



Ταξινόμηση

Ιωάννης Μαδεμλής

Προβλεπτική μοντελοποίηση

- Μία από τις βασικές κατηγορίες προβλημάτων της εξόρυξης γνώσης είναι η **προβλεπτική μοντελοποίηση**, με δύο υποπροβλήματα:
 - *Ταξινόμηση* (classification).
 - *Παλινδρόμηση* (regression).
- Πρόκειται επομένως για τους δύο βασικότερους τύπους προβλημάτων **επιβλεπόμενης μάθησης**.
 - Συνήθως, στην εξόρυξη γνώσης μας ενδιαφέρει μόνον η ταξινόμηση των προτύπων ενός συνόλου δεδομένων.
- Ταξινόμηση είναι η εργασία ανάθεσης στοιχείων/προτύπων σε μία από K προκαθορισμένες **κλάσεις**.

Ταξινόμηση

Ταξινομητές

- Τα στοιχεία είναι m -διάστατα διανύσματα γνωρισμάτων.
- Σε κάθε στοιχείο/πρότυπο αντιστοιχεί ένα επιπρόσθετο γνώρισμα-στόχος το οποίο είναι **διακριτό** και δηλώνει την κλάση (π.χ., ακέραιος στο $[1, K]$).
- Επομένως η ταξινόμηση είναι η μάθηση μίας συνάρτησης που απεικονίζει ένα σύνολο γνωρισμάτων με δεδομένες τιμές σε μία κλάση.
 - *Συνάρτηση*: **μοντέλο ταξινόμησης**.
 - Μηχανισμός αλγόριθμος ο οποίος την υλοποιεί: **ταξινομητής**.

Ταξινόμηση

Εκπαίδευση

- Ο ταξινομητής είναι μία συστηματική προσέγγιση κατασκευής του μοντέλου ταξινόμησης με βάση ένα γνωστό σύνολο προτύπων εισόδου γνωστής κλάσης, ονόματι **σύνολο εκπαίδευσης**.
- Δοθέντος ενός ταξινομητή και ενός συνόλου εκπαίδευσης, η διαδικασία μάθησης / κατασκευής του μοντέλου ταξινόμησης λέγεται **εκπαίδευση**.
- Ο εκπαιδευμένος ταξινομητής είναι πλέον σε θέση να ταξινομεί πρότυπα από το **σύνολο ελέγχου**, μία συλλογή διανυσμάτων/προτύπων άγνωστης κλάσης.
 - Προβλέπει την κλάση κάθε προτύπου, μέσω του μοντέλου ταξινόμησης το οποίο κωδικοποιήθηκε στη δομή του κατά την εκπαίδευση.
 - Η εκπαίδευση συνίσταται στην αυτόματη εύρεση των βέλτιστων αριθμητικών τιμών των *παραμέτρων* οι οποίες ορίζουν την ακριβή μορφή ενός μοντέλου ταξινόμησης, μέσω ενός αλγορίθμου εκπαίδευσης.

Ταξινόμηση

Γενίκευση

- Άρα ένας ταξινομητής πρέπει όχι μόνο να ταξινομεί ορθώς τα στοιχεία με τα οποία εκπαιδεύτηκε, αλλά και να προβλέπει με επαρκή ακρίβεια την κλάση νέων προτύπων ελέγχου.
 - Να έχει καλή **γενικευτική ικανότητα**.
- Η ποιότητα ενός εκπαιδευμένου ταξινομητή αξιολογείται με βάση την **ακρίβειά** του σε ένα σύνολο ελέγχου (accuracy):
 - τον λόγο των ορθών προβλέψεων προς το ολικό πλήθος προβλέψεων (Correct Classification Rate), ή
 - το **σφάλμα ταξινόμησης**, δηλαδή τον λόγο των λανθασμένων προβλέψεων προς το ολικό πλήθος προβλέψεων.

