University of Piraeus
Department of Informatics
Postgraduate Program
“Informatics”

PATTERN RECOGNITION (TME103)

Professors:
George Tsihrintzis
Dionysios Sotiropoulos

Course Assignment 2020
by

Alexander D. Davakis [MPPL18013]
alexdab@windowslive.com
John Kambilafkas [MPPL18024]
ikambilafkas@gmail.com
Andrew S. Kappos [MPPL18027]
kappos.a@hotmail.com

Index

Assignment Description

Ανάλυση και Πρόγνωση Αθλητικών Γεγονότων με χρήση Αλγορίθμων Μηχανικής Μάθησης

Στόχος της συγκεκριμένης εργασίας είναι η ανάπτυξη αλγορίθμων μηχανικής μάθησης για την πρόβλεψη του αποτελέσματος ενός ποδοσφαιρικού αγώνα. Το σύνολο των δεδομένων που θα χρησιμοποιήσετε βρίσκεται στην παρακάτω δικτυακή τοποθεσία, https://www.kaggle.com/hugomathien/soccer, υπό την μορφή μιας βάσης δεδομένων SQLite. Μάλιστα, η ωφέλιμη πληροφορία για την υλοποίηση των ζητούμενων μηχανισμών μάθησης είναι αποθηκευμένη στους πίνακες **Match** και **Team_Attributes**.

Θεωρώντας το σύνολο των διαφορετικών αγώνων της βάσης ως $\boldsymbol{M} = \{\boldsymbol{m_1}, \dots, \boldsymbol{m_n}\}$ και το αντίστοιχο σύνολο των διαφορετικών ομάδων ως $\boldsymbol{T} = \{\boldsymbol{t_1}, \dots, \boldsymbol{t_K}\}$, τότε η πιο αφηρημένη αναπαράσταση του κάθε αγώνα μπορεί να πραγματοποιηθεί ως μια διατεταγμένη τριάδα της μορφής $\boldsymbol{m} = \langle \boldsymbol{h}, \boldsymbol{a}, \boldsymbol{r} \rangle$, με $\boldsymbol{h} \neq \boldsymbol{a}$, όπου η $\boldsymbol{h} \in \boldsymbol{T}$ είναι η ομάδα που αγωνίζεται εντός έδρας (home team), $\boldsymbol{a} \in \boldsymbol{T}$ είναι η ομάδα που αγωνίζεται εκτός έδρας (away team) και $\boldsymbol{r} \in \{\boldsymbol{H}, \boldsymbol{D}, \boldsymbol{A}\}$ το αποτέλεσμα του αγώνα. Συγκεκριμένα, το **H** (home win) υποδηλώνει νίκη της ομάδας που αγωνίζεται εντός έδρας, το **D** (Draw) υποδηλώνει την ισόπαλη έκβαση του αγώνα και το **A** (away win) υποδηλώνει την νίκη της ομάδας που αγωνίζεται εκτός έδρας. Αν με $\boldsymbol{G_m(h)} \geq \boldsymbol{0}$ και $\boldsymbol{G_m(a)} \geq \boldsymbol{0}$ συμβολίσουμε το πλήθος των τερμάτων που επιτυγχάνονται από τις ομάδες εντός και εκτός έδρας κατά την διεξαγωγή του αγώνα $\boldsymbol{m}$, τότε η μεταβλητή έκβασης του αγώνα $\boldsymbol{r}$ μπορεί να ορισθεί συναρτήσει της διαφοράς των συνολικών τερμάτων $\boldsymbol{\Delta G_m(h,a)} = \boldsymbol{G_m(h)} - \boldsymbol{G_a(h)}$ ως:

$$\boldsymbol{r} = \begin{cases} \mathbf{H}, \boldsymbol{\Delta G_m(h,a)} > \mathbf{0}; \\ \mathbf{D}, \boldsymbol{\Delta G_m(h,a)} = \mathbf{0}; \\ \mathbf{A}, \boldsymbol{\Delta G_m(h,a)} < \mathbf{0}. \end{cases}$$

Η διαδικασία εκπαίδευσης των εμπλεκόμενων ταξινομητών θα πρέπει να βασιστεί σε ένα σύνολο χαρακτηριστικών γνωρισμάτων της κάθε ομάδας καθώς και σε ένα σύνολο προγνωστικών (odds) για την πιθανή έκβαση του κάθε αγώνα από έναν αριθμό στοιχηματικών εταιρειών. Συγκεκριμένα, κάθε ομάδα $\boldsymbol{t} \in \boldsymbol{T}$ είναι συσχετισμένη με ένα διάνυσμα χαρακτηριστικών $\boldsymbol{\varphi(t)} \in \mathbb{R}\boldsymbol{8}$ τα οποία αντιστοιχούν στις παρακάτω στήλες του πίνακα **Team_Attributes** {**buildUpPlaySpeed**, **buildUpPlayPassing**, **chanceCreationPassing**, **chanceCreationCrossing**, **chanceCreationShooting**, **defencePressure**, **defenceAggregation**, **defenceTeamWidth**}. Επιπλέον, κάθε αγώνας $\boldsymbol{m} \in \boldsymbol{M}$ είναι συσχετισμένος με τέσσερα διανύσματα προγνωστικών $\boldsymbol{\psi_\kappa(m)} \in \mathbb{R}^3$ με $\boldsymbol{k} \in \{\boldsymbol{B365}, \boldsymbol{BW}, \boldsymbol{IW}, \boldsymbol{LB}\}$ τα αντιστοιχούν στις παρακάτω στήλες του πίνακα **Match** {**B365H**, **B365D**, **B365A**,**BWH**, **BWD**, **BWA**, **IWH**, **IWD**, **IWA**, **LBH**, **LBD**, **LBA**}. Δηλαδή, το κάθε διάνυσμα $\boldsymbol{\psi_\kappa(m)} = \left[\boldsymbol{d_k^H(m)}, \boldsymbol{d_k^D(m)}, \boldsymbol{d_k^A(m)}\right]$ $\boldsymbol{\psi\kappa(m)}=[\boldsymbol{dkH(m)},\boldsymbol{dkD(m)},\boldsymbol{dkA(m)}]$ συγκεντρώνει τις στοιχηματικές αποδόσεις για κάθε πιθανή έκβαση του αγώνα $\boldsymbol{m}$ για κάθε στοιχηματική εταιρεία $\boldsymbol{k} \in \boldsymbol{B} = \{\boldsymbol{B365}, \boldsymbol{BW}, \boldsymbol{IW}, \boldsymbol{LB}\}$ $\in \boldsymbol{B}= \{\boldsymbol{B365},\boldsymbol{BW},\boldsymbol{IW},\boldsymbol{LB}\}$. Λάβετε υπόψιν πως υπάρχουν εγγραφές στον πίνακα **Match** για τις οποίες τα αντίστοιχα διανύσματα προγνωστικών έχουν μηδενικές τιμές. Οι συγκεκριμένες εγγραφές θα πρέπει να αφαιρεθούν.

Ερωτήματα:

I. Να υλοποιήσετε ένα γραμμικό νευρωνικό δίκτυο, ώστε ο εκπαιδευόμενος ταξινομητής να υλοποιεί μια συνάρτηση διάκρισης της μορφής $\boldsymbol{g_{\kappa}(\psi_{\kappa}(m))} : \mathbb{R}^{3} \rightarrow \{\boldsymbol{H}, \boldsymbol{D}, \boldsymbol{A}\}$ για κάθε στοιχηματική εταιρεία. Να αναγνωρίσετε την στοιχηματική εταιρεία τα προγνωστικά της οποίας οδηγούν σε μεγαλύτερη ακρίβεια ταξινόμησης.

