A picture containing text

Description automatically generated

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

|  |  |
| --- | --- |
| Ακαδημαϊκό έτος: | **2020-2021** |
| Εξεταστική: | **Σεπτέμβριος** |
| Μάθημα: | **Αναγνώριση Προτύπων** |
| Εξάμηνο: | **5ο** |
|  |  |

|  |  |
| --- | --- |
| **Ονοματεπώνυμο** | **ΑΜ** |
|  |  |
|  |  |
|  |  |

Περιεχόμενα

[Εισαγωγή 3](#_Toc82085668)

[Ερώτημα i 4](#_Toc82085669)

[Ανάλυση 4](#_Toc82085670)

[Υλοποίηση 4](#_Toc82085671)

[Εκτέλεση 4](#_Toc82085672)

[Ερώτημα ii 4](#_Toc82085673)

[Ανάλυση 4](#_Toc82085674)

[Υλοποίηση 4](#_Toc82085675)

[Εκτέλεση 4](#_Toc82085676)

[Ερώτημα iii 4](#_Toc82085677)

[Ανάλυση 4](#_Toc82085678)

[Υλοποίηση 4](#_Toc82085679)

[Εκτέλεση 4](#_Toc82085680)

# Εισαγωγή

Με τον όρο «Μηχανική Μάθηση» εννοούμε τη μελέτη αλγορίθμων που μπορούν να αυτό-βελτιώνονται μέσω εμπειρίας και με την αξιοποίηση δεδομένων. Οι αλγόριθμοι αυτοί έχουν σκοπό τη δημιουργία μοντέλων, που θα περιγράφουν ένα σετ δεδομένων, γνωστά ως «δεδομένα εκπαίδευσης», προκειμένου να κάνουν προβλέψεις ή να παίρνουν αποφάσεις χωρίς την συμβολή του ανθρώπου.

Το ζητούμενο στην εργασία είναι ο ορισμός τέτοιων μοντέλων με σκοπό τον καλύτερο διαχωρισμό των δεδομένων και την ομαδοποίηση τους σε κλάσεις. Θέλουμε, επομένως, να δημιουργήσουμε μοντέλα ταξινόμησης που για κάποιες τιμές/βάρη θα διαχωρίζουν με τον καλύτερο δυνατό τρόπο το σύνολο των δεδομένων μας.

Για να το επιτύχουμε αυτό, αναγκαία προϋπόθεση αποτελεί ο ορισμός των βασικών συναρτήσεων βάση των οποίων θα κατασκευάσουμε τα μοντέλα ταξινόμησης. Ένας ταξινομητής εκφράζεται μέσω μίας συνάρτησης πρόβλεψης, που περιγράφει όσο καλύτερα γίνεται το ζητούμενο, και μίας συνάρτησης κόστους/λάθους, που μας δίνει το πόσο απέχει η τιμή πρόβλεψης του ταξινομητή από την πραγματική τιμή. Στόχος, λοιπόν, για κάθε ερώτημα είναι η δημιουργία ενός μοντέλου μηχανικής μάθησης που με τις κατάλληλες τιμές/βάρη στη συνάρτηση πρόβλεψης να ελαχιστοποιείται το κόστος/λάθος για ένα σετ δεδομένων.

Γενικές έννοιες

10-fold-cross-validation

Ρύθμιση βαρών

Δεδομένα

# Ερώτημα i

Να υλοποιήσετε τον Αλγόριθμο Ελάχιστου Μέσου Τετραγωνικού Σφάλματος (**Least Mean Squares**), ώστε ο εκπαιδευμένος ταξινομητής να υλοποιεί την συνάρτηση διάκρισης της μορφήςγια κάθε στοιχηματική εταιρεία. Να αναγνωρίσετε την στοιχηματική εταιρεία τα προγνωστικά της οποίας οδηγούν σε μεγαλύτερη ακρίβεια ταξινόμησης.

