***ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ***

ΠΡΟΤΖΕΚΤ 2

* + ***ΑΛΓΟΡΙΘΜΟΣ HO-KASHYAP***

1. ***Μη γραµµικά διαχωρίσιµα πρότυπα***

Ο αλγόριθµος Ho-Kasyhap λύνει αποτελεσµατικά το πρόβληµα τις εκπαίδευσης των συναρτήσεων απόφασης στην περίπτωση των µη γραµµικά διαχωρίσιµων προτύπων. Ο αλγόριθµος ελαχιστοποίησης του σφάλµατος εκτίµησης είναι ένας επαναληπτικός αλγόριθµος ο οποίος συγκλίνει ακόµα και στις περιπτώσεις στις οποίες τα πρότυπα των παραδειγµάτων δεν είναι γραµµικά διαχωρισµένα.

Ας δούµε όµως πως προκύπτει ο αλγόριθµος. Οπως έχει προαναφερθεί µελετάµε το πρόβληµα ορισµού της συνάρτησης απόφασης δύο κατηγοριών µία και ο υπολογισµός των συντελεστών εξαρτάται µονάχα από τα παραδείγµατα δύο κατηγοριών κάθε φορά. Εστω g12(***x***) = -g21(***x***) και N1, N2 είναι ο αριθµός των παραδειγµάτων για τις δύο κατηγορίες αντίστοιχα.

Ορίζω σαν ***x*** την επέκταση του διανύσµατος κάθε πρότυπου που προκύπτει µε την προσθήκη µιάς νέας διάστασης µε τιµή 1, ***x*** = (***x***', 1) όπου ***x***' είναι το διάνυσµα του πρότυπου. Επίσης για κάθε ένα από τα διανύσµατα εκπαίδευσης της κατηγορίας ω2 αντιστρέφω το πρόσηµο των συντελεστών ***x*** = -***x***. Η µετατροπή αυτή µετασχηµατίζει το πρόβληµα εκπαίδευσης του συστήµατος στην εύρεση των συντελεστών ***w*** οι οποίοι να ικανοποιούν τις ακόλουθες ανισότητες:

***x***i ***w*** > 0, i = 1,...,M, M = N1+ N2

Ολες τις συναρτήσεις µπορούµε να τις οµαδοποιήσουµε σε:

***X w*** > ***0***, ***X*** = ( ***x***1, ***x***2,..., ***x***M)

Το πρόβληµα επίλυσης της ανίσωσης µπορεί να αντικατασταθεί από ένα πρόβληµα επίλυσης της εξίσωσης: ***X w*** = ***b***, ***b*** διάνυσµα θετικών αριθµών.

Αναλυτική λύση της εξίσωσης δεν µπορεί να επιτευχθεί, γιαυτό τον λόγο προσπαθούµε να επιτύχουµε µία επαναληπτική λύση η οποία να βελτιστοποιεί διαδοχικά µία συνάρτηση κριτηρίου. Η συνάρτηση αυτή θα µας δείχνει ταυτόχρονα πόσο έχουµε πλησιάσει την

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| επιθυµούµενη λύση του προβλήµατος | | | σε κάθε επανάληψη του αλγόριθµου. Η συνάρτηση | | |  |
| σφάλµατος δίνεται από την εξίσωση: | | |  |  |  |  |
| *Er*( *X* , *w*,*b*)= | 1 | ∑(*wT* *xi* −*bi* )2 = | | 1 | || *Xw* −*b* ||2 |  |
|  | 2 |  |
| 2 | | *i* |  |  |  |

Η βέλτιστη τιµή του ***w*** προκύπτει από την ελαχιστοποίηση της συνάρτησης του σφάλµατος.

Επειδή µπορούµε να ελαχιστοποιήσουµε την συνάρτηση µεταβάλλοντας δύο ανεξάρτητες µεταβλητές τις ***w***, και ***b*** υπολογίζουµε την παράγωγο της συνάρτησης σφάλµατος ως προς αυτές τις µεταβλητές.