II. Να υλοποιήσετε ένα πολυστρωματικό νευρωνικό δίκτυο, ώστε ο εκπαιδευόμενος ταξινομητής να υλοποιεί μια συνάρτηση διάκρισης της μορφής $\boldsymbol{g_{\kappa}(\psi_{\kappa}(m))} : \mathbb{R}^{3} \rightarrow \{\boldsymbol{H}, \boldsymbol{D}, \boldsymbol{A}\}$ για κάθε στοιχηματική εταιρεία. Να αναγνωρίσετε την στοιχηματική εταιρεία τα προγνωστικά της οποίας οδηγούν σε μεγαλύτερη ακρίβεια ταξινόμησης.

III. Να υλοποιήσετε ένα πολυστρωματικό νευρωνικό δίκτυο, ώστε ο εκπαιδευμένος ταξινομητής να υλοποιεί μια συνάρτηση διάκρισης της μορφής $\boldsymbol{g(\Phi(m))} : \mathbb{R}^{28} \rightarrow \{\boldsymbol{H}, \boldsymbol{D}, \boldsymbol{A}\}$**,** όπου το $\boldsymbol{\Phi(m)} \in \mathbb{R}^{28}$ αντιστοιχεί στο πλήρες διάνυσμα χαρακτηριστικών του κάθε αγώνα που δίνεται από την σχέση:

$$\boldsymbol{\Phi(m)} = [\boldsymbol{\varphi(h)}, \boldsymbol{\varphi(\alpha)}, \boldsymbol{\psi_{B365}(m)}, \boldsymbol{\psi_{BW}(m)}, \boldsymbol{\psi_{IW}(m)}, \boldsymbol{\psi_{LW}(m)}]$$

IV. Να εφαρμόσετε τον αλγόριθμο ομαδοποίησης c – means επάνω στο σύνολο των διανυσμάτων προγνωστικών $\boldsymbol{\Psi_{\kappa}} = \{\boldsymbol{\psi_{\kappa}(m)} \in \mathbb{R}^{3} : \boldsymbol{m} \in \boldsymbol{M}\}$ για κάθε στοιχηματική εταιρεία $\boldsymbol{k} \in \boldsymbol{B}$, θέτοντας την τιμή του c ίση με 3. Με το τρόπο αυτό, θα παράξετε μια διαφορετική διαμέριση του συνόλου των αγώνων $\boldsymbol{M}$ σε τρείς συστάδες για κάθε στοιχηματική εταιρεία. Λαμβάνοντας υπόψιν το αποτέλεσμα του κάθε αγώνα να υπολογίσετε την κατανομή των τριών αποτελεσμάτων εντός της κάθε συστάδας για κάθε στοιχηματική εταιρεία. Υπάρχει κάποιο αποτέλεσμα που να επικρατεί σε συχνότητα εντός της κάθε συστάδας;

Παρατηρήσεις:

I. Για κάθε ταξινομητή που θα υλοποιήσετε θα πρέπει να αναφέρετε την ταξινομιτική του ακρίβεια τόσο κατά τη φάση της εκπαίδευσης όσο και κατά τη φάση του ελέγχου σύμφωνα με την μέθοδο της 10-πλης διεπικύρωσης (**10 fold cross validation**).

II. Στο αρχείο **EuropeanDatabaseRetriever.m** σας παρέχεται κώδικας για την άντληση των δεδομένων από τη βάση SQLite.

III. Παραδοτέα της εργασίας αποτελούν ο **κώδικας** της υλοποίησης σας σε MATLAB ή Python καθώς και ένα συνοδευτικό **κείμενο τεκμηρίωσης**.

IV. Μπορείτε να εργασθείτε σε ομάδες των **δύο** ή **τριών ατόμων**.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ!

Πίσω στην κορυφή

Return To Index

Introduction

Machine Learning

Μηχανική Εκμάθηση [**Machine Learning**] καλείται η μελέτη αλγορίθμων Ηλεκτρονικών Υπολογιστών οι οποίοι βελτιώνονται αυτοματοποιημένα μέσα από την επαναλαμβανόμενη χρήση τους (εμπειρία) και τη χρήση δεδομένων (εκπαίδευση) και αποτελεί μέρος του κλάδου της Τεχνητής Νοημοσύνης.

Η Μηχανική Εκμάθηση στοχεύει στην υλοποίηση αλγοριθμικών συστημάτων στα οποία μηχανές μαθαίνουν και εκπαιδεύουν τους εαυτούς τους να επιτυγχάνουν σε εργασίες για τις οποίες δεν έχουν προγραμματιστεί ρητώς. Οι σχετικοί αλγόριθμοι ξεκινούν με είσοδο κάποια αρχικά δεδομένα και εκπαιδεύουν ένα μοντέλο βασισμένο σε δεδομένα δειγμάτων, αποκαλούμενα ως «δεδομένα εκπαίδευσης», προκειμένου να κάνουν προβλέψεις ή να παίρνουν αποφάσεις.

Ενώ για απλές εργασίες είναι δυνατόν να προγραμματίσουμε αποτελεσματικούς αλγόριθμους και δεν απαιτείται μάθηση από την πλευρά του υπολογιστή, για πιο ανεπτυγμένες εργασίες μπορεί να είναι πρόκληση για έναν άνθρωπο να δημιουργεί χειρωνακτικά τους απαραίτητους αλγόριθμους. Στην πράξη, αποδεικνύεται πως τα να βοηθήσουμε τη μηχανή να αναπτύξει το δικό της αλγόριθμο μπορεί να είναι πιο αποτελεσματικό από το να τις ορίσουμε κάθε απαραίτητο βήμα.

Οι αλγόριθμοι μηχανικής εκμάθησης χρησιμοποιούνται σε πληθώρα εφαρμογών, όπως **φιλτράρισμα e-mail** και **υπολογιστική όραση**, εργασίες για τις οποίες είναι δύσκολο ή ανέφικτο να κατασκευαστούν συμβατικοί αλγόριθμοι. Ένα υποσύνολο της μηχανικής εκμάθησης συνδέεται στενά με την Υπολογιστική Στατιστική η οποία επικεντρώνεται στο να κάνει **προβλέψεις** χρησιμοποιώντας υπολογιστές. Η μελέτη της Μαθηματικής Βελτιστοποίησης παραδίδει μεθόδους, θεωρία και πεδία εφαρμογών στο τομέα της μηχανικής εκμάθησης και η **εξόρυξη δεδομένων** είναι σχετικός τομέας μελέτης, επικεντρωνόμενος στη Διερευνητική Ανάλυση Δεδομένων μέσα από εκμάθηση δίχως επιτήρηση. Στην εφαρμογή της κατά μήκος επιχειρηματικών προβλημάτων, η μηχανική εκμάθηση αναφέρεται επίσης ως **Προγνωστική Ανάλυση**.

