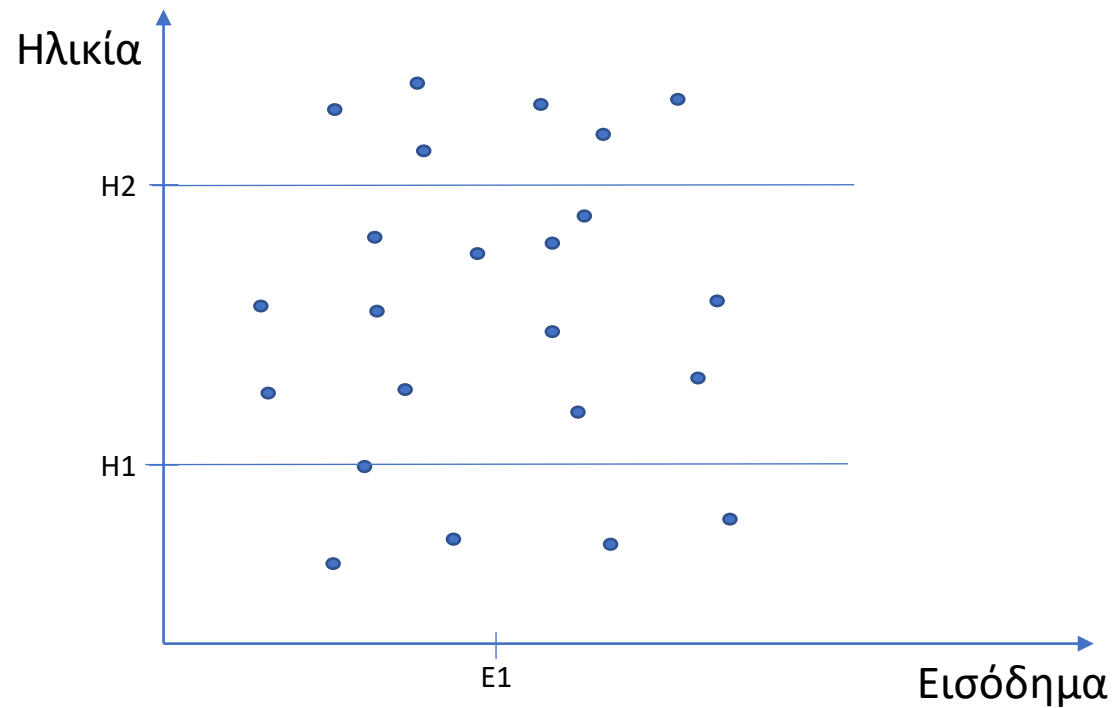
- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου
- Συλλέγει δεδομένα που περιέχουν την ηλικία και το εισόδημα.