Ταξινόμηση

Ετικέτες κλάσης

- Άρα για την αξιολόγηση ενός ταξινομητή απαιτείται να είναι γνωστές οι κλάσεις των *προτύπων ελέγχου*.
 - Απλώς, αυτές οι **ετικέτες κλάσης** δεν αξιοποιούνται κατά την εκπαίδευση.
 - Μόνον οι ετικέτες κλάσεις των *προτύπων εκπαίδευσης* χρησιμοποιούνται από τον αλγόριθμο εκπαίδευσης, ο οποίος κατασκευάζει το μοντέλο ταξινόμησης.
 - **Επιβλεπόμενη μάθηση.**
- Για να δημιουργηθεί ένα επιτυχημένο μοντέλο ταξινόμησης υψηλής ακρίβειας, πρέπει να υπάρχει πράγματι κάποια συσχέτιση μεταξύ:
 - των γνωρισμάτων ενός προτύπου εισόδου, και
 - της κλάσης του,
- Διαφορετικά το μοντέλο θα έχει πολύ μικρή γενικευτική ικανότητα.

Ταξινόμηση

Μεροληψία και διακύμανση

- Σε όλους τους παραμετρικούς ταξινομητές εμφανίζεται το δίλημμα **μεροληψίας-διακύμανσης** κατά την κατασκευή τους:
 - Η μεροληψία είναι ανάλογη με την «απλότητά» τους, δηλαδή τη δόμηση τους από ένα μικρό, άκαμπτο πλήθος «θεμέλιων λίθων» και παραμέτρων.
 - Η διακύμανση είναι ανάλογη με την «πολυπλοκότητά» τους, δηλαδή τη δόμηση τους από ένα μεγάλο, ευέλικτο πλήθος «θεμέλιων λίθων» και παραμέτρων.
- Ένας ταξινομητής υψηλής διακύμανσης υποφέρει από μικρή γενικευτική ικανότητα.
 - Υπερεκπαίδευση.
- Ένας ταξινομητής υψηλής μεροληψίας δεν καταφέρνει να κωδικοποιήσει στη δομή του το ορθό μοντέλο ταξινόμησης κατά την εκπαίδευσή του.
 - Υποεκπαίδευση.

Ταξινόμηση

Μεροληψία και διακύμανση

- Η διακύμανση και η μεροληψία είναι ιδιότητες του ταξινομητή αντιστρόφως ανάλογες μεταξύ τους.
 - Υψηλή μεροληψία – υπερβολικά λίγες παράμετροι για δοθέν σύνολο εκπαίδευσης.
 - Υψηλή διακύμανση – υπερβολικά πολλές παράμετροι για δοθέν σύνολο εκπαίδευσης.
- Κατά την οικοδόμηση του μοντέλου ταξινόμησης πρέπει να βρεθεί ένας βέλτιστος συνδυασμός μεροληψίας και διακύμανσης.
 - Μία χρυσή τομή στο πλήθος παραμέτρων οι οποίες ορίζουν τη δομή του, ώστε ο ταξινομητής να είναι σε θέση να κατηγοριοποιεί άγνωστα στοιχεία (στα οποία δεν εκπαιδεύτηκε) με επαρκή ακρίβεια.

Ταξινόμηση

Μεροληψία και διακύμανση

- Λύση: Συνηθίζεται να κατασκευάζεται πρώτα ένας πολύπλοκος ταξινομητής υψηλής διακύμανσης και ακολούθως να «απλοποιείται».
- Η διαδικασία απλοποίησης εξαρτάται από τον τύπο και τον αλγόριθμο παρασκευής του αρχικού μοντέλου ταξινόμησης.
- Στόχος της είναι να καταλήξει σε έναν τελικό ταξινομητή με καλή γενικευτική ικανότητα.
- Εναλλακτικά: Εκκινώντας από το απλούστερο δυνατό μοντέλο, η διαδικασία κατασκευής ενός πιο πολύπλοκου ταξινομητή σταματά **πρόωρα**.
 - Μόλις ο ρυθμός αύξησης της γενικευτικής ικανότητας του τρέχοντος μοντέλου αρχίζει να μειώνεται.