Ο τομέας της Μηχανικής Εκμάθησης αξιοποιεί διάφορες προσεγγίσεις ώστε να διδάξει τους υπολογιστές να επιτυγχάνουν εργασίες για τις οποίες δεν υπάρχουν πλήρως ικανοποιητικοί καθολικοί αλγόριθμοι. Όταν υπάρχει τεράστιο πλήθος δυνατών απαντήσεων, μία προσέγγιση είναι να ορίσουμε μερικές από τις σωστές απαντήσεις ως έγκυρες. Αυτό μπορεί να χρησιμοποιηθεί από τον υπολογιστή σα δεδομένα εκπαίδευσης για να βελτιώσει τους αλγορίθμους που χρησιμοποιεί. Για παράδειγμα, προκειμένου να εκπαιδευτεί ένα σύστημα ώστε ν αναγνωρίζει ψηφιακούς χαρακτήρες, χρησιμοποιείται συχνά το σύνολο δεδομένων MNIST που περιλαμβάνει χειρόγραφα ψηφία.

Πίσω στην κορυφή

Return To Index

Προσεγγίσεις Μηχανικής Εκμάθησης

Οι προσεγγίσεις στη μηχανική εκμάθηση παραδοσιακά χωρίζονται σε 3 ευρείες κατηγορίες, ανάλογα με τη φύση του σήματος ή της ανάδρασης που είναι διαθέσιμα στο αυτοδιδασκόμενο σύστημα :

- Εκμάθηση Υπό Επιτήρηση [Supervised Learning] : Ο υπολογιστής παραλαμβάνει εισόδους με κατηγοριοποιήσεις, δηλαδή παραδειγματικές εισόδους και τις επιθυμητές εξόδους τους, και μαθαίνει στοχεύοντας στο να αναπτύξει γενικό κανόνα ο οποίος χαρτογραφεί εισόδους σε εξόδους.
- Εκμάθηση Δίχως Επιτήρηση [Unsupervised Learning] : Ο υπολογιστής παραλαμβάνει εισόδους δίχως κατηγοριοποιήσεις και μαθαίνει στοχεύοντας στο να βρει δομή στην είσοδο. Η εκμάθηση δίχως επιτήρηση μπορεί να είναι αυτοσκοπός (πχ, ανακάλυψη κρυμμένων μοτίβων σε δεδομένα) ή μέσο προς κάποιο σκοπό (εκμάθηση χαρακτηριστικών)
- Εκμάθηση Υπό Ασθενή Επιτήρηση [Weakly Supervised Learning] : Ο υπολογιστής παραλαμβάνει εισόδους με μερική, ασθενή κατηγοριοποίηση, δηλαδή παραδειγματικές εισόδους στις οποίες έχουν ανατεθεί ατελείς έξοδοι, και μαθαίνει στοχεύοντας στο να αναπτύξει γενικό κανόνα ο οποίος χαρτογραφεί εισόδους σε εξόδους.

  Η μερική κατηγοριοποίηση που παρέχεται έχει συμβεί με μικρό κόστος και προέλθει από θορυβώδεις, περιορισμένες και ανακριβείς πηγές όμως έχει αποδειχθεί ότι μπορεί να οδηγεί στην ανάπτυξη ισχυρών προβλεπτικών μοντέλων.
- Εκμάθηση Ενίσχυσης [Reinforcement Learning] : Ο υπολογιστής παραλαμβάνει συνεχώς εισόδους διαδρώντας με ένα δυναμικό περιβάλλον, από το οποίο παρέχεται ανάδραση η οποία είναι ανάλογη με ανταμοιβές, και μαθαίνει στοχεύοντας στη μεγιστοποίηση της ανάδρασης-ανταμοιβής. Αυτή η εκμάθηση αξιοποιείται για την εκτέλεση συγκεκριμένου σκοπού όπως το να οδηγήσει ένα όχημα ή να παίξει ένα ανταγωνιστικό παιχνίδι.

Έχουν αναπτυχθεί και άλλες προσεγγίσεις οι οποίες δεν ταιριάζουν τέλεια σε αυτή την τριπλή κατηγοριοποίηση ενώ μερικές φορές ένα σύστημα μηχανικής εκμάθησης αξιοποιεί περισσότερες από μία προσεγγίσεις όπως για παράδειγμα η μοντελοποίηση θεμάτων, η μείωση διαστασιμότητας και η μετα-μάθηση.

Πλέον, η Βαθιά Εκμάθηση [Deep Learning] έχει γίνει η κυρίαρχη προσέγγιση για πολλές εξελισσόμενες εργασίες στο πεδίο της μηχανικής εκμάθησης.

Πίσω στην κορυφή

Return To Index

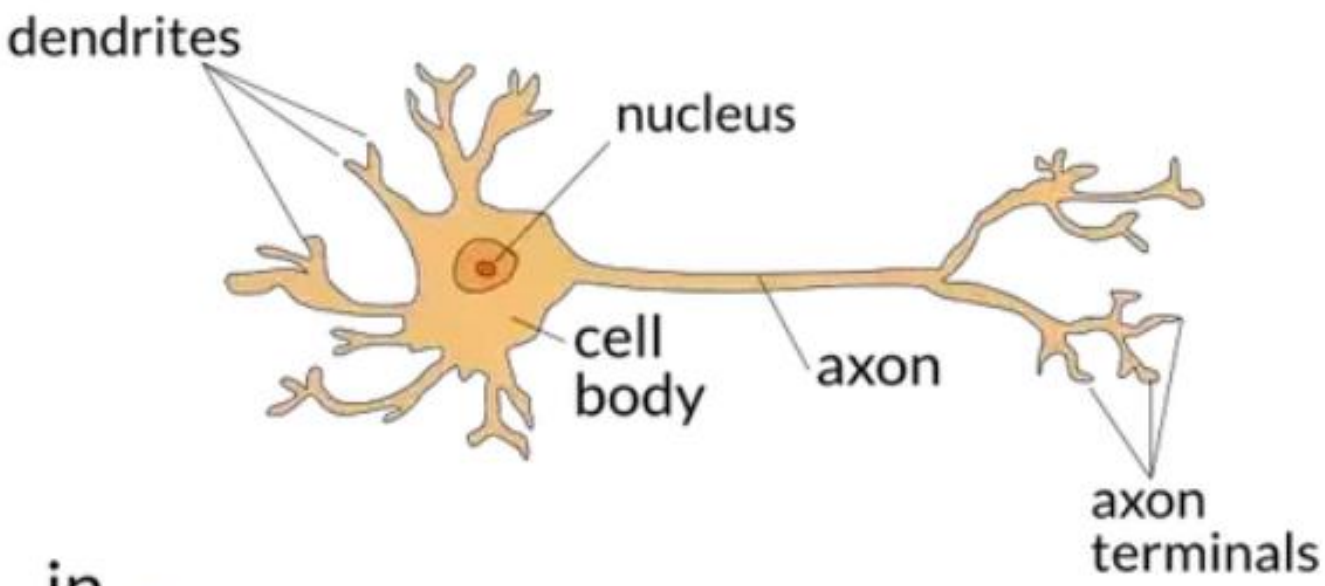
(Artificial) Neural Networks and Classifiers