Εισαγωγικά Θέματα

Παράδειγμα

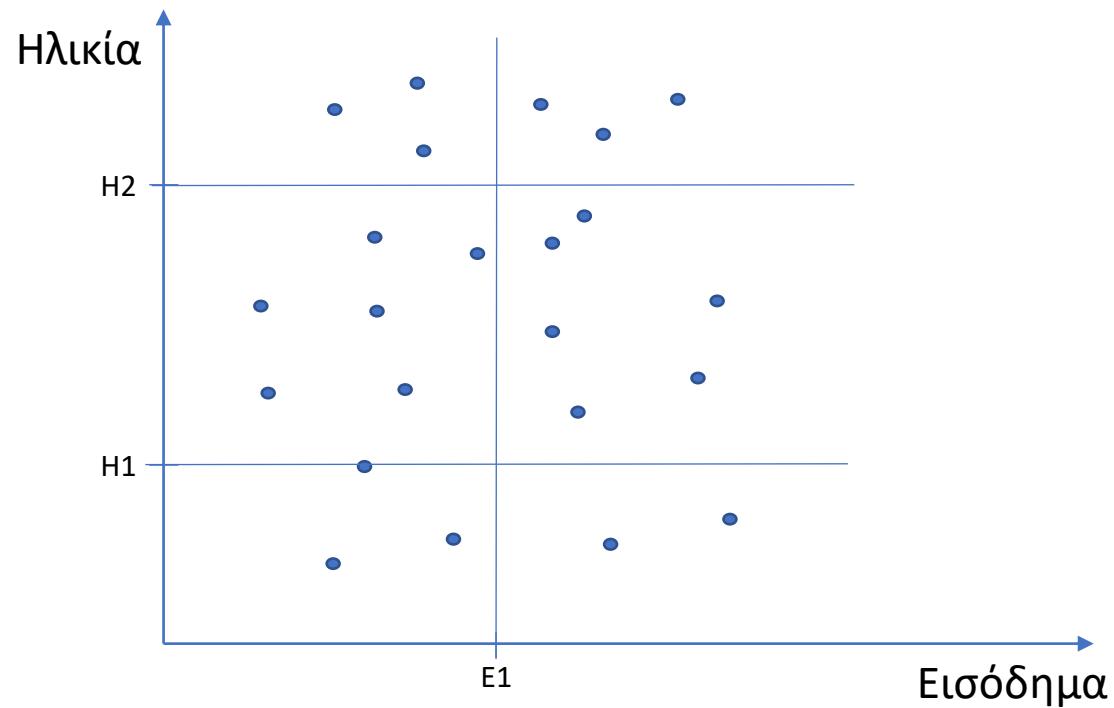
- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου
- Συλλέγει δεδομένα που περιέχουν την ηλικία και το εισόδημα.



Εισαγωγικά Θέματα

Παράδειγμα

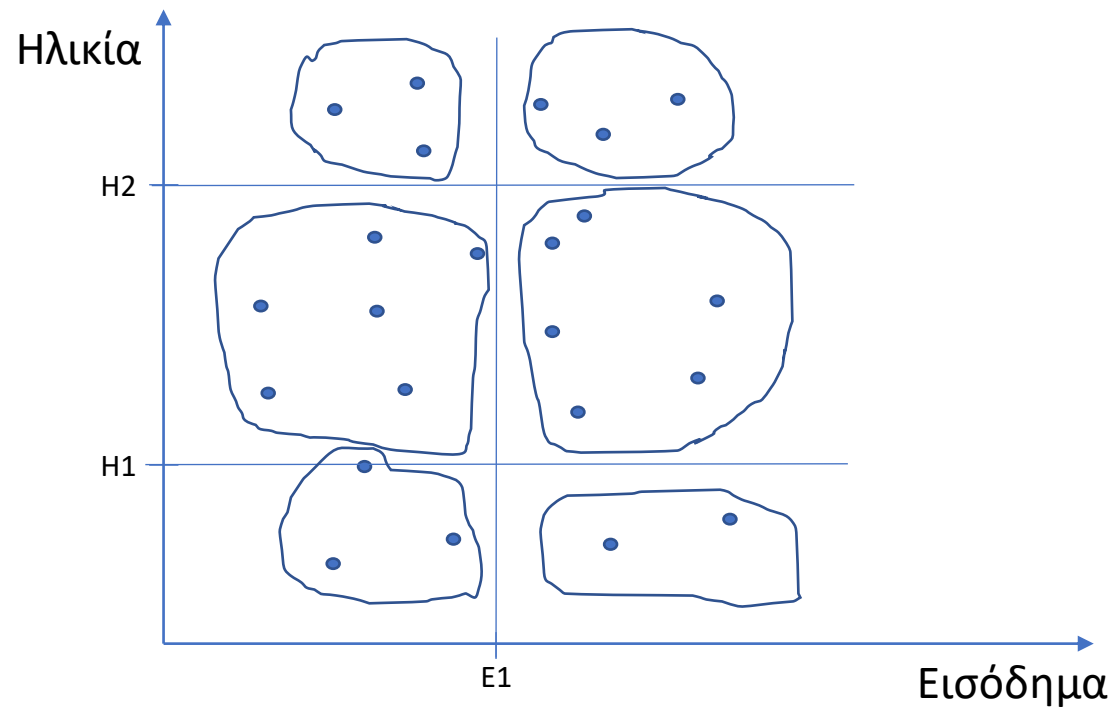
- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου
- Συλλέγει δεδομένα που περιέχουν την ηλικία και το εισόδημα.