## Ανάλυση

Για το ερώτημα αυτό, θέλουμε να κατασκευάσουμε ένα μοντέλο/ταξινομητή που θα χωρίζει το σύνολο των αποδόσεων κάθε στοιχηματικής εταιρίας σε 3 υπερεπίπεδα («Νίκη Εντός», «Ισοπαλία», «Νίκη Εκτός»). Έπειτα, βάση των προβλέψεων του ταξινομητή για κάθε στοιχηματική. Η συνάρτηση πρόβλεψης που θα χρησιμοποιηθεί για κάθε ένα υπερεπίπεδο ορίζεται ως:

y = w0\*x0+w1\*x1+w2\*x2+w3\*x3, όπου

‘wᵢ’: τα βάρη που πρέπει να υπολογιστούν για την ελαχιστοποίηση του σφάλματος του ταξινομητή

‘xᵢ’: οι αποδόσεις κάθε στοιχηματικής. Το ‘x0’=1 αποτελεί ένα σταθερό όρο και χρησιμοποιείται για να συμπεριλαμβάνεται πάντα η τιμή ‘w0’ που αποτελεί το bias του ταξινομητή. Στο ‘x1’ αποδίδεται η απόδοση για τη περίπτωση της «Νίκης Εντός», στο ‘x2’ για τη περίπτωση της «Ισοπαλίας» και στο ‘x3’ για τη περίπτωση της «Νίκης Εκτός».

Επομένως, για κάθε υπερεπίπεδο θα πρέπει να υπολογισθεί το διάνυσμα βαρών που ελαχιστοποιεί το σφάλμα του ταξινομητή. Συγκεκριμένα στο «Ερώτημα i» θέλουμε να ελαχιστοποιήσουμε το μέσο τετραγωνικό σφάλμα ανάμεσα στην επιθυμητή και τη πραγματική έξοδο του ταξινομητή. Άρα η συνάρτηση κόστους ορίζεται ως:

J = Ε [(d(xᵢ) – y(xᵢ)) ^2], όπου

‘d(xᵢ)’: η επιθυμητή έξοδος του ταξινομητή, δηλαδή το πραγματικό αποτέλεσμα του αγώνα i.

‘y(xᵢ)’: η πραγματική έξοδος του ταξινομητή, δηλαδή η πρόβλεψη του ταξινομητή για τον αγώνα i.

ΣΧΕΔΙΟ

Όπως αναφέρθηκε και νωρίτερα, ο τρόπος με τον οποίο θα υπολογίσουμε τα βέλτιστα βάρη για την συνάρτηση h ειναι μέσω της ελαχιστοποίησης της

συνάρτησης κόστους. Πολλές φορές όμως η ελαχιστοποίηση αυτή δεν είναι

τόσο εύκολη. Μια πολύ συχνή τακτική σύγκλισης προς την ελάχιστη τιμή

της συνάρτησης κόστους είναι με την χρήση του αλγορίθμου Gradient

Descent. O συγκεκριμένος αλγόριθμος προσπαθεί να βρεί τα βάρη της h

κατα τα οποία η παράγωγος της συνάρτησης κόστους συγκλίνει προς το

μηδέν. Ωστόσο στην περίπτωση μας η χρήση του παραπάνω αλγορίθμου δεν

είναι απαραίτητη. Παρατηρόντας την συνάρτηση κόστους μας βλέπουμε πως

αποτελείται απο μια τετραγωνική συνάρτηση. Γραφικά αυτό σημαίνει πως

είναι μια παραβολή. Όπως γνωρίζουμε μια παραβολίκη συνάρτηση είναι

κυρτή και έχει μονάχα ενα ολικό ελάχιστο. Αυτο μας δίνει την δυνατότητα να

λύσουμε αναλυτικά το σύστημα όλων των εξίσωσεων για κάθε set τιμών i .

Προφανώς τα βάρη που θα πάρουμε σαν λύση του συστήματος αποτελούν

και τα βέλτιστα. Επειδή το σύστημα το οποίο απαιτείται να λυθεί

περιλαμβάνει πολλές εξισώσεις θα χρησιμοποίησουμε πίνακες και

διανύσματα για να βρόυμε το βέλτιστο δίανυσμα βαρών.

## Υλοποίηση

## Εκτέλεση

# Ερώτημα ii

## Ανάλυση

## Υλοποίηση

## Εκτέλεση

## Ερώτημα iii

## Ανάλυση

## Υλοποίηση

## Εκτέλεση