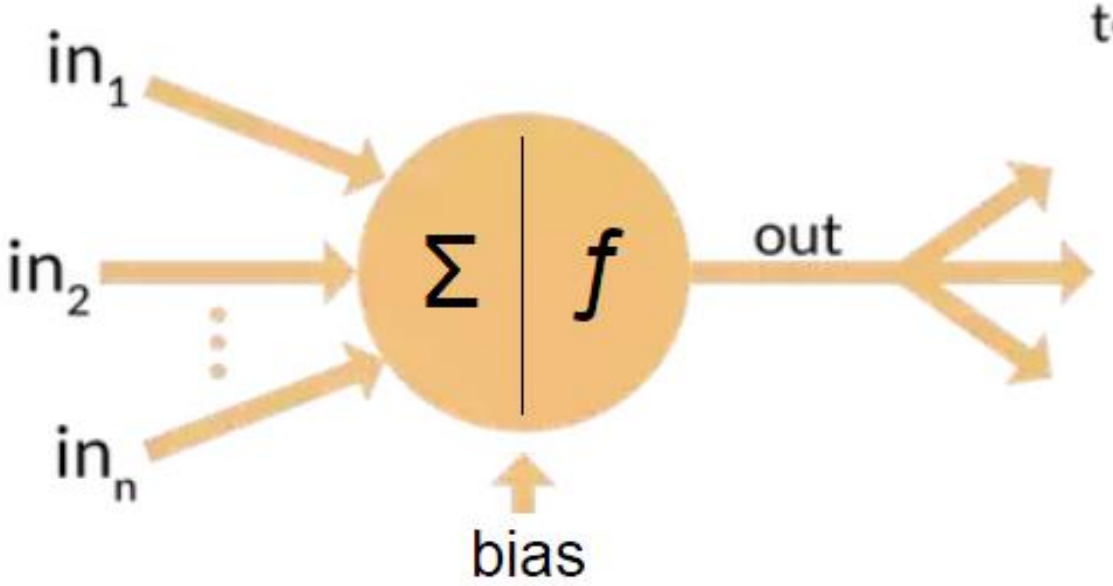
Τα τεχνητά **Νευρωνικά Δίκτυα** είναι τεχνητά συστήματα με δομή εμπνευσμένη από τη λειτουργία του βιολογικού νευρικού συστήματος και του εγκεφάλου. Αυτά και οι Ταξινομητές είναι τα κύρια εργαλεία της Μηχανικής Εκμάθησης.

Η βασική μονάδα του βιολογικού νευρωνικού συστήματος είναι ο **νευρώνας** [**neuron**], ο οποίος δομείται ως εξής :

- Ο πυρήνας [kernel] δέχεται σήματα από άλλους νευρώνες μέσω καναλιών εισόδου και τα επεξεργάζεται ώστε να δημιουργήσει ένα καινούργιο σήμα. Εφόσον το σήμα έχει επαρκή ισχύ ενεργοποιείται η έξοδος του νευρώνα και παράγεται ένα σήμα εξόδου το οποίο μεταδίδεται μέσω ενός καναλιού εξόδου.
- Οι δενδρίτες [dendrites] είναι τα κανάλια εισόδου ενός νευρώνα.
- Άξονας [axon] καλείται το κανάλι εξόδου κάθε νευρώνα και συνδέεται με τους δενδρίτες άλλων νευρώνων.
- Σύναψη [synapse] καλείται η σύνδεση μεταξύ ενός άξονα και ενός δενδρίτη. Η ισχύς της μεταβάλλει ανάλογα το μεταδιδόμενο σήμα.

Ο ανθρώπινος εγκέφαλος αποτελείται από δεκάδες δισεκατομμύρια νευρώνες.





Πίσω στην κορυφή
Return To Index

Σε ένα τεχνητό νευρωνικό δίκτυο, το κύριο επεξεργαστικό στοιχείο-κόμβος είναι μια τεχνητή απομίμηση του βιολογικού νευρώνα. Ένας τεχνητός τέτοιος νευρώνας είναι ένα χρονικώς αναλλοίωτο σύστημα χωρίς μνήμη, με πολλές εισόδους και μία έξοδο και δομείται ως εξής :

- Ο αντίστοιχος πυρήνας του περιλαμβάνει τη συνάρτηση ενεργοποίησης f [activation function] η οποία μπορεί να είναι είτε μια συνάρτηση κατωφλιού που παράγει μη-μηδενική τιμή αν το άθροισμα των εισόδων input_i είναι επαρκώς ισχυρό, είτε μια συνεχής συνάρτηση. Η συνάρτηση δέχεται ως είσοδο το άθροισμα των εισόδων x_i πολλαπλασιασμένες με ένα βάρος [weight] w_i και αθροισμένο με μια σταθερά κατωφλιού [bias] $\theta = b \cdot w_b$ και παράγει έξοδο ανάλογα με τον τύπο της
$$f = f\left(\sum_{i=0}^{n}(x_i \cdot w_i + b \cdot w_b)\right).$$
- Οι αντίστοιχοι δενδρίτες είναι ορισμένου πλήθους n είσοδοι δεδομένων, σε μορφή διανύσματος $\text{input}_i = x_i$. Όλες οι αρχικές είσοδοι του νευρωνικού δικτύου αποτελούν το στρώμα εισόδου για το δίκτυο.
- Ο αντίστοιχος άξονας είναι η έξοδος output της συνάρτησης ενεργοποίησης, της οποίας η τιμή είτε διαδίδεται στην είσοδο του επόμενου στρώματος μέσω μιας σύναψης, είτε εξέρχεται από το σύστημα, συνήθως ως μέρος ενός διανύσματος εξόδου. Όλες οι τελικές έξοδοι του νευρωνικού δικτύου αποτελούν το στρώμα εξόδου για το δίκτυο.
- Η αντίστοιχη σύναψη είναι η σύνδεση μεταξύ νευρώνων-κόμβων του νευρωνικού δικτύου. Όπως και στους βιολογικούς εγκεφάλους, η σύνδεση ελέγχεται από τη δύναμη ή το εύρος σύνδεσης μεταξύ δυο κόμβων (συναπτικό βάρος). Πολλαπλές συνάψεις μπορούν να συνδέουν τους ίδιους νευρώνες, με κάθε σύναψη να έχει διαφορετικό επίπεδο επίδρασης (σκανδάλη) πάνω στην πυροδότηση του νευρώνα κι ενεργοποίηση του επόμενου. Μαθηματικά, αναπαρίσταται ως διάνυσμα βάρους.

Ταξινομητής είναι ένας αλγόριθμος ταξινόμησης δεδομένων εισόδου σε ομάδες (κλάσεις) εξόδου χρησιμοποιώντας μία μαθηματική συνάρτηση. Ένα νευρωνικό δίκτυο μπορεί να αποτελείται από πολλές μονάδες ταξινομητών ως νευρώνων του και να εκτελεί σύνθετη ταξινόμηση με χρήση πολλαπλών κριτηρίων κατά συνέπεια μαθαίνοντας πολλά χαρακτηριστικά από τα δεδομένα εισόδου του.

Ταξινομητές χρησιμοποιούνται και από Μηχανές Διανυσμάτων Υποστήριξης [**S**upport **V**ector **M**achines] οι οποίες ειδικεύονται αποκλειστικά στην ταξινόμηση των δεδομένων εισόδου τους. Οι μηχανές αυτές είναι μοντέλα εκμάθησης υπό επιτήρηση με σχετιζόμενους αλγόριθμους οι οποίοι αναλύουν δεδομένα προς ταξινόμηση και στατιστική ανάλυση παλινδρόμησης, συνήθως γραμμικής.