∂*E*( *X* , *w*,*b*) = *X* *T* ( *Xw* −*b*)

∂*w*

Με γνωστή την τιµή ***b***, η ελάχιστη τιµή της συνάρτησης βρίσκεται στο σηµείο στο οποίο ισχύει:

∂*E*( *X* , *w*,*b*) = −( *Xw* −*b*)

∂*b*

∂*E*( *X* , *w*,*b*)=0⇒ *X T* ( *Xw* −*b*)=0⇒ *w* =( *X T X* )−1 *X T b*

∂*w*

Η έκφραση ( ***X***T ***X*** )-1 ***X***T ονοµάζεται γενικευµένος αντίστροφος πίνακας του ***X***. Ο συντελεστής ***b*** επαναπροσδιορίζεται βάσει της µεθόδου βηµατικής σύγκλισης:

*b* = *b* −*c* ∂*E*( *X* , *w*,*b*)

∂*b*

Για να έχουµε αποδεκτή λύση δεν πραγµατοποιούµε την απλή αντικατάσταση της παραγώγου που έχουµε υπολογίσει διότι γνωρίζουµε ότι οι συνιστώσες του διανύσµατος ***b*** πρέπει να είναι πάντα θετικοί αριθµοί. Με την αναδροµική σχέση που δίνουµε στην επόµενη εξίσωση επαναπροσδιορίζουµε τους συντελεστές του διανύσ µατος ***b*** µόνο στην περίπτωση κατά την οποία αυτοί θα αυξηθούν αριθµητικά σε κάποιες από τις συνιστώσες τους.

***b*** = ***b*** + c ( ***X w*** - ***b*** + | ***X w*** - ***b***|)

Με την τεχνική αυτή προσεγγίζουµε την επίλυση της ανίσωσης που επιθυµούµε, ικανοποιώντας ταυτόχρονα και τους περιορισµούς που θέσαµε.

Με βάση τα παραπάνω παραθέτουµε τον αλγόριθµο υπολογισµού των συναρτήσεων απόφασης από παραδείγµατα που βασίζεται στην ελαχιστοποίηση του σφάλµατος εκτίµησης. Συνήθης επιλογή του συντελεστή c είναι κάποιος µικρός θετικός πραγµατικός αριθµός 0 < c < 1. Για αυτές τις τιµές ο αλγόριθµος, που είναι γνωστός και µε το όνοµα Ho-Kashyap, συγκλίνει ακόµα και αν τα διανύσµατα δεν είναι γραµµικά διαχωρίσιµα.

***Αρχικές τιµές***.Τοποθετούµε τυχαίουςµικρούς θετικούς αριθµούς στους συντελεστές τουδιανύσµατος ***b***.

***Υπολογισµός των συντελεστών της συνάρτησης απόφασης***.Με την σχέση πουακολουθεί πραγµατοποιούµε µία εκτίµηση του ***w***: ***w*** = ( ***X***T ***X*** )-1 ***X***T ***b.***

***Ελεγχος σύγκλισης*** .Αν ***X w*** > ***0*** τότε τα παραδείγµατα είναι γραµµικά διαχωρίσιµα οπότεκαι ο αλγόριθ µος τερ µατίζει. Αν η παραπάνω συνθήκη δεν ικανοποιείται και οι µεταβολές του ***w*** κατά δυο διαδοχικές προσεγγίσεις έχουν µειωθεί σηµαντικά τότε έχουµε κατάσταση σύγκλισης του αλγόριθµου. Σε αυτή την περίπτωση γνωρίζουµε ότι τα παραδείγµατα δεν είναι γραµµικά διαχωρίσιµα.

***Επαναπροσδιορισµός του b.*** Η νέα τιµή του διανύσµατος ***b*** υπολογίζεται από τηνακόλουθη αναδροµική σχέση: ***b*** = ***b*** + c ( ***X w*** - ***b*** + | ***X w*** - ***b*** |).

