Project Υπολογιστική Νοημοσύνη Β Μέρος

Β1. α) Κωδικοποίηση για τα άτομα του πληθυσμού:

Για να κωδικοποιήσουμε τα άτομα του πληθυσμού θα χρησιμοποιήσουμε δυαδική κωδικοποίηση.

Πρώτα, θα πραγματοποιήσουμε Standardization στα δεδομένα, προκειμένου να εξαλείψουμε τα outliers και έπειτα θα πραγματοποιήσουμε Normalization στα δεδομένα, ώστε να τα φέρουμε στο εύρος [0,1]. Με αυτό τον τρόπο, έχουμε αφαιρέσει κατά ένα μεγάλο βαθμό τα outliers που μπορεί να υπήρχαν στα δεδομένα μας, τα οποία θα επηρέαζαν μετρήσεις, όπως ο μέσος όρος ανά κλάση (sitting, walking, etc), τον οποίο θα χρειαστούμε για να αξιολογήσουμε τα άτομα του πληθυσμού, αλλά μειώσαμε και τον αριθμό των bit που θα χρειαστούν για να κωδικοποιήσουμε κάθε τις τιμές των x,y,z των 4 αισθητήρων.

Υποθέτουμε επιθυμητή ακρίβεια λύσεων: 4 δεκαδικά ψηφία. Άρα, σε συνδυασμό με τα παραπάνω, έχουμε:

Διάστημα τιμών των x, y, z είναι το [0,1], με μήκος 1. Το οποίο θα πρέπει να διαχωριστεί σε τουλάχιστον 1\*104=10000 ίσα υπο-διαστήματα για να επιτευχθεί η επιθυμητή ακρίβεια των τεσσάρων δεκαδικών ψηφίων. Άρα θα χρειαστούμε 14 δυαδικά ψηφία για καθένα από τα x, y, z των τεσσάρων αισθητήρων.

β) Με 14 bit μπορούμε να αναπαραστήσουμε 214=16384 διαφορετικούς αριθμούς, ενώ εμείς θέλουμε 10000, αφού τόσα είναι και τα υπο-διαστήματα στα οποία θα χωρίσουμε το διάστημα [0, 1]. Εάν κατά τη διασταύρωση ή τη μετάλλαξη προκύψει για κάποια μεταβλητή (δηλαδή 14αδα bit) ενός χρωμοσώματος τιμή η οποία είναι στο εύρος μεταξύ [10001, 16384], μπορούμε να αντιστοιχίσουμε, τυχαία και με ίση πιθανότητα, την πλεονάζουσα τιμή σε κάποια από τις νόμιμες τιμές. Με αυτό τον τρόπο διατηρούμε την τυχαία φύση του γενετικού αλγορίθμου αλλά μπορεί να “χάσει” πληροφορία με το να επιλέξει τιμή λιγότερο κατάλληλη από αυτή που είχε πριν.

Προκειμένου να αποφύγουμε τις πλεονάζουσες τιμές θα μπορούσαμε να χωρίσουμε το διάστημα [0, 1] σε 16384 υπο-διαστήματα, εύρους 0.6103515625

γ) Για την δημιουργία αρχικού πληθυσμού διαλέγουμε τυχαία και ισοπίθανα τιμές από το διάστημα [0, 1], τις κωδικοποιούμε με τον τρόπο που αναφέραμε στο α) ερώτημα και ανά 12 τις ενώνουμε ώστε να δημιουργήσουμε ένα χρωμόσωμα.

(Τυχαίες τιμές στο διάστημα [0,1] ή βλέπω σε ποιες τιμές κινείται ο κάθε αισθητήρας και βγάζω σε αυτά τα διαστήματα τυχαίες τιμες;]

δ) Καθώς έχουμε κωδικοποιήσει τις τιμές των αισθητήρων με δυαδική κωδικοποίηση, Οι μετρικές Ευκλείδειας απόστασης, απόστασης Manhattan και συσχέτισης Pearson δεν είναι κατάλληλες για την αξιολόγηση της απόστασης μεταξύ ενός χρωμοσώματος και του μέσου όρου των 5 κλάσεων. Καθώς έχουμε πολύ μεγάλα μητρώα, η Ευκλείδεια απόσταση θα αξιολογεί ως όμοια άτομα τα οποία μοιράζονται πολλές 0 τιμές, χωρίς όμως αυτό να τα κάνει απαραίτητα όμοια. Η απόσταση Manhattan, θα επιστρέφει τον αριθμό των διαφορετικών bit μεταξύ δύο μητρώων, κάτι που δεν είναι επιθυμητό, καθώς μπορεί να τύχει μεταξύ δύο διαφορετικών στάσεων του σώματος, ο αριθμός των bit που διαφέρουν να είναι πολύ κοντά μεταξύ τους, και να μην μπορούμε έτσι να ξεχωρίσουμε σε ποια κλάση είναι πιο κοντά. Η συσχέτιση Pearson μετρά την γραμμική συσχέτιση μεταξύ μεταβλητών συνεχόμενων τιμών, υποθέτοντας ότι αυτές ακολουθούν κανονική κατανομή και καθώς τα δυαδικά δεδομένα δεν ακολουθούν κάποιο γραμμικό μοτίβο, τα αποτελέσματα θα ήταν ανακριβή.

Με τη χρήση ομοιότητας συνημιτόνου μπορούμε να βρούμε την ομοιότητα μεταξύ δύο μητρώων στο χώρο, που είναι η μεταξύ γωνία συνημιτόνου των δύο μητρώων. Η τιμή της είναι μεταξύ [-1,1]. Για τιμή 1, τα δύο μητρώα είναι ευθέως ανάλογα, δηλαδή όμοια. Για τιμή 0, τα δύο μητρώα είναι κάθετα, δηλαδή δεν έχουν ομοιότητα και για τιμή -1 είναι ακριβώς αντίθετα. Με αυτό τον τρόπο μπορούμε να δούμε εύκολα, κατά πόσο ένα χρωμόσωμα είναι κοντά ή όχι στην επιθυμητή κλάση (sitting) και να ενεργήσουμε ανάλογα.