Πίσω στην κορυφή

Return To Index

k-Fold Cross-Validation

Διεπικύρωση [**Cross-validation**], καλούμενη και Εκτίμηση Περιστροφής [Rotation Estimation] ή Δοκιμή Εκτός Δείγματος [Out-of-Sample Testing], είναι ένα σύνολο τεχνικών επικύρωσης για (στατιστικά) μοντέλα. Αξιολογούν το πώς το μοντέλο πρόβλεψης που αναπτύχθηκε από εκπαίδευση βασιζόμενο σε συγκεκριμένα, γνωστά δεδομένα μπορεί να προβλέπει με επιτυχία όταν βασίζεται σε άγνωστα δεδομένα (γενικοποίηση). Αξιοποιείται στη Θεωρία Στατιστικής Εκμάθησης της Μηχανικής Εκμάθησης, σε εφαρμογές εκμάθησης υπό επιτήρηση.

Η Διεπικύρωση μπορεί να είναι Εξαντλητική ή Μη-Εξαντλητική, ανάλογα με το αν υπολογίζεται κάθε τρόπος διαμέρισης του αρχικού δείγματος.

Η **k-πλή Διεπικύρωση** [**k-Fold Cross-validation**] είναι μια μέθοδος Μη-Εξαντλητικής Διεπικύρωσης όπου το αρχικό δείγμα διαμερίζεται τυχαία σε k ίσα υποδείγματα. Κάθε υποδείγμα επιλέγεται διαδοχικά ως σύνολο επικύρωσης για τον έλεγχο του εκπαιδευόμενου μοντέλου, με τα υπόλοιπα k-1 υποδείγματα να χρησιμοποιούνται ως δεδομένα εκπαίδευσης. Η ζητούμενη εκτίμηση γενικοποίησης συνήθως θα προέλθει από το μέσο όρο των αποτελεσμάτων αυτών και συχνά η σταθερά k επιλέγεται να ισούται με 10.

Πίσω στην κορυφή

Return To Index

C-means Clustering

Συσταδοποίηση [**Clustering**] ή **Ανάλυση Συστάδας** [**Cluster Analysis**] ονομάζεται η εργασία όπου ομαδοποιούμε αντικείμενα σε συστάδες έτσι ώστε κάθε συστάδα περιέχει σχετικά αντικείμενα, ομοιότερα των άλλων συστάδων. Η ομοιότητα αυτή είναι αφηρημένος όρος και ορίζεται από την εκάστοτε κατάσταση στην οποία δουλεύουμε, με μέτρα ομοιότητας την απόσταση, τη συνδεσιμότητα και την ένταση.

Η συσταδοποίηση Αποτελεί την κύρια εργασία της ανάλυσης διερευνητικών δεδομένων και συνήθη τεχνική της ανάλυσης στατιστικών δεδομένων. Χρησιμοποιείται σε αρκετά πεδία όπως :

- αναγνώριση προτύπου
- ανάλυση εικόνας
- ανάκτηση πληροφορίας
- πηγαία κωδικοποίηση
- βιοπληροφορική
- συμπίεση δεδομένων
- μείωση ρυθμού bit
- γραφικά υπολογιστών
- μηχανική εκμάθηση

Χρησιμοποιεί ποικίλους αλγορίθμους οι οποίοι διαφέρουν σημαντικά στο πώς ορίζουν τις συστάδες και τον αποδοτικό υπολογισμό τους.

Μια μορφή συσταδοποίησης είναι η **Ασαφής Συσταδοποίηση** όπου κάθε στοιχείο μπορεί να ανήκει σε περισσότερες από μία συστάδες ενώ η αντίθετη μορφή της καλείται σκληρή ή μη-ασαφής συσταδοποίηση [Hard / Non-Fuzzy clustering]. Πιο συγκεκριμένα, μέσα από την ασαφή συσταδοποίηση, σε κάθε εξεταζόμενο στοιχείο ανατίθεται ένας βαθμός για το κατά πόσο ανήκει σε κάθε συστάδα ανάλογα με την απόσταση του από το κέντρο. Έτσι, ένα στοιχείο στην άκρη της συστάδας ανήκει λιγότερο σε αυτή από ότι ανήκει ένα στοιχείο που βρίσκεται κοντά στο κέντρο της.

Η **Συσταδοποίηση C-means** [**F**uzzy **C**-**M**eans clustering] είναι ένας δημοφιλής αλγόριθμος ασαφούς συσταδοποίησης, πολύ όμοιος με τον αλγόριθμο k-means ο οποίος αξιοποιείται κατά κόρον στην επεξεργασία σήματος. Η διαδικασία που ακολουθεί είναι η ακόλουθη :

i. Ορίζεται το πλήθος των συστάδων
ii. Αναθέτει σε κάθε αντικείμενο συστάδας με τυχαίο τρόπο ένα συντελεστή
iii. Επαναλαμβάνει τα ακόλουθα έως ότου ο αλγόριθμος συγκλίνει, δηλαδή η αλλαγή των συντελεστών μεταξύ διαδοχικών επαναλήψεων είναι μικρότερη του κατωφλιού ευαισθησίας ε :
    a. Για κάθε συστάδα, υπολογίζει το κεντροειδές της
    b. Για κάθε αντικείμενο, υπολογίζει τους συντελεστές συμμετοχής του σε κάθε συστάδα

Πίσω στην κορυφή
Return To Index

Presentation and Execution of the Assignment's code

Η εργασία εκπονείται με τη χρήση 4 αρχείων :

- AssignmentCode.m, το οποίο περιλαμβάνει τον κώδικα που καθοδηγεί την εκπόνηση της εργασίας. Είναι γραμμένος σε sections και χωρίζεται γενικώς στην επεξεργασία και προετοιμασία των δεδομένων και στην κλήση του αρχείου Answer.m ανά ερώτημα.
- EuropeanSoccerDatabaseRetriever.m και EuropeanSoccerDatabaseRetrieverImproved.m, τα οποία περιέχουν τον κώδικα ο οποίος ανακτά τα ζητούμενα δεδομένα από το αρχείο database.sqlite. Η έκδοση 'Improved' περιέχει τροποποιημένο κώδικα βασισμένο στο πρώτο, που όμως είναι βελτιστοποιημένος ως προς τα ζητούμενα της εργασίας.
- database.sqlite, το οποίο είναι η βάση δεδομένων European Soccer Database από την ιστοσελίδα της kaggle που κατεβάζουμε κατά τις οδηγίες της εκφώνησης
- Answer.m, το οποίο εκτελεί τον υπολογισμό και παρουσίαση καθεμιάς από τις απαντήσεις των τεσσάρων ερωτημάτων της εργασίας.

Ο κώδικας αναπτύχθηκε βελτιστοποιημένα με γνώμονα την ταχύτητα και την αποδοτικότητα της χρήσης του Matlab. Αξιοποιεί τις ιδιαίτερες εσωτερικές δομές της πλατφόρμας όπως λογική δεικτοδότηση [logical indexing], αποφυγή βρόγχων, προδιάθεση πόρων, τοπικότητα συναρτήσεων, επικαιρότητα εντολών και χρήση τομέων ώστε να τρέχει όσο το δυνατόν γρηγορότερα και ασφαλέστερα κατά τις οδηγίες της πλατφόρμας.

