# ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
Τομέας Επικοινωνιών, Ηλεκτρονικής & Συστημάτων Πληροφορικής
Εργαστήριο Διαχείρισης και Βέλτιστου Σχεδιασμού Δικτύων - NETMODE
Ηρώων Πολυτεχνείου 9, Ζωγράφου, 157 80 Αθήνα, Τηλ: 210-772.2503, Fax: 210-772.1452
e-mail: maglaris@netmode.ntua.gr, URL: http://www.netmode.ntua.gr


## ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ
**(ΔΠΜΣ Επιστήμη Δεδομένων & Μηχανική Μάθηση)**

**ΕΝΔΕΙΚΤΙΚΕΣ ΑΣΚΗΣΕΙΣ 1, ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ 2022**

## Logistic Regression

Η τεχνική Logistic Regression χρησιμοποιείται σε περιπτώσεις δυαδικής ταξινόμησης. Στα πλαίσια του παραδείγματος που παρατίθεται, έχουμε δύο ανεξάρτητες μεταβλητές, τις X1 και X2. Η είσοδος των δύο αυτών μεταβλητών προκύπτει τυχαία από κατανομές Gauss. Η έξοδος παίρνει τιμές 0 και 1. Το training dataset με το οποίο και θα εκπαιδεύσουμε το μοντέλο είναι το εξής:

| *X1* | *X2* | *Y* |
|---|---|---|
| 2.7810836 | 2.550537003 | 0 |
| 1.465489372 | 2.362125076 | 0 |
| 3.396561688 | 4.400293529 | 0 |
| 1.38807019 | 1.850220317 | 0 |
| 3.06407232 | 3.005305973 | 0 |
| 7.627531214 | 2.759262235 | 1 |
| 5.332441248 | 2.088626775 | 1 |
| 6.922596716 | 1.77106367 | 1 |
| 8.675418651 | -0.2420686549 | 1 |
| 7.673756466 | 3.508563011 | 1 |

Η λογιστική κατανομή παίρνει ως είσοδο πραγματικές τιμές και προβλέπει την πιθανότητα η έξοδος να ανήκει στην αρχική κλάση (*κλάση 0*). Αν λοιπόν η πιθανότητα είναι μεγαλύτερη του *0.5*, τότε η έξοδος μπορεί να θεωρηθεί ότι θα πάει στην *κλάση 0*, αλλιώς η πρόβλεψη που θα πάρουμε ως έξοδο, ανήκει στην *κλάση 1*.

Για το συγκεκριμένο σύνολο δεδομένων, η λογιστική παλινδρόμηση έχει 3 συντελεστές όπως θα είχε και μία γραμμική παλινδρόμηση, για παράδειγμα:

$$Y = b_0 + b_1 * X1 + b_2 * X2 \quad (1)$$

Στόχος είναι να βρεθούν οι καλύτερες τιμές για τους συντελεστές (b0,b1,b2) με βάση το *training dataset*.

Σε αντίθεση όμως με την γραμμική παλινδρόμηση, η έξοδος μετασχηματίζεται σε μία πιθανότητα με βάση την λογιστική συνάρτηση:

$$p(class = 0) = 1/(1 + EXP(-Y)) \quad (2)$$

Για να υπολογίσουμε τις τιμές των παραμέτρων θα ακολουθήσουμε τη μέθοδο *Stochastic Gradient Descent (SGD)*. Για το σκοπό αυτό, ο χρήστης θα πρέπει να ορίζει το *ρυθμό μάθησης (learning rate)* και το συνολικό αριθμό των *εποχών (epochs) εκπαίδευσης*. Στα πλαίσια του παραδείγματος θεωρούμε ότι οι αρχικές τιμές των συντελεστών είναι μηδενικές και ο ρυθμός μάθησης είναι *γ=0.3*.

Σε κάθε εποχή τα βήματα είναι τα εξής:

1. Υπολογίζουμε μία πρόβλεψη με βάση τις τρέχουσες τιμές των συντελεστών.
2. Υπολογισμός των νέων τιμών των συντελεστών με βάση το σφάλμα της πρόβλεψης.

