**ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ - Ιωάννης Γ. Τσούλος**

**C:\Users\Evita\Documents\Thesis\Stoixeia Ypologistikhs Noumosunhs - Metaptyxiako (Tsoulos)\1. Χαρακτηριστικά και Πρότυπα\lecture1\_data.pdf**

**Χαρακτηριστικό** ορίζουμε κάθε τιμή που αναπαριστά μία ιδιότητα

Τα χαρακτηριστικά μπορούν να είναι

1. Συνεχείς τιμές
2. Διακριτές τιμές
3. Αλφαριθμητικές τιμές

Παραδείγματα χαρακτηριστικών μπορούν να είναι:

1. Η θερμοκρασία από έναν αισθητήρα (συνεχής τιμή)
2. Η ηλικία ενός ανθρώπου (διακριτή τιμή)
3. Η πιστοληπτική ικανότητα ενός δανειολήπτη (αλφαριρθμητική τιμή)

[ ΜΕΤΑΤΡΟΠΕΣ ΤΙΜΩΝ ]

1.

Οι αλφαριθμητικές τιμές αν είναι πεπερασμένες σε πλήθος

μετατρέπονται σε διακριτές πχ. με απαρίθμηση.

2.

Οι συνεχείς τιμές μπορούν να μετατραπούν σε διακριτές

τιμές με χρήση ορίων.

3.

Συνήθως οι διακριτές τιμές είναι κατάλληλες για δένδρα

απόφασης, ενώ οι συνεχείς είναι περισσότερο κατάλληλες

σε τεχνητά νευρωνικά δίκτυα.

ΠΑΡΑΔΕΙΓΜΑΤΑ

Μετατροπή χρώματος φρούτων:

αρχικές τιμές: ΚΙΤΡΙΝΟ, ΠΡΑΣΙΝΟ, ΚΟΚΚΙΝΟ

Διακριτές τιμές: 0, 1, 2

Συνήθως δεν υπάρχει κάποιο θέμα με τις αριθμητικές τιμές

που επιλέγονται

Συνεχείς τιμές θερμοκρασvιών πχ 16.7

Αρχικές τιμές: Συνεχείς τιμές σvτο διάσvτημα [-20,40]

Δημιουργία 6 ομάδων:

[-20,-10],(-10,0],(0,10],(10,20],(20,30],(30,40].

Ανάθεσvη σvε κάθε ομάδα μιας διακριτής τιμής: [0,1,2,3,4,5]

Για παράδειγμα η θερμοκρασvία 16.7 είναι σvτην τέταρτη

ομάδα και έτσvι παίρνει την τιμή 3.

Είναι κρίσvιμο το εύρος του διασvτήματος και σvε πολλές

περιπτώσvεις απαιτείται και η σvυμβουλή ενός ειδικού σvτο

πεδίο για τον καθορισvμό του

**Πρότυπα** είναι τα σύνολα χαρακτηριστικών. Κάθε πρότυπο είναι μία ξεχωριστή καταγραφή. Δεν είναι υποχρεωτικό όλα τα χαρακτηριστικά να είναι αποκλειστικά συνεχή ή αποκλειστικά διακριτά.

Το σύνολο προτύπων ονομάζεται dataset.

Συνήθως μαζί με κάθε πρότυπο υπάρχει και ένας

χαρακτηρισvμός όπως για παράδειγμα η ποιότητα ενός

μπουκαλιού κρασvιού.

Η **έλλειψη τιμών** σε ορισμένα χαρακτηριστικά Προκύπτει από λαθός καταχωρήσvεις πολλές φορές Μπορεί να προκύψει από δεδομένα σvτα οποία έχουν γίνει

κατα λάθος διαγραφές. Πολλές φορές προκαλείται από ασvτοχία υλικού σvε περίπτωσvη αισvθητήτων για παράδειγμα

**[ΤΡΟΠΟΙ ΕΠΙΛΥΣΗΣ ΧΑΜΕΝΩΝ ΤΙΜΩΝ]**

1.

1 Διαγραφή ολόκληρης της γραμμής. Μπορεί να μειώσvει

αρκετά τις εγγραφές και δεν χρησvιμοποιείτα σvυχνά.

2.

2 Αναζήτησvη της πραγματικής τιμής. Αυτό μπορεί να γίνει

από τον ειδικό που έφτιαξε το σvύνολο δεδομένων.

3.

3 Χρήσvη σvταθεράς σvτις χαμένες τιμές. Αντικατάσvτασvη

χαμένων τιμών με κάποια σvταθερά πχ 0.0 αλλά μπορεί να

προκαλέσvει θόρυβο σvτα δεδομένα.

4.

4 Αντικατάσvτασvη με τον μέσvο όρο. Αντικαθίσvτανται οι

χαμένες τιμές με τον μέσvο όρο της σvτήλης. Είναι η πιο

κοινή μέθοδος