Επιλέξαμε να αξιοποιήσουμε 3 αρχεία κώδικα έναντι ενός, ώστε να ισορροπήσουμε την ταχύτητα με την αναγνωσιμότητα του κώδικα, καθότι η τελευταία δοκιμαζόταν ακόμη και με τη χρήση τομέων. Για το μέγεθος, τις κλήσεις που πραγματοποιούνται και τη μνήμη που δεσμεύεται κατά την εκτέλεση, η χρήση αυτού του πλήθους αρχείων αποδείχτηκε και μετρήθηκε αξιοπρεπής.

Execution Instructions

Η εργασία απαιτεί τα αρχεία AssignmentCode.m, Answer.m, database.sqlite και EuropeanSoccerDatabaseRetrieverImproved.m, να είναι εναποθετημένα στον ίδιο φάκελο, σε μονοπάτι αγγλικών χαρακτήρων και χωρίς κενά καθώς και περίπου 306Mb αποθηκευτικού χώρου στον ίδιο φάκελο. Η εργασία εκκινείται απλώς με την εκτέλεση του κώδικα του αρχείου *AssignmentCode.m*, είτε ανά section [Run Section : Ctr+Enter κι έπειτα Advance : Ctl+Down], είτε ολοσχερώς με την εντολή Run[F5]. Αυτό θα καλέσει τα scripts *EuropeanSoccerDatabaseRetrieverImproved.m* και *Answer.m* κατά τους διάφορους τομείς της εκτέλεσης τα οποία θα καταγράψουν τις απαντήσεις τους στην κονσόλα ελέγχου και θα αποθηκεύσουν σχετικά .png αρχεία στους αντίστοιχους φάκελους (Answer_1_plots, Answer_2_plots και Answer_3_plots).

Πίσω στην κορυφή

Return To Index

AssignmentCode.m

Το αρχείο *AssignmentCode.m* περιλαμβάνει σε τμήματα τον κώδικα επεξεργασίας και προετοιμασίας των δεδομένων και των κλήσεων του script Answer με κατάλληλες παραμέτρους, προς απάντηση του εκάστοτε ερωτήματος.

Είναι χωρισμένος στους τομείς :

- Δήλωσης της Εργασίας και των εργαζόμενων φοιτητών, εκκίνησης χρονομετρητή, δήλωσης αρχείου αποθήκευσης των εξόδων της κονσόλας και καταστολής προειδοποιήσεων.
- Προετοιμασίας των δεδομένων του 1ου και 2ου ερωτήματος της εργασίας και αποθήκευσή τους στα αρχεία *AssignmentMaterials.mat* και *FeaturesBCompanies.mat*
- Προετοιμασίας των δεδομένων του 3ου ερωτήματος της εργασίας και αποθήκευσή τους στο αρχείο *FeaturesMulti.mat*
- Κλήσης του *Answer* προς απάντηση του 1ου ερωτήματος, με κατάλληλες παραμέτρους
- Κλήσης του *Answer* προς απάντηση του 2ου ερωτήματος, με κατάλληλες παραμέτρους
- Κλήσης του *Answer* προς απάντηση του 3ου ερωτήματος, με κατάλληλες παραμέτρους
- Κλήσης του *Answer* προς απάντηση του 4ου ερωτήματος, με κατάλληλες παραμέτρους
- Δήλωσης τέλους της Εργασίας, παύσης χρονομετρητή, κλείσιμου της καταγραφής και επαναφοράς προειδοποιήσεων.

Στους υπολογισμούς και στις διάφορες απαντήσεις,
το αποτέλεσμα «Νίκη της Εντός Έδρας ομάδας» [Home Win] αντιστοιχεί στο ψηφίο «1»,
το αποτέλεσμα «Νίκη της Εκτός Έδρας ομάδας» [Away Win] αντιστοιχεί στο ψηφίο «2» και
το αποτέλεσμα «Ισοπαλία» [Draw] αντιστοιχεί στο ψηφίο «3»

EuropeanSoccerDatabaseRetrieverImproved.m

Το αρχείο *EuropeanSoccerDatabaseRetrieverImproved.m* περιλαμβάνει τον κώδικα ανάκτησης των σχετικών με την εργασία δεδομένων από τη βάση δεδομένων *database.sqlite*. Σε βελτιστοποίηση του αρχικού παρεχόμενου αρχείου *EuropeanSoccerDatabaseRetriever.m*, το αρχείο ανακτά μονάχα τους πίνακες Match και TeamAttributes, αφαιρώντας από αυτούς τις στήλες των μη χρησιμοποιούμενων στοιχηματικών εταιρειών και τις εγγραφές που περιλαμβάνουν μηδενικές αποδόσεις.

Answer.m

Το αρχείο Answer.m περιλαμβάνει σε τμήματα τον κώδικα ο οποίος απαντά στα ερωτήματα. Καλείται με συγκεκριμένες παραμέτρους ανά ερώτημα :
τον αριθμό του ερωτήματος, τα δεδομένα προς χρήση, την παράμετρο ταξινόμησης και τις επιλογές ταξινόμησης.

Είναι χωρισμένος στους τομείς :

- Κεντρική συνάρτηση **Answer** η οποία
  - δημιουργεί το μονοπάτι αποθήκευσης των αρχείων τύπου .png που παρουσιάζουν τις ακρίβειες των διάφορων διαμερίσεων ανά fold του cross-validation
  - φορτώνει τα δεδομένα ως struct και κατασκευάζει περιγραφικές μεταβλητές
  - προδιαθέτει τους πόρους των πινάκων μη δυναμικού μεγέθους
  - καλεί τις συναρτήσεις υπολογισμού μεταφέροντας τις παραμέτρους της,
  - εκτελεί τους υπολογισμούς του Ερωτήματος IV (η τυχαιότητα έχει σκοπίμως διατηρηθεί ώστε να επιβεβαιώνεται η συχνότητα εμφάνισης του ζητούμενου αποτελέσματος).
  - παραλαμβάνει τα αποτελέσματα και υπολογίζει κι ορίζει τις ζητούμενες συνολικές ακρίβειες και
  - τα παρουσιάζει κατάλληλα.

- Επιπλέον συνάρτηση υπολογισμού **NNMulticlass** που
  - αφαιρεί την τυχαιότητα των αποτελεσμάτων ώστε να μπορεί να ελεγχθεί η ορθότητα του κώδικα
  - δημιουργεί το μονοπάτι αποθήκευσης των αρχείων
  - κατασκευάζει το Νευρωνικό Δίκτυο κατά τις απαιτήσεις του κάθε ερωτήματος 1 έως 3
  - εκτελεί τη διαμέριση του συνόλου εισόδου κατά k-Fold Cross-validation για k=10
  - για κάθε Fold εκπαιδεύει το νευρωνικό δίκτυο
  - καλεί την **plotconfusion** για την αποθήκευση των διάφορων ακριβειών σε ανάλογο φάκελο και
  - επιστρέφει τις ακρίβειες ανά Fold στην **Answer.**

- Επιπλέον συνάρτηση υπολογισμού **ClassifierTemplateLinearLogLas** που
  - αφαιρεί την τυχαιότητα των αποτελεσμάτων ώστε να μπορεί να ελεγχθεί η ορθότητα του κώδικα
  - δημιουργεί το μονοπάτι αποθήκευσης των αρχείων
  - κατασκευάζει το γραμμικό ταξινομητή κατά τις απαιτήσεις του ερωτήματος 1
  - εκτελεί τη διαμέριση του συνόλου εισόδου κατά k-Fold Cross-validation για k=10
  - για κάθε Fold εκπαιδεύει το γραμμικό ταξινομητή
  - καλεί την **confusionchart** για την αποθήκευση των διάφορων ακριβειών σε ανάλογο φάκελο και
  - επιστρέφει τις ακρίβειες ανά Fold στην **Answer.**

