

ΕΠΛ 442: Μηχανική Μάθηση

Radial Basis Function Network

Deadline: 04/12/2024

Υλοποιήστε ένα Νευρωνικό Δίκτυο τύπου Συνάρτησης Αξονικής Βάσης (Radial Basis Function, RBF) χρησιμοποιώντας Γκαουσιανές συναρτήσεις βάσης στους κρυφούς νευρώνες. Ακολουθώντας χρησιμοποιήστε το κώδικα σας για να κατασκευάσετε και να εκπαιδεύσετε ένα RBF δίκτυο το οποίο θα χρησιμοποιηθεί για την πρόβλεψη πληροφοριών για το πιο κάτω πρόβλημα. Σημειώστε ότι το πρόβλημα είναι πρόβλημα παλινδρόμησης (regression) και όχι κατηγοριοποίησης. Τα δεδομένα για την εκπαίδευση και έλεγχο του δικτύου περιγράφονται στη παράγραφο «Δεδομένα» πιο κάτω. Ο αριθμός των κέντρων, ο αριθμός των επαναλήψεων, οι ρυθμοί μάθησης καθώς και οι κανονικές αποκλίσεις (πλάτη γκαουσιανών) θα είναι μεταβλητές που θα καθορίζει ο χρήστης.

Είσοδος Προγράμματος

Παράμετροι δομής και εκπαίδευσης δικτύου: Το όνομα του αρχείου πρέπει να είναι parameters.txt και να περιλαμβάνει τις απαραίτητες μεταβλητές:

numHiddenLayerNeurons
numInputNeurons
numOutputNeurons
learningRates
sigmas
maxIterations
centresFile
trainFile
testFile

Έξοδος Προγράμματος

1. Αρχείο (results.txt) για συγκεκριμένη εκπαίδευση του RBF με τρεις στήλες. Η πρώτη θα είναι ο μετρητής επαναλήψεων, η δεύτερη το λάθος εκμάθησης (training error) για κάθε επανάληψη και η τρίτη το λάθος ελέγχου (testing error) για κάθε επανάληψη.
2. Αρχείο (weights.txt) που θα περιέχει όλα τα βάρη των ακμών του νευρωνικού σας δικτύου μετά το τέλος της εκπαίδευσης.

Παραδοτέα:

1. Κώδικας με σχόλια, training.txt, test.txt, parameters.txt, centerVectors.txt
2. Γραπτό κείμενο μέχρι 6 σελίδες.

Σημειώσεις:

1. Η ακριβής μορφή των δικτύων, οι παράμετροι εισόδου όπως και διάφορες λεπτομέρειες της υλοποίησης του κώδικα είναι στη δική σας δικαιοδοσία.
2. Χρησιμοποιήστε την παραλλαγή του RBF αλγορίθμου με μεταβλητά κέντρα.
3. Για την κατασκευή του training.txt και test.txt διαχωρίστε τυχαία τα δεδομένα σας έτσι ώστε το test.txt να περιέχει 10 γραμμές και το training.txt τις υπόλοιπες 21.
4. Τα αρχικά κέντρα που θα χρησιμοποιηθούν από τα νευρώνια του κρυφού επιπέδου θα δίδονται στο RBF δίκτυο μέσω του αρχείου centerVectors.txt.

Δεδομένα:

Selwood data set:

Το παρεχόμενο αρχείο περιέχει δεδομένα για 31 μόρια που έχουν ελεγχθεί εργαστηριακά για συγκεκριμένη βιολογική δράση. Η πρώτη γραμμή του αρχείου περιέχει το όνομα του και ακολουθείται από μια κενή γραμμή. Η επόμενη γραμμή περιέχει το όνομα/περιγραφή των δεδομένων κάθε στήλης. Κάθε μια από τις επόμενες 31 γραμμές αντιστοιχεί σε ένα μόριο. Η πρώτη στήλη περιέχει τον κωδικό/όνομα του μορίου, η δεύτερη τη βιολογική δράση (Y) όπως έχει μετρηθεί πειραματικά και οι επόμενες 53 στήλες το διάνυσμα που περιγράφει το μόριο. Το διάνυσμα απαρτίζεται από μοριακά χαρακτηριστικά που έχουν υπολογιστεί κυρίως από τη μοριακή δομή. Το εύρος των τιμών κάθε χαρακτηριστικού είναι διαφορετικό γι αυτό και επιβάλλεται η κανονικοποίηση των δεδομένων πριν τη χρήση. Οι στήλες διαχωρίζονται με κόμμα. Όπου η βιολογική δράση παρουσιάζεται ως < -1 θεωρήστε ότι είναι -1.

Τα αρχεία (δεδομένα, περιγραφή και δημοσίευση) βρίσκονται στο moodle.