A picture containing text, sign

Description automatically generated

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΙΔΙΚΕΥΣΗ: “ΠΡΟΗΓΜΕΝΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ”

Wa. Po. – Water Potability

(Analytics as a Service)

ΕΥΦΥΗ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

 Επιβλέπων Καθηγητής: Δημοσθένης Κυριαζής

Ιωάννα Κανδή – Κωνσταντίνος Μαυρογιώργος

ΜΕ2136 – ΜΕ2144

{[ioannakandi, kostismvg}@gmail.com](mailto:ioannakandi@gmail.com)

Ιούνιος 2022

Πίνακας Περιεχομένων

[Κατάλογος Εικόνων 1](#_Toc98767823)

[Κατάλογος Πινάκων 1](#_Toc98767824)

[Εισαγωγή 2](#_Toc98767825)

[Ορισμός 3](#_Toc98767826)

[Μεθοδολογία 4](#_Toc98767827)

[Γραμμική & Λογιστική Παλινδρόμηση 5](#_Toc98767828)

[Πλεονεκτήματα & Μειονεκτήματα της Λογιστικής Παλινδρόμησης 6](#_Toc98767829)

[Συμπεράσματα 7](#_Toc98767830)

[Πηγές 8](#_Toc98767831)

# Κατάλογος Εικόνων

[Εικόνα 1: Σιγμοειδής συνάρτηση 4](#_Toc98767834)

[Εικόνα 2: Συντελεστές παλινδρόμησης 4](#_Toc98767835)

[Εικόνα 3: Απεικόνιση σιγμοειδούς συνάρτησης 4](#_Toc98767836)

[Εικόνα 4 Διαγραμματική απεικόνιση γραμμικής & λογιστικής παλινδρόμησης 5](#_Toc98767837)

# Κατάλογος Πινάκων

[Πίνακας 1:Σύγκριση γραμμικής & λογιστικής παλινδρόμησης 5](#_Toc98767847)

[Πίνακας 2: Πλεονεκτήματα & μειονεκτήματα χρήσης λογιστικής παλινδρόμησης 6](#_Toc98767848)

# Εισαγωγή

Η εκπόνηση εργασιών στα πλαίσια των μαθημάτων της Μεταπτυχιακής Εξειδίκευσης Σπουδών με τίτλο «Προηγμένα Πληροφοριακά Συστήματα» ήταν απαραίτητη για την μέγιστη δυνατή εξοικείωση των φοιτητών με το αντικείμενο του Προγράμματος. Η παρούσα εργασία αφορά το μάθημα εαρινού εξαμήνου με τίτλο «Ευφυή Πληροφοριακά Συστήματα και Τεχνητή Νοημοσύνη», και πραγματεύεται περιληπτικά τη δημιουργία μιας containerized εφαρμογής που αξιοποιεί την βιβλιοθήκη Keras για την πραγματοποίηση πρόβλεψης σχετικά με τη ποσιμότητα δειγμάτων νερού. Στο κείμενο που ακολουθεί περιγράφεται το σύνολο δεδομένων που αξιοποιήθηκε, οι τεχνολογίες που εφαρμόστηκαν και ο τρόπος με τον οποίο λειτουργεί η εφαρμογή . Τέλος, παρατίθενται τα συμπεράσματα που προκύπτουν από την εκπόνηση της εργασίας αυτής.

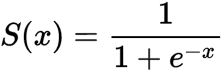
# Ορισμός

Η Λογιστική Παλινδρόμηση (Logistic Regression) αποτελεί έναν αλγόριθμο ταξινόμησης (classification algorithm), ο οποίος επεξεργάζεται -εξαρτημένες- κατηγορικές μεταβλητές, με δύο (2) μόνο τιμές: είτε ένα (1) ή μηδέν (0), είτε ‘ναι’ ή ‘όχι’ (yes/no) , είτε Τ (True) ή F (False) [1]. Οι τύποι της λογιστικής παλινδρόμησης - ανάλογα με τον τύπο της εξαρτημένης μεταβλητής - ανάγονται σε τρεις (3) κατηγορίες:

* Δυαδική (binary): ΝΑΙ/΄ΟΧΙ
* Τακτική (ordinal): π.χ. πάρα πολύ/πολύ/μέτρια/λίγο/πολύ λίγο
* Πολυωνυμική (polynominal)[[1]](#footnote-1): π.χ. το χρώμα ενός αμαξιού. Μπορεί να είναι, μπλε, κόκκινο, κίτρινο, πράσινο κ.ο.κ.

# Μεθοδολογία

Η λογιστική παλινδρόμηση βασίζεται στη σιγμοειδή συνάρτηση (sigmoid function) η οποία ορίζεται από τον τύπο :



Εικόνα 1: Σιγμοειδής συνάρτηση

Όπου x είναι η μεταβλητή εισόδου και το S(x) είναι το αντίστοιχο αποτέλεσμα της μεταβλητής αυτής.