Task I

Το ερώτημα απαιτεί να κατασκευαστεί νευρωνικό δίκτυο για κάθε μία από τις στοιχηματικές εταιρείες, το οποίο θα λαμβάνει ως είσοδο τις αποδόσεις κάθε αποτελέσματος [προγνωστικά] ανά αγώνα, θα εκπαιδεύει τον εαυτό του μέσω γραμμικής συνάρτησης και θα εξάγει πρόβλεψη για το αποτέλεσμα. Η ταξινομητική του ακρίβεια θα αξιολογείται μέσω k-Fold Cross-validation και θα καταγράφεται για τις διάφορες φάσεις της εκπαίδευσής του ενώ θα ορίζεται η γενική ακρίβεια ανά εταιρεία. Η απάντηση του ερωτήματος προκύπτει από τη σύγκριση των γενικών ακριβειών των εταιρειών.

Η απάντηση υλοποιείται με την κατασκευή νευρωνικού δικτύου τύπου patternet με γραμμική συνάρτηση μεταφοράς που ταξινομεί σε πολλαπλές κλάσεις. Η αξιολόγηση θα πραγματοποιείται μέσω 10-Fold Cross-validation και ανά Fold καταγράφονται οι ακρίβειες κατά τις φάσεις εκπαίδευσης, ελέγχου και επικύρωσης καθώς και η συνολική ακρίβεια. Για ακρίβεια ανά εταιρεία επιλέγουμε τη μέγιστη γενική ακρίβεια από όλα τα Folds.

Προς χάριν πληρότητας της απάντησης, επίσης κατασκευάσαμε έναν γραμμικό ταξινομητή τύπου templateLinear, ο οποίος μπορεί να δέχεται δεδομένα πολλαπλών διαστάσεων και να ταξινομεί σε πολλαπλές κλάσεις. Μαθαίνει μέσω λογιστικής παλινδρόμησης με ποινή πολυπλοκότητας τύπου ridge : $\frac{\lambda}{2}\sum_{j=1}^{p}\beta_j^2$.

Τα αποτελέσματα του υπολογισμού αναγνωρίζουν ως ακριβέστερες εταιρείες,
από τον γραμμικό ταξινομητή, τις **BW** και **LB** με ακρίβεια 54,54%,
από το απλό γραμμικό νευρωνικό δίκτυο, την **BW** με ακρίβεια 34,42%, ενώ
από μονονευρωνικό γραμμικό νευρωνικό δίκτυο, την **IW** με ακρίβεια 46.95%,
από μονοστρωματικό γραμμικό νευρωνικό δίκτυο, όλες με ακρίβεια 47% και
από πολυστρωματικό γραμμικό νευρωνικό δίκτυο, όλες με ακρίβεια 47.22%.

Task II

Το ερώτημα απαιτεί να κατασκευαστεί πολυστρωματικό νευρωνικό δίκτυο για κάθε μία από τις στοιχηματικές εταιρείες, το οποίο θα λαμβάνει ως είσοδο τις αποδόσεις κάθε αποτελέσματος [προγνωστικά] ανά αγώνα, θα εκπαιδεύει τον εαυτό του και θα εξάγει πρόβλεψη για το αποτέλεσμα. Η ταξινομητική του ακρίβεια θα αξιολογείται μέσω k-Fold Cross-validation και θα καταγράφεται για τις διάφορες φάσεις της εκπαίδευσής του ενώ θα ορίζεται η γενική ακρίβεια ανά εταιρεία. Η απάντηση του ερωτήματος προκύπτει από τη σύγκριση των γενικών ακριβειών των εταιρειών.

Η απάντηση υλοποιείται με την κατασκευή πολυστρωματικού νευρωνικού δικτύου τύπου patternet με συνάρτηση μεταφοράς την Scaled Conjugate Gradient που ταξινομεί σε πολλαπλές κλάσεις. Η αξιολόγηση θα πραγματοποιείται μέσω 10-Fold Cross-validation και ανά Fold καταγράφονται οι ακρίβειες κατά τις φάσεις εκπαίδευσης, ελέγχου και επικύρωσης καθώς και η συνολική ακρίβεια. Για ακρίβεια ανά εταιρεία επιλέγουμε τη μέγιστη γενική ακρίβεια από όλα τα Folds.

Τα αποτελέσματα του υπολογισμού αναγνωρίζουν ως ακριβέστερες εταιρείες,
από το πολυστρωματικό νευρωνικό δίκτυο, την **LB** με ακρίβεια 54,16%.

Πίσω στην κορυφή

Return To Contents

Task III

Το ερώτημα απαιτεί να κατασκευαστεί πολυστρωματικό νευρωνικό δίκτυο, το οποίο θα λαμβάνει ως είσοδο συγκεκριμένα χαρακτηριστικά της κάθε ομάδας καθώς και τις αποδόσεις κάθε αποτελέσματος [προγνωστικά] από όλες τις σχετικές στοιχηματικές εταιρείες ανά αγώνα, θα εκπαιδεύει τον εαυτό του μέσω γραμμικής συνάρτησης και θα εξάγει πρόβλεψη για το αποτέλεσμα. Η ταξινομητική του ακρίβεια θα αξιολογείται μέσω k-Fold Cross-validation και θα καταγράφεται για τις διάφορες φάσεις της εκπαίδευσής του ενώ θα ορίζεται η γενική ακρίβεια. Η απάντηση του ερωτήματος προκύπτει από την αναφορά της γενικής ακρίβειας του μοντέλου.

Η απάντηση υλοποιείται με την κατασκευή πολυστρωματικού νευρωνικού δικτύου τύπου patternet με συνάρτηση μεταφοράς την Scaled Conjugate Gradient που ταξινομεί σε πολλαπλές κλάσεις. Η αξιολόγηση θα πραγματοποιείται μέσω 10-Fold Cross-validation και ανά Fold καταγράφονται οι ακρίβειες κατά τις φάσεις εκπαίδευσης, ελέγχου και επικύρωσης καθώς και η συνολική ακρίβεια. Για ακρίβεια ανά εταιρεία επιλέγουμε τη μέγιστη γενική ακρίβεια από όλα τα Folds.

Τα αποτελέσματα του υπολογισμού αναγνωρίζει ως ακρίβεια του εκπαιδευμένου νευρωνικού δικτύου 54,89%.

Task IV

Το ερώτημα απαιτεί να εκτελεστεί ασαφής συσταδοποίηση στο σύνολο των προγνωστικών ανά στοιχηματική εταιρεία και να καταγραφεί το πλήθος των αποτελεσμάτων ανά συστάδα. Η απάντηση του ερωτήματος προέρχεται από τη σύγκριση των πληθών αυτών.

Η απάντηση υλοποιείται με την εφαρμογή της συνάρτησης **fcm,** η οποία κάνει ασαφή c-means συσταδοποίηση, στο σύνολο των προγνωστικών για την κάθε εταιρεία. Έπειτα, ορίζουμε σαφή ομαδοποίηση ανά cluster ανάλογα με τη συμμετοχή του κάθε προγνωστικού σε κάθε κέντρο (centroid) όπως επιστρέφεται από την **fcm** και μετράμε το πλήθος κάθε ταμπέλας-αποτελέσματος σε κάθε cluster.