***Επανάληψη του αλγόριθµου***.Το επαναληπτικό τµήµα του αλγόριθµου εκτελείται από τοδεύτερο βήµα.

ΕΚΤΕΛΕΣΗ ΤΗΣ ΑΣΚΗΣΗΣ

***1. ∆ιάβασµα προτύπων από ιατρικών εργαστηριακές εξετάσεις για την διάγνωση ζακχαρώδη διαβήτη σε γυναίκες***

Στο αρχείο "\\ee-ntserver\PatRec\pima-indians-diabetes.data"

υπάρχουν σε φόρ µα "ASCII" δεδοµένα που προέρχονται από εξετάσεις αίµατος γυναικών που έχουν εξετασθεί για πιθανή ασθένεια ζαχαρώδη διαβήτη. Οι µετρήσεις αυτές έχουν πραγµατοποιηθεί σε 768 γυναίκες Ινδιάνους στις ΗΠΑ. Σε κάθε γραµµή του αρχείου υπάρχουν πληροφορίες που συσχετίζονται µε κάθε εξεταζόµενη γυναίκα και περιέχει πληροφορίες που συσχετίζονται µε τα συµπτώµατα που ανιχνεύονται µε φυσικές και εργαστηριακές εξετάσεις.

Οι τρείς πρώτες γραµµές του αρχείου είναι οι ακόλουθες:

6 148 72 35 0 33.6 0.627 50 1

1 85 66 29 0 26.6 0.351 31 0

8 183 64 0 0 23.3 0.672 32 1

Οι αριθµοί σε κάθε γραµµή παριστούν τα ακόλουθα στοιχεία:

|  |  |
| --- | --- |
| ***Πεδίο*** | ***Περιεχόµενο του πεδίου*** |
| 1 | Αριθµός κυήσεων |
| 2 | ∆οκιµασία φόρτισης του πλάσµατος στο αίµα µε γλυκόζη |
| 3 | ∆ιαστολική πίεση του αίµατος |
| 4 | Πάχος πτυχών του δέρµατος στην περιοχή του τρικεφάλου |
| 5 | Συγκέντρωςη ινσουλίνης στο αίµα |
| 6 | Μάζα σώµατος |
| 7 | Συνάρτηση κληρονοµικότητας |
| 8 | Ηλικία |
| 9 | Ταξινόµηση: 0 Υγιής, 1 διαβητικός |
|  |  |

Τοποθετείστε τα παραδείγµατα του αρχείου που αντιστοιχούν σε υγιείς στον πίνακα X και τα παραδείγµατα που αντιστοιχούν σε ασθενείς στον πίνακα C.

***2. Υπολογίστε τους συντελεστές των γραµµικών συναρτήσεων απόφασης µε***

***τον αλγόριθµο Ho-Kashyap.***

Για να κατασκευάσετε ένα αυτόµατο διαγνωστικό σύστηµα για αυτόµατη διάγνωση πάθησης διαβήτη χρησιµοποιώντας γραµµικές συναρτήσεις απόφασης πρέπει αρχικά να υπολογίσετε τους συντελεστές των γραµµικών συναρτήσεων. Επιλέξτε έναν συντελεστή εκπαίδευσης και υπολογίστε το διάνυσµα των συντελεστών της γραµµικής συνάρτησης απόφασης µε την βοήθεια του αλγόριθµου Ho-Kashap.

***3. Υπολογίστε το σφάλµα του συστήµατος ταξινόµησης***

Χρησιµοποιώντας όλα τα παραδείγµατα εκπαίδευσης υπολογίστε το ελάχιστο σφάλµα του συστήµατος ταξινόµησης.

***4. Μελετήστε την επίδραση που έχει ο συντελεστής εκπαίδευσης στον αλγόριθµο Ho-kashyap***

Κατασκευάστε την γραφική παράσταση που δίνει την αξιοπιστία του γραµµικού συστήµατος ταξινόµησης για διαφορετικές τιµές του συντελεστή εκπαίδευσης. Κατασκευάστε επίσης και τον αριθµό των επαναλήψεων του αλγόριθµου σαν συνάρτηση του συντελεστή εκπαίδευσης.

***5. Τι παρατηρείτε από τα αποτελέσµατα των µετρήσεων***

Περιγράψτε και αιτιολογείστε τα αποτελέσµατα που πήρατε από τις µετρήσεις του ελάχιστου σφάλµατος.