Με βάση τα παραπάνω για την πρώτη εποχή θα έχουμε ως εξής:

*Τιμές των συντελεστών*: $b_0 = 0.0$, $b_1 = 0.0$, $b_2 = 0.0$

*Πρώτο ζεύγος του training dataset*: X1=2.7810836, X2=2.550537003, Y=0

Χρησιμοποιώντας την παρακάτω εξίσωση θα υπολογίζουμε την πρόβλεψη:

$$p_{pred1} = 1/(1 + EXP(-(b_0 + b_1 * X1 + b_2 * X2)))$$

$$p_{pred1} = 1/(1 + EXP(-(0.0 + 0.0 * 2.7810836 + 0.0 * 2.550537003)))$$

$$\boldsymbol{p_{pred1} = 0.5}$$

Έχοντας υπολογίσει την πρόβλεψη, είμαστε σε θέση να υπολογίσουμε τις νέες τιμές των συντελεστών. Οι τιμές των συντελεστών υπολογίζονται με την χρήση της παρακάτω εξίσωσης:

$$b = b + \gamma * (Y - p_{pred}) * p_{pred} * (1 - p_{pred}) * X$$

Το *b* συμβολίζει τον συντελεστή για τον οποίο ανανεώνουμε την τιμή του.

Το *γ* είναι ο ρυθμός μάθησης και προσδιορίζει πόσο αλλάζει η τιμή των συντελεστών κάθε φορά που ανανεώνεται η τιμή τους. Στα πλαίσια του παραδείγματος ο ρυθμός μάθησης είναι ίσος με *0.3*

Παρατήρηση: Στην περίπτωση όπου υπολογίζουμε την τιμή του συντελεστή $b_0$, υποθέτουμε ότι η τιμή του X είναι ίση με 1.0 ανεξάρτητα αν δεν έχει τιμή εισόδου.

Με βάση τα παραπάνω θα έχουμε ως εξής:

$$b_0 = b_0 + \gamma * (Y - p_{pred}) * p_{pred} * (1 - p_{pred}) \ =>$$

$$b_0 = b_0 + 0.3 * (0 - 0.5) * 0.5 * (1 - 0.5) \ =>$$

$$b_0 = -0.0375$$

Με αντίστοιχο τρόπο θα βρούμε τις τιμές των συντελεστών $b_1$ και $b_2$. Συνεπώς οι τιμές που θα έχουμε είναι οι εξής:

$$b_1 = b_1 + 0.3 * (0 - 0.5) * 0.5 * (1 - 0.5) * 2.7810836 \ =>$$

$$b_1 = -0.104290635$$

Και

$$b_2 = b_2 + 0.3 * (0 - 0.5) * 0.5 * (1 - 0.5) * 2.550537003 \ =>$$

$$b_2 = -0.09564513761$$

Η παραπάνω διαδικασία (SGD) θα πρέπει να επαναληφθεί για να ανανεωθεί το μοντέλο για κάθε ζεύγος του training dataset. Στα πλαίσια του παραδείγματος θα πρέπει να επαναλάβουμε την διαδικασία για 10 εποχές. Εφόσον επαναληφθεί η διαδικασία για 10 φορές τότε οι τιμές των συντελεστών που θα προκύψουν είναι οι εξής:

$$b_0 = -0.4066054641$$

$$b_1 = \ 0.8525733164$$

$$b_2 = \ -1.104746259$$

Συνεπώς το μοντέλο μετά τις 10 εποχές θα έχει την παρακάτω μορφή με βάση τους συντελεστές που έχουν προκύψει:

$$Y = -0.4066054641 + \ 0.8525733164 * X1 + -1.104746259 * X2$$

Με βάση το παραπάνω μοντέλο τώρα μπορεί να υπολογιστεί η έξοδος για κάθε ζεύγος τιμών του training dataset. Άρα οι τιμές της εξόδου (Y) που θα προκύψουν είναι οι εξής:

