****

Πολυτεχνική Σχολή

Τμήμα Μηχανικών Η/Υ & Πληροφορικής

**ΘΕΩΡΊΑ ΑΠΟΦΆΣΕΩΝ**

PROJECT 2022-23

ΜΈΡΟΣ Α’

ΝΑΣΙΈΛΗ ΒΑΣΙΛΙΚΉ

Α.Μ. 1090034

*Προεπεξεργασία δεδομένων:*

*Να αναφέρετε, πόσα είναι τα χαρακτηριστικά κάθε δείγματος και πόσα δείγματα εκπαίδευσης περιέχει το αρχείο.*

Το σύνολο δεδομένων ILPD περιέχει 11 στήλες και 583 γραμμές/δείγματα (Ξεκινώντας από το 0). Συγκεκριμένα, συγκρατεί με σειρά στις στήλες τα εξής δεδομένα (σύμφωνα με την επίσημη ιστοσελίδα):

1. Age of the patient

2. Gender of the patient

3. Total Bilirubin

4. Direct Bilirubin

5. Alkaline Phosphotase

6. Alamine Aminotransferase

7. Aspartate Aminotransferase

8. Total Protiens

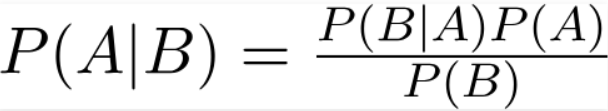
9. Albumin

10. A/G Ratio Albumin and Globulin Ratio

11. Selector field used to split the data into two sets

***Ερώτημα 2.*** *Στο μάθημα συζητήθηκε εκτεταμένα ο ταξινομητής Bayes. Στη βιβλιογραφία, υπάρχει μια παραλλαγή του που λέγεται Αφελής Ταξινομητής Bayes (Naïve Bayes), με την υπόθεση ότι τα χαρακτηριστικά είναι στατιστικά ανεξάρτητα. Αναζητήστε τη σχετική βιβλιογραφία στο Internet, και να κάνετε μια σύντομη παρουσίαση του αλγορίθμου. Στη συνέχεια να κάνετε μια σύγκριση με τον Ταξινομητή Bayes.*

Ο ταξινομητής Naive Bayes είναι ένα πιθανολογικό μοντέλο μηχανικής μάθησης που ταξινομεί στοιχεία σε κλάσεις ,εκμεταλλευόμενος την θεωρία πιθανοτήτων και το θεώρημα Bayes, το οποίο περιγράφει την πιθανότητα ενός χαρακτηριστικού, βασισμένη σε προηγούμενη γνώση προϋποθέσεων / άλλων χαρακτηριστικών.



Χρησιμοποιώντας ως βάση το θεώρημα Bayes, μπορούμε να βρούμε την πιθανότητα να συμβεί το γεγονός Α, δεδομένου ότι έχει ήδη συμβεί το Β. Εδώ, το Β είναι η απόδειξη και το Α είναι η υπόθεση. Η υπόθεση που γίνεται είναι ότι τα χαρακτηριστικά είναι ανεξάρτητα. Δηλαδή η παρουσία (ή απουσία) ενός συγκεκριμένου χαρακτηριστικού κλάσης δεν επηρεάζει την παρουσία (ή απουσία) κάποιου άλλου χαρακτηριστικού της ίδιας κλάσης. Σε αυτή την επεξήγηση οφείλεται και η υιοθεσία του ορισμού ‘αφελής’.

‘Bayesian analysis allows us to answer questions for which frequentist statistical approaches were not developed. In fact the very idea of assigning a probability to a hypothesis is not part of the frequentist paradigm.’[[1]](#footnote-1)

***Ερώτημα 3.*** *Με χρήση της μεθόδου 5-fold cross validation, να εκπαιδεύσετε τον Naïve Bayes ταξινομητή, να παρουσιάστε και να σχολιάσετε την απόδοσή του. Μπορείτε να χρησιμοποιήσετε κατάλληλες συναρτήσεις του matlab ή οποιαδήποτε εφαρμογή επιθυμείτε (ή να υλοποιήσετε δικό* σας κώδικα). Για την αξιολόγηση της απόδοσης του ταξινομητή να χρησιμοποιήσετε τις μετρικές του ερωτήματος 4, παρακάτω.

1. Berrar, Daniel. “Bayes’ Theorem and Naive Bayes Classifier,” 2018.

   https://doi.org/10.1016/B978-0-12-809633-8.20473-1. [↑](#footnote-ref-1)