Ταξινόμηση

Υπερεκπαίδευση

- Αν δεν συνυπολογιστεί η διακύμανση κατά την κατασκευή / εκπαίδευση του ταξινομητή, τότε το τελικό μοντέλο ταξινόμησης πιθανόν να πάσχει από υπερεκπαίδευση.
- Μία κατάσταση όπου:
 - έχει «απομνημονεύσει» στην εσωτερική του δομή τα στοιχεία του συνόλου εκπαίδευσης και είναι σε θέση να τα ταξινομεί άριστα...
 - ...χωρίς όμως να έχει πραγματικά οικοδομήσει από αυτά ένα γενικό μοντέλο ταξινόμησης, το οποίο θα του επέτρεπε να ταξινομεί και τα άγνωστα στοιχεία του συνόλου ελέγχου.
- Ένα υπερεκπαιδευμένο μοντέλο έχει μάθει τον **θόρυβο** (τις ιδιαιτερότητες) του συνόλου εκπαίδευσης, αντί για τη χρήσιμη γνώση.
 - Π.χ., εσφαλμένες ετικέτες κλάσης, σύνολο εκπαίδευσης μη αντιπροσωπευτικό του προβλήματος, κλπ.

Ταξινόμηση

Σύνορα απόφασης

- Θεωρούμε τον διανυσματικό χώρο των m -διάστατων προτύπων εκπαίδευσης.
 - Κάθε γνώρισμα είναι μία διάσταση.
 - Για $m = 2, 3$, κάθε πρότυπο μπορεί να εικονιστεί στη γραφική αναπαράσταση του χώρου αυτού ως σημείο.
- Πότε υπάρχει πράγματι κάποιο εξαγώγιμο μοντέλο ταξινόμησης το οποίο μπορούμε να μάθουμε;
 - Όταν πράγματι οι τιμές κάποιων γνωρισμάτων καθορίζουν στην πλειονότητα των περιπτώσεων την κλάση του εκάστοτε προτύπου.
- Τότε μία υπερεπιφάνεια (καρτεσιανή καμπύλη αν $m = 2$), ονόματι **σύνορο απόφασης**, διαχωρίζει τα περισσότερα στοιχεία/πρότυπα μίας κλάσης από τα περισσότερα στοιχεία μίας άλλης.

Ταξινόμηση

Σύνορα απόφασης

- Τα σύνορα απόφασης είναι δυνατόν να τέμνονται και να σχηματίζουν **περιοχές απόφασης** μες στον m -διάστατο διανυσματικό χώρο των προτύπων.
- Κατά την εκπαίδευση ενός ταξινομητή, το υπό κατασκευή μοντέλο ταξινόμησης μαθαίνει τις εν λόγω περιοχές απόφασης.
 - Πώς; Διά της ανάθεσης κατάλληλων τιμών στις παραμέτρους που ορίζουν τη δομή του, ώστε να κωδικοποιηθούν ορθώς στο εσωτερικό του τα υπάρχοντα όρια απόφασης.
- Ένα καλό μοντέλο όμως πρέπει να γενικεύει αυτά τα όρια απόφασης και σε περιοχές του m -διάστατου χώρου οι οποίες δεν καλύπτονται από το σύνολο εκπαίδευσης.

Ταξινόμηση

Σύνορα απόφασης

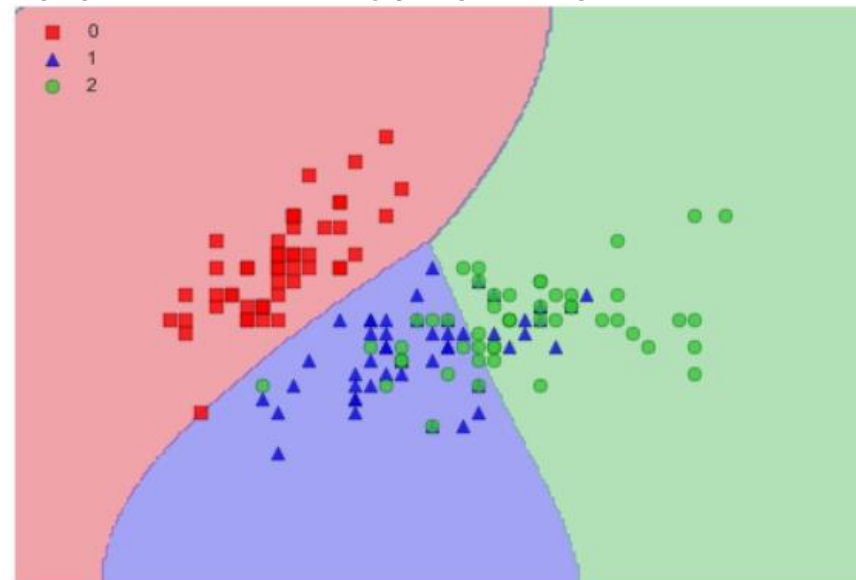
- Τα σύνορα απόφασης στη γενική περίπτωση είναι υπερεπιφάνειες στον διανυσματικό χώρο \mathbb{R}^m .
 - Αν $m = 2$, τότε ο χώρος των προτύπων είναι το ευκλείδειο επίπεδο και τα σύνορα απόφασης είναι απλές καμπύλες.
 - Αν τα πρότυπα δύο κλάσεων είναι επαρκώς απομακρυσμένα μεταξύ τους στον χώρο \mathbb{R}^m , τότε οι δύο κλάσεις μπορούν να διαχωριστούν από ένα απλό υπερεπίπεδο.
 - Ευθεία, αν $m = 2$.
 - Δύο κλάσεις τα πρότυπα των οποίων μπορούν να διαχωριστούν πλήρως μεταξύ τους από ένα υπερεπίπεδο απόφασης καλούνται **γραμμικώς διαχωρίσιμες**.
 - Ένας ταξινομητής ο οποίος μαθαίνει ως μοντέλο ταξινόμησης υπερεπίπεδα απόφασης, καλείται **γραμμικός ταξινομητής**.
 - Λειτουργεί καλά μόνο για γραμμικώς διαχωρίσιμες κλάσεις.