Η μεταβλητή x ορίζεται, επίσης, και ως γραμμικός συνδυασμός του συνόλου των ανεξάρτητων μεταβλητών που συμμετέχουν στο μοντέλο παλινδρόμησης.



Εικόνα 2: Συντελεστές παλινδρόμησης

Οι συντελεστές παλινδρόμησης β0,…,βκ δύνανται να είναι θετικοί ή αρνητικοί. Ανάλογα με το πρόσημο, γίνεται λόγος είτε για αύξηση της πιθανότητας του ενδεχομένου, είτε για μείωση της πιθανότητας εμφάνισης του ενδεχομένου αντίστοιχα [2].

Και είναι της μορφής:

Diagram

Description automatically generated

Εικόνα 3: Απεικόνιση σιγμοειδούς συνάρτησης

# Γραμμική & Λογιστική Παλινδρόμηση

Στη βιβλιογραφία, παρατηρείται σύγκριση μεταξύ της γραμμικής και της λογιστικής παλινδρόμησης. Παρακάτω παρατίθεται ένας πίνακας σύγκρισης της γραμμικής (linear regression) και της λογιστικής παλινδρόμησης (logistic regression) [3].

|  |  |
| --- | --- |
| Γραμμική Παλινδρόμηση | Λογιστική Παλινδρόμηση |
| Διαχείριση ακραίων τιμών (outliers) | Κατάταξη τιμών (classification) |
| Αφορά συνεχείς τιμές (continuous output) | Αφορά ξεχωριστές τιμές (discreet output) |
| Εύρεση βέλτιστου ευθύγραμμου τμήματος | Διαχείριση των αποτελεσμάτων διακριτά σε σιγμοειδή συνάρτηση (Sigmoid curve) |
| Τα αποτελέσματα μπορούν να ξεπεράσουν τις τιμές 0 και 1 | Οι τιμές βρίσκονται εντός των ορίων 0 και 1 |

Πίνακας 1:Σύγκριση γραμμικής & λογιστικής παλινδρόμησης

Chart

Description automatically generated with medium confidence

Εικόνα 4 Διαγραμματική απεικόνιση γραμμικής & λογιστικής παλινδρόμησης

# Πλεονεκτήματα & Μειονεκτήματα της Λογιστικής Παλινδρόμησης

Παρακάτω παρατίθενται τα πλεονεκτήματα και μειονεκτήματα χρήσης της λογιστικής παλινδρόμησης [4].

|  |  |
| --- | --- |
| Πλεονεκτήματα | Μειονεκτήματα |
| Μεγάλη ακρίβεια για μικρά σύνολα δεδομένων | Μη αποδοτικό για δεδομένα με πολλά χαρακτηριστικά |
| Εύκολη υλοποίηση και εφαρμογή | Χρήση μόνο για διακριτές τιμές |
| Εύκολα επεκτάσιμο (update) | Απαραίτητη η προεπεξεργασία δεδομένων |

Πίνακας 2: Πλεονεκτήματα & μειονεκτήματα χρήσης λογιστικής παλινδρόμησης

# Συμπεράσματα

Συνοψίζοντας, η παρούσα εργασία ήταν απαραίτητη για να επιτευχθεί μια πρώτη επαφή με σημαντικές έννοιες του μαθήματος. Πιο συγκεκριμένα, αναλύθηκε η Λογιστική Παλινδρόμηση (Logistic Regression), και πραγματοποιήθηκε σύγκρισης μεταξύ αυτής και της Γραμμικής Παλινδρόμησης (Linear Regression). Τέλος, καταγράφηκαν τα πλεονεκτήματα, καθώς και τα μειονεκτήματα χρήσης της ανωτέρω μεθόδου.

# Πηγές

1. Ahmed A, Jalal A, Kim K. A novel statistical method for scene classification based on multi-object categorization and logistic regression. Sensors. 2020 Jan;20(14):3871.
2. Jiang F, Zhidong GU, Zengshan LI, Xiaodong WA. A method of predicting visual detectability of low-velocity impact damage in composite structures based on logistic regression model. Chinese Journal of Aeronautics. 2021 Jan 1;34(1):296-308.
3. Mourtada J, Gaïffas S. An improper estimator with optimal excess risk in misspecified density estimation and logistic regression. Journal of Machine Learning Research. 2022 Jan 1;23(31):1-49.
4. What is Logistic Regression? , IBM, <https://www.ibm.com/topics/logistic-regression>

1. Οι πολυωνυμική παλινδρόμηση αφορά αποτελέσματα/μεταβλητές οι οποίες δεν έχουν κάποια φυσική συνοχή, σε αντίθεση με την τακτική, στην οποία παρατηρείται η έννοια της κατάταξης/ανισότητας. [↑](#footnote-ref-1)