

# Εξόρυξη δεδομένων με χρήση τεχνικών μηχανικής μάθησης

Τμ. Μηχανικών Πληροφορικής και Υπολογιστών



# Μετασχηματισμός των Δεδομένων

# Μετασχηματισμός των Δεδομένων

- Μείωση των δεδομένων (data reduction)
- Μετατροπή των δεδομένων (data transformation)

# Data reduction

## Μείωση στηλών (Correlation)

Τι ορίζουμε ως Συσχέτιση.

Είναι η σχέση που μπορεί να έχει η μια τιμή με μια άλλη η οποία εξαρτάται από το πλήθος των μεταβλητών. Δηλαδή αν έχουμε συσχέτιση μεταξύ 2 τιμών τότε έχουμε απλή συσχέτιση ενώ αν έχουμε μεταξύ περισσότερων τότε έχουμε πολλαπλή συσχέτιση.

# Data reduction

## Μείωση στηλών (Correlation)

**Συσχέτιση** δύο τυχαίων μεταβλητών ορίζουμε τη συναρτησιακή σχέση εξάρτησης της μίας μεταβλητής ως προς την άλλη. Αν οι μεταβλητές είναι δύο τότε έχουμε απλή συσχέτιση, ενώ αν είναι περισσότερες έχουμε την πολλαπλή συσχέτιση.

# Data reduction

## Μείωση στηλών (Correlation)

Ορίζουμε τον **Συντελεστή Συσχέτισης** με σκοπό να μετρήσουμε το βαθμό συσχέτισης των δυο μεταβλητών. Όταν ο συντελεστής συσχέτισης ισούται με 1 υποδηλώνεται η απόλυτη συσχέτιση των δυο μεταβλητών ενώ με 0 υποδηλώνεται ότι δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών.

# Data reduction

## Δημιουργία νέου πεδίου (PCA)

Μέσω της Ανάλυσης Πρωταρχικών Συνιστωσών – Principal Component Analysis (PCA), μπορούμε από ένα σύνολο μεταβλητών να δημιουργήσουμε ένα αντιπροσωπευτικό υποσύνολο

# Μετατροπή των δεδομένων (transformation)

Κατά τη διαδικασία του μετασχηματισμού μπορούμε να μετατρέψουμε τα δεδομένα από τη μια μορφή στην άλλη.

Για παράδειγμα μπορούμε να μετατρέψουμε αριθμητικά δεδομένα σε True – False.

Αν θεωρήσουμε ότι οι τιμές που μπορεί να πάρει είναι 0-10, τότε θα πρέπει να ορίσουμε το εύρος των τιμών για τις οποίες θέλουμε να παίρνει τιμή true και για ποιες τιμή false.



# Μετασχηματισμός των δεδομένων (transformation)

## Κανονικοποίηση (Normalization)

Σε περίπτωση που οι τιμές για τις οποίες καλούμαστε να επεξεργαστούμε εμφανίζουν μεγάλη διακύμανση τότε υπάρχει η δυνατότητα της κανονικοποίησης σε συγκεκριμένο εύρος τιμών. Συνήθως η κλίμακά μας είναι από 0 έως 1.

ΤΕΛΟΣ ΕΝΟΤΗΤΑΣ