| *X1* | *X2* | *Y* |
|---|---|---|
| 2.7810836 | 2.550537003 | 0.2987569857 |
| 1.465489372 | 2.362125076 | 0.145951056 |
| 3.396561688 | 4.400293529 | 0.08533326531 |
| 1.38807019 | 1.850220317 | 0.2197373144 |
| 3.06407232 | 3.005305973 | 0.2470590002 |
| 7.627531214 | 2.759262235 | 0.9547021348 |
| 5.332441248 | 2.088626775 | 0.8620341908 |
| 6.922596716 | 1.77106367 | 0.9717729051 |
| 8.675418651 | -0.2420686549 | 0.9992954521 |
| 7.673756466 | 3.508563011 | 0.905489323 |

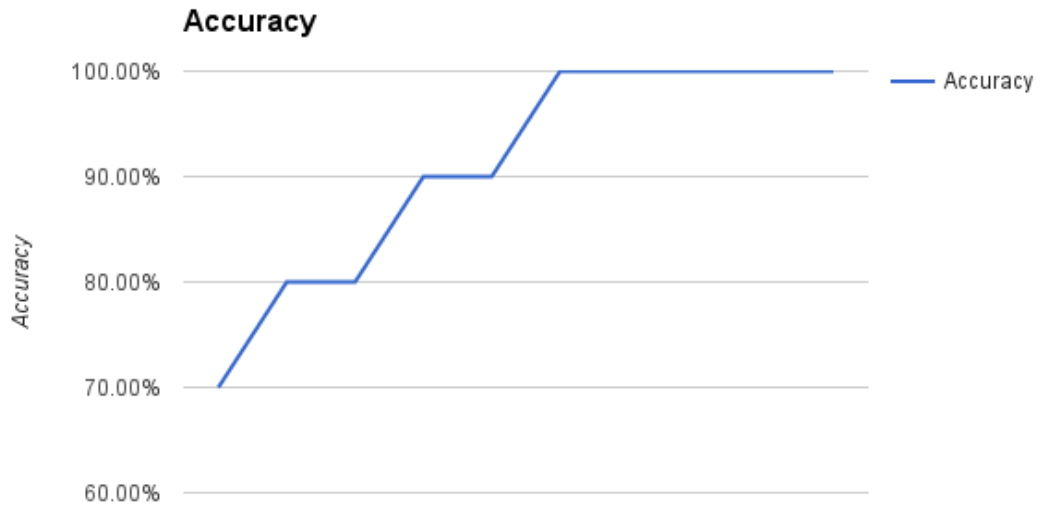
Οι παραπάνω τιμές εξόδου (Y) συμβολίζουν την πιθανότητα κάθε ζεύγος τιμών να ανήκει στην *κλάση 0*. Οι τιμές Y μπορούν να μετατραπούν σε τιμές 0 ή 1 με βάση το παρακάτω κανόνα:

$$p_{pred} = IF \ (Y < 0.5) Then \ 0 \ Else \ 1$$

Με βάση τον απλό αυτό κανόνα θα έχουμε τον αρχικό πίνακα που μας είχε δοθεί:

| *X1* | *X2* | *Y* |
|---|---|---|
| 2.7810836 | 2.550537003 | 0 |
| 1.465489372 | 2.362125076 | 0 |
| 3.396561688 | 4.400293529 | 0 |
| 1.38807019 | 1.850220317 | 0 |
| 3.06407232 | 3.005305973 | 0 |
| 7.627531214 | 2.759262235 | 1 |
| 5.332441248 | 2.088626775 | 1 |
| 6.922596716 | 1.77106367 | 1 |
| 8.675418651 | -0.2420686549 | 1 |
| 7.673756466 | 3.508563011 | 1 |

Σχετικά με την ακρίβεια του μοντέλου παρατηρούμε τα εξής:




Φαίνεται ξεκάθαρα πως το μοντέλο κατάφερε να είναι 100% ακριβές στο τέλος της 6ης εποχής.
Ο τύπος για να προσδιορίσουμε την ακρίβεια του μοντέλου είναι ο εξής:

$$accuracy = \left({}^{coreect\ predictions}/_{num\ predictions\ made}\right) * 100$$