Ταξινόμηση

Σύνορα απόφασης

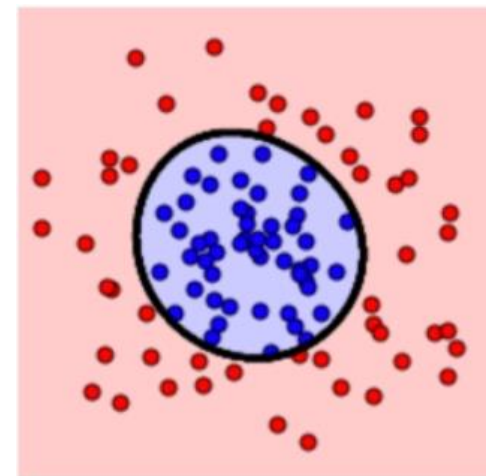
- $m = 2, k = 3$, μη γραμμικώς διαχωρίσιμες κλάσεις:

Μετά την εκπαίδευση, ο ταξινομητής έχει μάθει ατελείς καμπύλες απόφασης – δύσκολο πρόβλημα ταξινόμησης.



- $m = 2, k = 2$, ισχυρά μη γραμμικώς διαχωρίσιμες κλάσεις:

Δύσκολο πρόβλημα ταξινόμησης – ένας καλός ταξινομητής πρέπει να μάθει σχεδόν κυκλική καμπύλη απόφασης.