Ο υπολογισμός αναγνωρίζει ως εμφανιζόμενα με μεγαλύτερη συχνότητα τα αποτελέσματα «1» ή «2» εντός κάθε συστάδας και πάντα το αποτέλεσμα «1» συνολικά.

Return to Page Top

Return To Index

Code Execution Screenshots

Preparing the data for the 1st, 2nd and 4th Tasks

```
Command Window

Retrieving relevant data from the database...

Establishing connection to database.sqlite

Closed connection to database : database.sqlite

Clearing workspace of the unneeded variables...

Response column for the predictive models has been built.

Building the various sets needed for the training of the models...

Relevant files have been saved. Data preparation for answering the 1st, 2nd and 3rd Question has finished. This Section has completed successfully.

fx>>
```

Preparing the data for the 3rd Task

```
Command Window

Clearing Workspace and loading required data...

Checking information matching between the various used data...

For the given material, there are only 23 matches, that is 0.10% of the entire database, in which their team's attributes were recorded.

We now process the needed information regarding the attributes of the teams...

Relevant Table has been created successfully.

Retrieveing the corresponding data from the rest of the tables...

Vectors F(H) and F(A) have been calculated.

Relevant files have been saved. Data preparation for answering the 3rd Question has finished. This Section has completed successfully.

fx>>
```

Answering Task I

```
Command Window

  Question 1
For the given sample, under Linear Classifier proccessing, BW reaches best accuracy, at 54.54% :

Answer =

  4×12 table

    Evaluated Set    Estimated Maximum Accuracy %    Fold no1 Accuracy %    Fold no2 Accuracy %    Fold no3 Accuracy %    Fold no4 Accuracy %    Fold no5 Accu
    _____________    ____________________________    ___________________    ___________________    ___________________    ___________________    ____________

       "B365"                "54.32"                      "52.63"                "53.4"                 "53.14"                "52.47"                "53.72"
       "BW"                  "54.54"                      "52.45"                "52.96"                "53"                   "52.25"                "53.45"
       "IW"                  "54.5"                       "52.4"                 "53.09"                "52.87"                "52.16"                "53.72"
       "LB"                  "54.54"                      "52.27"                "53.14"                "52.87"                "52.29"                "53.72"


For the given sample, under Neural Network proccessing, BW reaches best accuracy, at 34.42% :

Answer =

  4×12 table

    Evaluated Set    Estimated Maximum Accuracy %    Fold no1 Accuracy %    Fold no2 Accuracy %    Fold no3 Accuracy %    Fold no4 Accuracy %    Fold no5 Accu
    _____________    ____________________________    ___________________    ___________________    ___________________    ___________________    ____________

       "B365"                "31.52"                      "31.52"                "29.02"                "27.24"                "26.61"                "25.86"
       "BW"                  "34.42"                      "34.42"                "31.24"                "30.31"                "28.62"                "26.03"
       "IW"                  "25.68"                      "25.29"                "24.25"                "23.99"                "25.1"                 "25.32"
       "LB"                  "33.53"                      "33.53"                "32.71"                "30.44"                "30.89"                "28.48"

fx>>
```

Return to Page Top

Πίσω στο Ευρετήριο

Answering Task II

Command Window

```
 Question 2
For the given sample, under Neural Network proccessing, LB reaches best accuracy, at 54.16% :

Answer =

  4×12 table
```

Evaluated Set	Estimated Maximum Accuracy %	Fold no1 Accuracy %	Fold no2 Accuracy %	Fold no3 Accuracy %	Fold no4 Accuracy %	Fold no5 Accu
"B365"	"54.12"	"54.01"	"53.14"	"52.02"	"51.76"	"53.76"
"BW"	"54.03"	"53.52"	"52.83"	"51.89"	"51.36"	"53.49"
"IW"	"54.01"	"53.52"	"52.91"	"52.02"	"51.67"	"53.54"
"LB"	"54.16"	"53.87"	"53.05"	"52.43"	"51.8"	"53.23"

```
fx>>
```

Answering Task III

Command Window

```
 Question 3
For the given sample, under Neural Network proccessing, MultiFeatures reaches best accuracy, at 54.89% :

Answer =

  1×12 table
```

Evaluated Set	Estimated Maximum Accuracy %	Fold no1 Accuracy %	Fold no2 Accuracy %	Fold no3 Accuracy %	Fold no4 Accuracy %	Fold no5 Ac
"MultiFeatures"	"54.89"	"50.23"	"54.08"	"50.7"	"54.89"	"53.4

```
fx>>
```

Πίσω στην κορυφή Συστήματος

Πίσω στο Ευρετήριο

Answering Task IV

Command Window

```
 Question 4
The clusters for each company are as follows :

Answer =

  4×4 table

    Evaluated Sets    Cluster1    Cluster2    Cluster3
    ______________    ________    ________    ________

        "B365"          "1"         "2"         "2"
        "BW"            "2"         "2"         "1"
        "IW"            "1"         "3"         "2"
        "LB"            "1"         "2"         "2"


Within cluster C1 prevail(s) the result(s) of 1
Within cluster C2 prevail(s) the result(s) of 2
Within cluster C3 prevail(s) the result(s) of 2
Within the entire cluster set, prevail(s) the result(s) of 2 >>
```

Command Window

```
 Question 4
The clusters for each company are as follows :

Answer =

  4×4 table

    Evaluated Sets    Cluster1    Cluster2    Cluster3
    ______________    ________    ________    ________

        "B365"          "2"         "2"         "1"
        "BW"            "2"         "1"         "2"
        "IW"            "1"         "3"         "2"
        "LB"            "1"         "2"         "2"


Within cluster C1 prevail(s) the result(s) of 1 2
Within cluster C2 prevail(s) the result(s) of 2
Within cluster C3 prevail(s) the result(s) of 2
Within the entire cluster set, prevail(s) the result(s) of 2 >>
```

Command Window

```
 Question 4
The clusters for each company are as follows :

Answer =

  4×4 table

    Evaluated Sets    Cluster1    Cluster2    Cluster3
    ______________    ________    ________    ________

        "B365"          "1"         "2"         "2"
        "BW"            "1"         "2"         "2"
        "IW"            "1"         "2"         "3"
        "LB"            "2"         "2"         "1"


Within cluster C1 prevail(s) the result(s) of 1
Within cluster C2 prevail(s) the result(s) of 2
Within cluster C3 prevail(s) the result(s) of 2
Within the entire cluster set, prevail(s) the result(s) of 2 >>
```

Command Window

```
 Question 4
The clusters for each company are as follows :

Answer =

  4×4 table

    Evaluated Sets    Cluster1    Cluster2    Cluster3
    ______________    ________    ________    ________

        "B365"          "1"         "2"         "2"
        "BW"            "2"         "1"         "2"
        "IW"            "3"         "1"         "2"
        "LB"            "2"         "2"         "1"


Within cluster C1 prevail(s) the result(s) of 2
Within cluster C2 prevail(s) the result(s) of 1 2
Within cluster C3 prevail(s) the result(s) of 2
Within the entire cluster set, prevail(s) the result(s) of 2 >>
```

Πίσω στην κορυφή Συστήματ

Return To Index