Εισαγωγικά Θέματα

Παράδειγμα

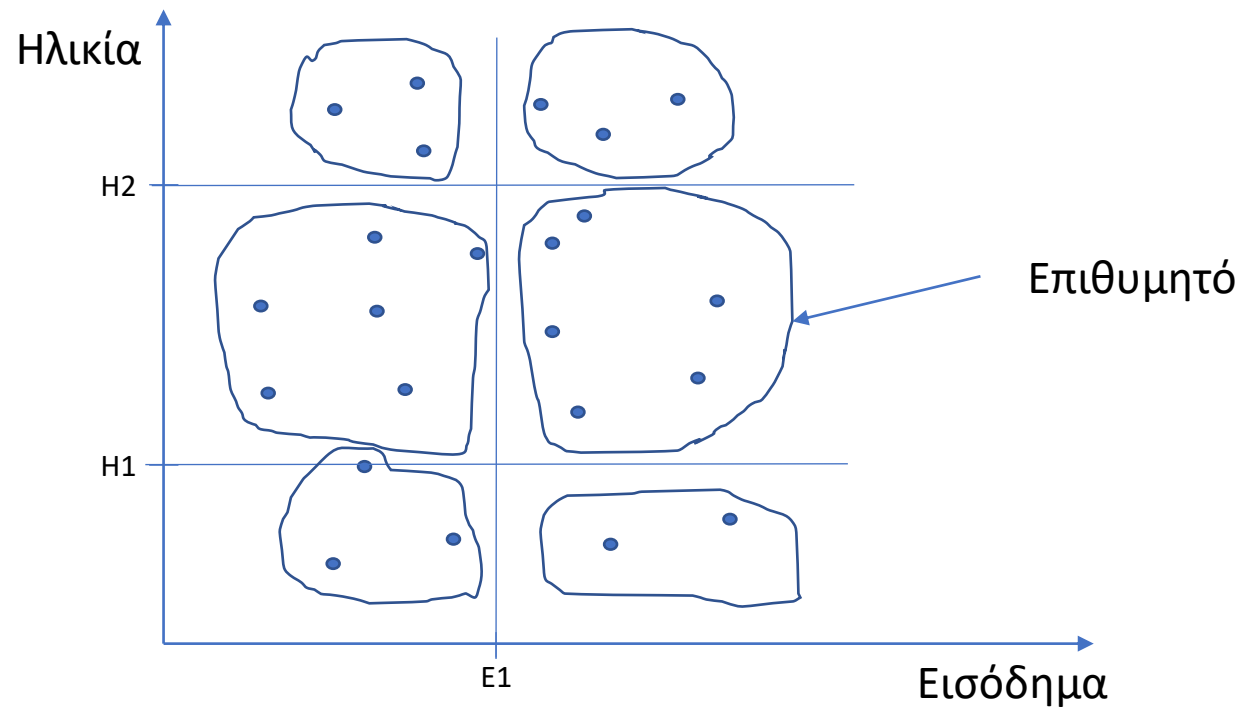
- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου
- Συλλέγει δεδομένα που περιέχουν την ηλικία και το εισόδημα.