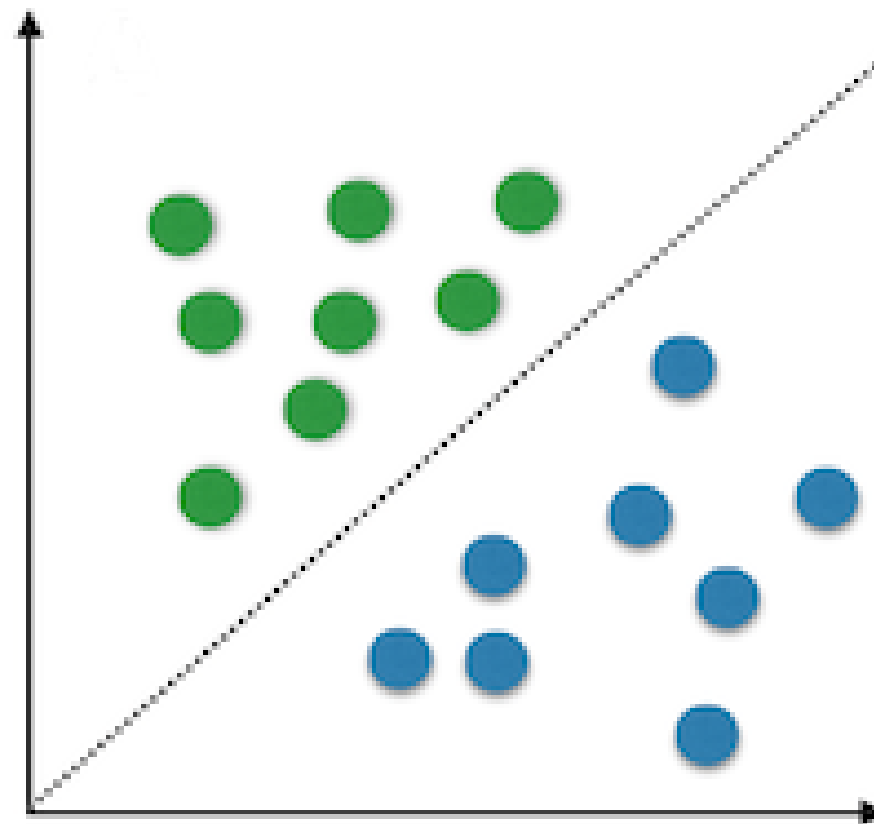


Ταξινόμηση

Σύνορα απόφασης

- $m = 2, k = 2$, γραμμικώς διαχωρίσιμες κλάσεις:

Κατά την εκπαίδευση, ο ταξινομητής πρέπει να μάθει μία ευθεία απόφασης η οποία διαχωρίζει πλήρως τις δύο κλάσεις – πολύ εύκολο πρόβλημα ταξινόμησης.



Ταξινόμηση

Σύνορα απόφασης

- Τι συμβαίνει κατά τη λειτουργία του εκπαιδευμένου ταξινομητή με άγνωστα πρότυπα ελέγχου;
 - Ουσιαστικά αναζητείται η περιοχή απόφασης όπου τοποθετούνται μες στον m -διάστατο χώρο και τους ανατίθεται η κλάση που αντιστοιχεί στην περιοχή αυτή, σύμφωνα με το μοντέλο ταξινόμησης που ανακαλύφθηκε.
- Η μεροληψία του ταξινομητή εκφράζει τη μέση απόσταση των συνόρων απόφασής του από τα πραγματικά σύνορα απόφασης στο σύνολο εκπαίδευσης.
 - Ένας υπερβολικά απλοϊκός ταξινομητής, άκαμπτος και με λίγες παραμέτρους, αδυνατεί να κωδικοποιήσει στη δομή του πολύπλοκα σύνορα απόφασης και παρουσιάζει μεγάλη μεροληψία.

Ταξινόμηση

Σύνορα απόφασης

- Η διακύμανση εκφράζει την εξάρτηση των προβλέψεων του ταξινομητή από μικρές διακυμάνσεις της εισόδου.
 - Αλλιώς την αστάθειά του.
- Σε μία εκπαίδευση υψηλής διακύμανσης, διαφορετικά σύνολα εκπαίδευσης για το ίδιο πρόβλημα μπορεί να οδηγούν σε πολύ διαφορετικά σύνορα απόφασης.
 - Αν ο ταξινομητής είναι υπερβολικά ευέλικτος και πολύπλοκος είναι δυνατόν να αποθηκεύσει τις «ιδιοτροπίες» του εκάστοτε συνόλου εκπαίδευσης στη δομή του.
 - Ενώ θα λειτουργεί πολύ επιτυχημένα για τα πρότυπα με τα οποία εκπαιδεύτηκε, θα παρουσιάζει πολύ χαμηλή γενικευτική ικανότητα.

Ταξινόμηση

Σφάλμα γενίκευσης

- Τελικό σφάλμα γενίκευσης.

- Άθροισμα της μεροληψίας, της διακύμανσης και του θορύβου.
- Ο θόρυβος εδώ είναι μία εγγενής συνιστώσα τυχαιότητας στο σύνολο δεδομένων, η οποία δηλώνει κάποια έλλειψη συσχέτισης μεταξύ των τιμών των γνωρισμάτων και της κλάσης ενός προτύπου.

Ταξινόμηση

- Πώς μετρούμε το σφάλμα γενίκευσης; Με προσεγγιστική εκτίμηση.
 - Μετά την εκπαίδευση, μετρούμε την ακρίβεια στο σύνολο ελέγχου.
 - Αντιθέτως, η ακρίβεια στο σύνολο εκπαίδευσης έχει νόημα μόνο κατά τη διάρκεια της εκπαίδευσης.

