

Κεχρής Κωνσταντίνος 3150071

Κουλός Νικόλαος 3150079

3 Εργασία:

Υλοποιήσαμε τον αλγόριθμο του αφελή ταξινομητή Bayes για πολυμεταβλητές Bernoulli (Κάθε αντικείμενο στο dataset παίρνει διακριτές τιμές). Χρησιμοποιήσαμε το car evaluation dataset το οποίο το βρήκαμε στο UCI Machine Learning Repository μέσω του οποίου γίνονται εκτιμήσεις για αυτοκίνητα. Το dataset περιέχει 1728 παραδείγματα προς εκπαίδευση και το κάθε αυτοκίνητο έχει 6 attributes.

- 1) buying: vhigh, high, med, low.
- 2) maint: vhigh, high, med, low.
- 3) Doors: 2, 3, 4, 5more.
- 4) Persons: 2, 4, more.
- 5) lug_boot: small, med, big.
- 6) safety: low, med, high.

Κάθε αυτοκίνητο μπορεί να καταταχθεί στις εξής τέσσερις κατηγορίες: unacc, acc, good και vgood.

Αρχικά διαβάζουμε το αρχείο train.txt το οποίο περιέχει το dataset. Αποθηκεύουμε τα attributes σε πίνακες και ξεκινάμε τον υπολογισμό. Πρέπει να υπολογίσουμε 4 γινόμενα και να τα συγκρίνουμε. Για τον λόγο αυτό υπολογίζουμε πόσες φορές εμφανίζεται η κάθε κατηγορία και αποθηκεύουμε το αποτέλεσμα στις αντίστοιχες μεταβλητές PC1, PC2, PC3, PC4.

Κατόπιν εκτελούμε μια for loop για να υπολογίσουμε τα τέσσερα ζητούμενα γινόμενα. Το κάθε γινόμενο ισούται με: $Rate = Rate * (C_i / PC_i + k)$. Το Rate είναι το ζητούμενο ποσοστό τα C_i είναι μετρητές (πόσες φορές υπάρχει το χαρακτηριστικό δεδομένου της ζητούμενης κατηγορίας) Τα C_i ξεκινάνε από 1 λόγω της ψευτοεμφάνισης. Στον αριθμητή έχουμε PC_i (πόσες φορές υπάρχει η ζητούμενη κατηγορία) και το k ποικίλλει ανάλογα με το κάθε χαρακτηριστικό πχ για την κατηγορία buying θα ναι $k = 4$ διότι το χαρακτηριστικό Buying μπορεί να πάρει 4 τιμές.

ΥΓ: Υλοποιήσαμε μόνο έναν από τους προβλεπόμενους 2 αλγόριθμους λόγω μεγάλου φόρτου εργασίας από άλλα μαθήματα.