Εισαγωγικά Θέματα

Παράδειγμα

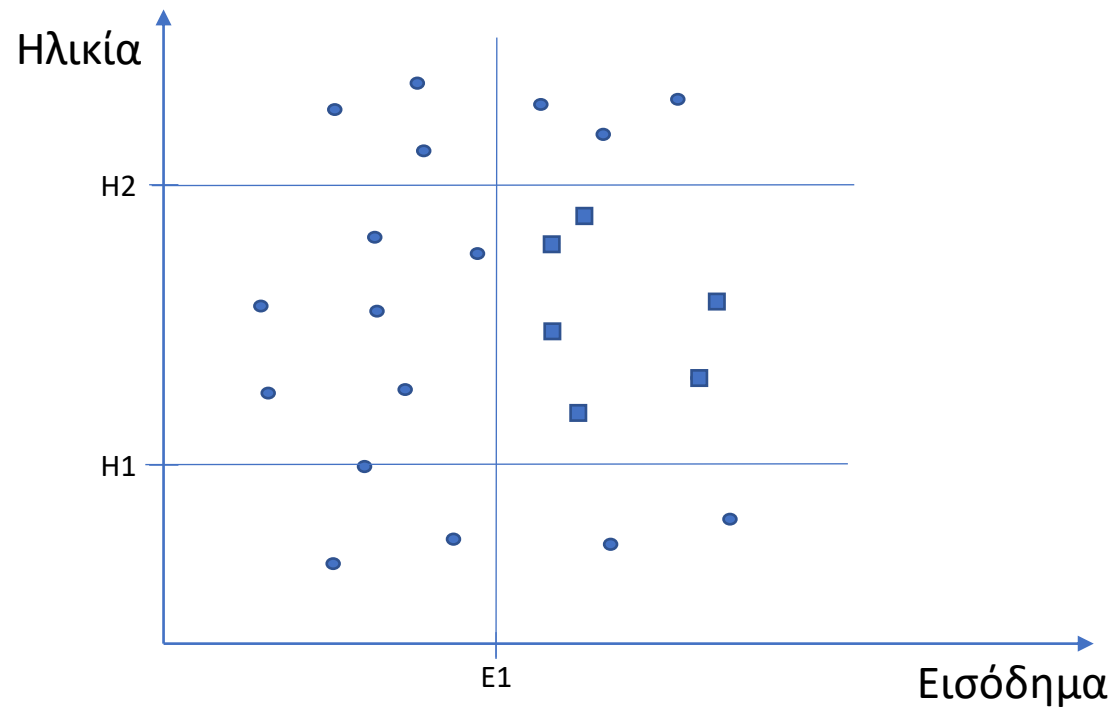
- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου
- Συλλέγει δεδομένα που περιέχουν την ηλικία και το εισόδημα.



Εισαγωγικά Θέματα

Παράδειγμα

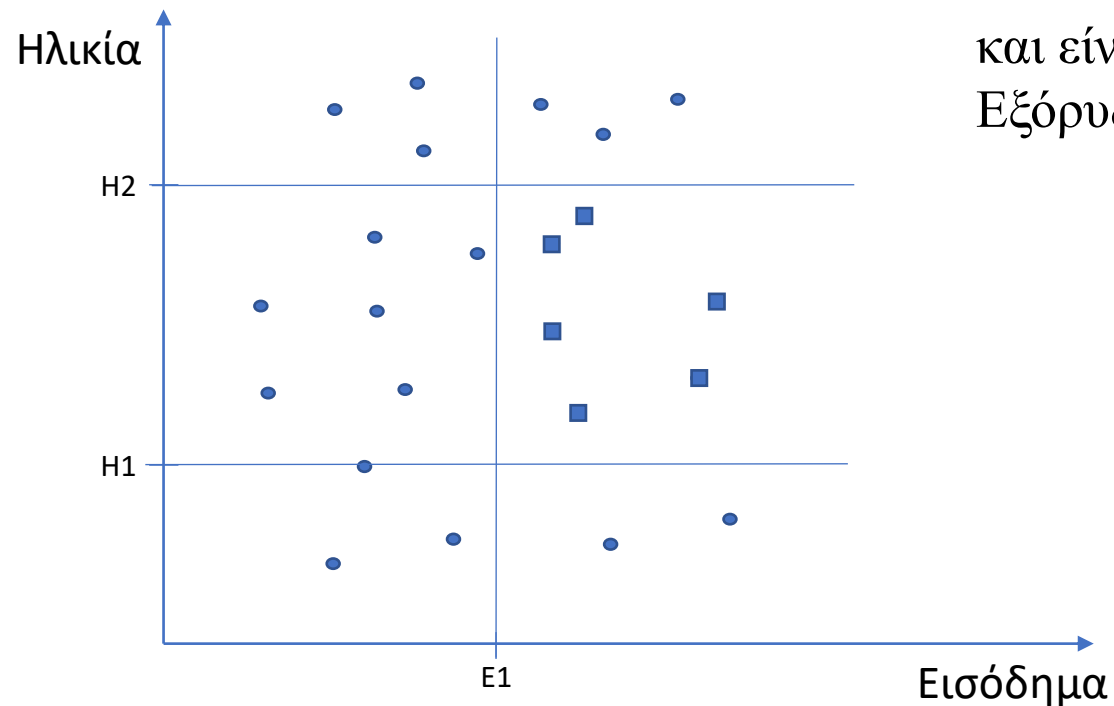
- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου
- Συλλέγει δεδομένα που περιέχουν την ηλικία και το εισόδημα.