Σύνολο δεδομένων

- Τυπικό πρωτόκολλο εκπαίδευσης και αξιολόγησης ταξινομητών.
- Δοθέντος ενός συνόλου δεδομένων (N πρότυπα και αντίστοιχες N ετικέτες κλάσεις τους) επαρκώς μεγάλου και αντιπροσωπευτικού του προβλήματος...
- ...το διασπούμε μέσω *τυχαίας δειγματοληψίας* σε τρία ανισομεγέθη υποσύνολα:
 - **Σύνολο εκπαίδευσης** (training set),
 - **Σύνολο επικύρωσης** (validation set),
 - **Σύνολο ελέγχου** (test set).

Επικύρωση

Σύνολο δεδομένων

- Η ισορροπία μεταξύ διακύμανσης και μεροληψίας εξαρτάται από ένα σύνολο επιλεγμένων τιμών **υπερπαραμέτρων**, οι οποίες καθορίζουν την πολυπλοκότητα του ταξινομητή και τον βαθμό προσαρμογής του στο εκάστοτε σύνολο εκπαίδευσης.
- Οι βέλτιστες τιμές υπερπαραμέτρων είναι αυτές οι οποίες δίνουν τη μέγιστη γενικευτική ικανότητα σε άγνωστα δεδομένα.
- Όμως οι τιμές των υπερπαραμέτρων προσδιορίζονται *χειροκίνητα από εμάς πριν από την εκπαίδευση* και δεν εντοπίζονται αυτομάτως από τον αλγόριθμο εκπαίδευσης.
 - Κατ' αντιδιαστολή με τις τιμές παραμέτρων του μοντέλου, οι οποίες κωδικοποιούν την εξαχθείσα γνώση περί περιοχών απόφασης.

Επικύρωση

Σύνολο δεδομένων

- Το σφάλμα ταξινόμησης του μοντέλου στα στοιχεία του συνόλου εκπαίδευσης ονομάζεται **σφάλμα εκπαίδευσης**.
 - Λειτουργεί ως οδηγός του αλγορίθμου εκπαίδευσης κατά τη διάρκεια της εκπαίδευσης του ταξινομητή.
- Το σφάλμα ταξινόμησης στα στοιχεία του συνόλου επικύρωσης λέγεται **σφάλμα επικύρωσης**.
 - Λειτουργεί ως οδηγός κατά τη διαδικασία επιλογής βέλτιστων τιμών υπερπαραμέτρων.
- Το σφάλμα ταξινόμησης του τελικού ταξινομητή στα στοιχεία του συνόλου ελέγχου καλείται **σφάλμα γενίκευσης**.
 - Χαρακτηρίζει την ποιότητα του τελικού κατασκευασθέντος μοντέλου ταξινόμησης.

Επικύρωση

Σύνολο δεδομένων

- Σύνοψη: εκπαιδεύουμε στο σύνολο εκπαίδευσης πολλαπλά διαφορετικά μοντέλα με διαφορετικές τιμές υπερπαραμέτρων και στο τέλος επιλέγουμε όποιο δίνει το ελάχιστο σφάλμα επικύρωσης.
- Δοκιμάζουμε το επιλεγμένο βέλτιστο μοντέλο στο σύνολο ελέγχου και καταγράφουμε το σφάλμα γενίκευσης.
- Στη συνέχεια μπορούμε να χρησιμοποιήσουμε το μοντέλο σε νέα άγνωστα δεδομένα για τα οποία δεν διαθέτουμε καθόλου ετικέτες κλάσεις, έχοντας μία καλή εκτίμηση της γενικευτικής του ικανότητας.

Επικύρωση

Σύνολο δεδομένων

- Γιατί δεν μπορούμε να επιλέξουμε βέλτιστες τιμές υπερπαραμέτρων με βάση το σφάλμα γενίκευσης στο σύνολο ελέγχου;
 - Θα υπερεκτιμούσαμε τη γενικευτική ικανότητα του επιλεγμένου μοντέλου σε άγνωστα δεδομένα.
- Γιατί δεν μπορούμε να επιλέξουμε βέλτιστες τιμές υπερπαραμέτρων με βάση το σφάλμα εκπαίδευσης στο σύνολο εκπαίδευσης;
 - Θα είχαμε επιλέξει στο τέλος ένα μοντέλο υπερεκπαιδευμένο στο σύνολο εκπαίδευσης.
- Γι' αυτό χρειαζόμαστε ξεχωριστό σύνολο επικύρωσης και υπολογισμό ξεχωριστού σφάλματος επικύρωσης.

