
ΚΕΦΑΛΑΙΟ 20

Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων

Ασκήσεις - Ερωτήσεις

1. Ποιες είναι οι κυριότερες ομοιότητες και διαφορές μεταξύ μηχανικής μάθησης και ανακάλυψης γνώσης σε βάσεις δεδομένων;
2. Τα συστήματα OLAP (online analytical processing) θεωρούνται συστήματα ανακάλυψης γνώσης σε βάσεις δεδομένων;
3. Ποιες τεχνικές μηχανικής μάθησης χρησιμοποιούνται συνήθως στην ανακάλυψη γνώσης σε βάσεις δεδομένων;
4. Ποιοι είναι οι κύριοι στόχοι ενός συστήματος ανακάλυψης γνώσης;
5. Που συνεισφέρουν τα συστήματα αποθήκευσης δεδομένων (data warehouse) στη διαδικασία ανακάλυψης γνώσης σε βάσεις δεδομένων;
6. Ποιες είναι οι κύριες τεχνικές μείωσης διαστάσεων των δεδομένων και ποια είναι η κύρια διαφορά τους;
7. Με ποιον τρόπο η ύπαρξη περιττών χαρακτηριστικών επηρεάζει τη διαδικασία ανακάλυψης γνώσης;
8. Τι είναι και πότε εφαρμόζεται η διαδικασία της διακριτοποίησης;
9. Ποια μέθοδο διακριτοποίησης θεωρείτε ότι είναι καλύτερη;

10. Τι τύπους δεδομένων πρέπει να υποστηρίζει μια βάση δεδομένων ώστε να γίνεται αποδοτική ανακάλυψη γνώσης;
11. Σε ποιες περιπτώσεις εφαρμόζεται η κατανομημένη εξόρυξη σε δεδομένα;
12. Τι δυσκολίες αντιμετωπίζονται στην ανακάλυψη γνώσης από ροές δεδομένων;
13. Να κανονικοποιήσετε τις τιμές {20, 40, 60, 80, 120} χρησιμοποιώντας:
- Κανονικοποίηση ελαχίστου-μεγίστου, στο διάστημα [0, 1]
 - Κανονικοποίηση z-score
 - Κανονικοποίηση δεκαδικής κλιμάκωσης
14. Έστω το παρακάτω σύνολο δεδομένων. Διακριτοποιήστε τη μεταβλητή Πληθωρισμός σε δύο τιμές (χαμηλός, υψηλός) με τη μέθοδο της εντροπίας. Η εξαρτημένη μεταβλητή είναι η Ανεργία. Δεν χρειάζεται να γίνουν πράξεις.

Δημόσια Έσοδα	Πληθωρισμός (%)	Ανεργία
Μικρά	2	Σταθερή
Μεσαία	3	Πτωτική
Μεγάλα	4	Πτωτική

15. Ο ακόλουθος πίνακας περιέχει 14 πειραματικά δεδομένα και τις αντίστοιχες τιμές της εξαρτημένης μεταβλητής.

64	65	68	69	70	71	72	75	80	81	83	85	88	90
yes	no	yes	yes	yes	no	yes	yes	no	yes	yes	no	yes	yes

Να διακριτοποιήσετε τα δεδομένα:

- με μια απλή μέθοδο
- με τη μέθοδο διαχωρισμού ίσων διαστημάτων (επιλέξτε τον κατάλληλο αριθμό)
- με τη μέθοδο διαχωρισμού ίσης συχνότητας
- με τη μέθοδο της εντροπίας