Εισαγωγικά Θέματα

Παράδειγμα

- Μία αυτοκινητοβιομηχανία θέλει να λανσάρει ένα καινούργιο και ακριβό μοντέλο
- Θέλει να αναλύσει την στρατηγική της όσον αφορά σε ποιο γκρουπ ανθρώπων πρέπει να στοχεύσει την διαφήμιση του μοντέλου
- Συλλέγει δεδομένα που περιέχουν την ηλικία και το εισόδημα.



Αυτό λέγεται πρόβλημα **κατηγοριοποίησης** και είναι ένα από τα βασικά προβλήματα της Εξόρυξης Δεδομένων

Εισαγωγικά Θέματα

Τι μπορούν να μάθουν οι υπολογιστές:

Εισαγωγικά Θέματα

Τι μπορούν να μάθουν οι υπολογιστές:

- **Γεγονός:** Απλή δήλωση Αλήθειας ή Ψεύδους

Εισαγωγικά Θέματα

Τι μπορούν να μάθουν οι υπολογιστές:

- **Γεγονός:** Απλή δήλωση Αλήθειας ή Ψεύδους
- **Έννοια:** Μία ομάδα αντικειμένων (δηλ. προτύπων) που σχετίζονται με σύμβολα ή συμβάντα (δηλ. γεγονότα), τα οποία έχουν κοινά χαρακτηριστικά.

Εισαγωγικά Θέματα

Τι μπορούν να μάθουν οι υπολογιστές:

- **Γεγονός:** Απλή δήλωση Αλήθειας ή Ψεύδους
- **Έννοια:** Μία ομάδα αντικειμένων (δηλ. προτύπων) που σχετίζονται με σύμβολα ή συμβάντα (δηλ. γεγονότα), τα οποία έχουν κοινά χαρακτηριστικά.
- **Διαδικασία:** Μία σειριακή εκτέλεση ενεργειών για την επίτευξη ενός στόχου (π.χ. την επίλυση ενός προβλήματος)

Εισαγωγικά Θέματα

Τι μπορούν να μάθουν οι υπολογιστές:

- **Γεγονός:** Απλή δήλωση Αλήθειας ή Ψεύδους
- **Έννοια:** Μία ομάδα αντικειμένων (δηλ. προτύπων) που σχετίζονται με σύμβολα ή συμβάντα (δηλ. γεγονότα), τα οποία έχουν κοινά χαρακτηριστικά.
- **Διαδικασία:** Μία σειριακή εκτέλεση ενεργειών για την επίτευξη ενός στόχου (π.χ. την επίλυση ενός προβλήματος)
- **Αρχή:** Κανόνες που δηλώνονται αξιωματικά (αποτελεί το υψηλότερο επίπεδο μάθησης ενός υπολογιστή)

Καλό Απόγευμα