Επικύρωση

Σύνολο δεδομένων

- Κατά κανόνα το σύνολο εκπαίδευσης είναι μεγάλο, ενώ τα σύνολα επικύρωσης και ελέγχου είναι μικρά.
 - Π.χ., ποσοστιαία διαμέριση του αρχικού συνόλου δεδομένων σε τρία υποσύνολα με αναλογία 60%/20%/20% ή 70%/15%/15%.
- Όμως η διάσπαση σε **σταθερά** υποσύνολα εκπαίδευσης και επικύρωσης μέσω τυχαίας δειγματοληψίας ενδέχεται να οδηγήσει σε στατιστική μεροληψία δειγματοληψίας.
 - Π.χ., τυχαίνει όλα τα πρότυπα μίας κλάσης να είναι μόνο στο σύνολο εκπαίδευσης ή μόνο στο σύνολο επικύρωσης.

Επικύρωση

Διασταυρούμενη επικύρωση

- Λύση: διασταυρούμενη επικύρωση (cross-validation).
- Στοχεύει στην αποφυγή της μεροληψίας δειγματοληψίας κατά την επικύρωση.
 - Διαμερίζουμε το αρχικό σύνολο δεδομένων μόνο σε σταθερό σύνολο εκπαίδευσης και σταθερό σύνολο ελέγχου.
 - Το σύνολο εκπαίδευσης διαμερίζεται σε k ξένα υποσύνολα (π.χ., $k = 3, 5, 7, 10$, κλπ.).
 - Κυκλικά χρησιμοποιείται ένα εξ αυτών, **διαφορετικό κάθε φορά**, ως σύνολο επικύρωσης, ενώ τα υπόλοιπα λειτουργούν από κοινού ως σύνολο εκπαίδευσης.
 - Σε κάθε κύκλο κατασκευάζεται ο ταξινομητής εκ του μηδενός.
 - Άρα εν τέλει θα έχουμε k διαφορετικά μοντέλα εκπαιδευμένα σε διαφορετικά δεδομένα.

Επικύρωση

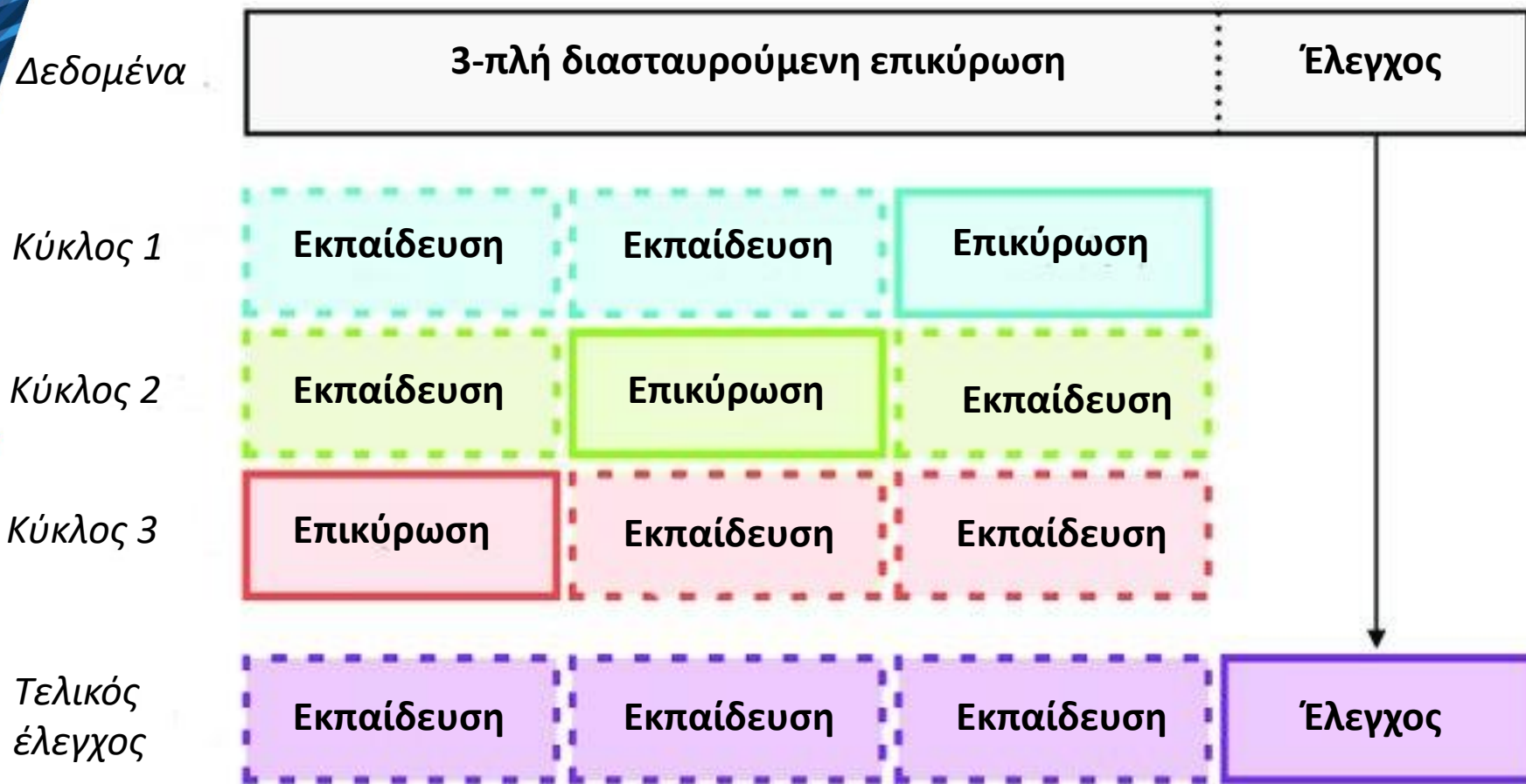
Διασταυρούμενη επικύρωση

- Λύση: διασταυρούμενη επικύρωση (cross-validation).
 - Τελικά υπολογίζεται ο μέσος όρος μ των k σφαλμάτων επικύρωσης που παρουσιάζουν οι k παραχθέντες ταξινομητές.
 - Οι k κύκλοι εκπαίδευσης-επικύρωσης-μέσου όρου επαναλαμβάνονται ξεχωριστά για κάθε διαφορετική εκδοχή του ταξινομητή.
 - Καθεμία από τις l εκδοχές αντιστοιχεί σε διαφορετικές τιμές υπερπαραμέτρων.
 - Άρα, εν τέλει εκπαιδεύονται lk ξεχωριστά μοντέλα ταξινόμησης από τα οποία υπολογίζονται l μέσα σφάλματα επικύρωσης.
 - Εν τέλει, επιλέγουμε την εκδοχή (σύνολο τιμών υπερπαραμέτρων) με το ελάχιστο σφάλμα επικύρωσης και μετρούμε το σφάλμα γενίκευσής της στο σταθερό σύνολο ελέγχου.

Επικύρωση

Διασταυρούμενη επικύρωση

Επικύρωση



Διασταυρούμενη επικύρωση

- Παρατηρήσεις:

- Υψηλό υπολογιστικό κόστος, χρησιμοποιείται συνήθως μόνο για μικρά σύνολα δεδομένων (σχετικά λίγα πρότυπα) όπου είναι περισσότερο εμφανής η μεροληψία δειγματοληψίας.
- Μετά την επιλογή βέλτιστων υπερπαραμέτρων, μπορούμε να εκπαιδεύσουμε από την αρχή το εν λόγω βέλτιστο μοντέλο στο **πλήρες** σύνολο εκπαίδευσης, πριν μετρήσουμε το σφάλμα γενίκευσής του στο σύνολο ελέγχου.
- Εντελώς ανεξάρτητα από την ανωτέρω διαδικασία επικύρωσης, ο μηχανισμός της διασταυρούμενης επικύρωσης μπορεί να αξιοποιηθεί **και** για τη μέτρηση της γενικευτικής ικανότητας!
 - Έτσι αποφεύγεται η μεροληψία διακύμανσης κατά τη διαμέριση μεταξύ συνόλου εκπαίδευσης και συνόλου ελέγχου.
 - Ακόμη μεγαλύτερη υπολογιστική πολυπλοκότητα.

Επικύρωση

Thank you for your attention!

Q & A

Contact: imademlis@aueb.gr