

Αξιολόγηση Ταξινομητών

Μέθοδοι Εκπαίδευσης Ταξινομητών

- **Μέθοδος επανατοποθέτησης** (resubstitution method)
Χρησιμοποίησε τα ίδια δεδομένα για εκπαίδευση και δοκιμή.
Έχουμε υποεκτίμηση του σφάλματος.
Η εκτίμηση βελτιώνεται για μεγάλες τιμές του N.
- **Μέθοδος διαχωρισμού** (holdout method)
Δοθέντος ενός συνόλου N διανυσμάτων εισόδου/εξόδου χώρισέ τα σε N1 διανύσματα εκπαίδευσης και N2 διανύσματα ελέγχου ($N1 + N2 = N$).
Πρόβλημα: Λιγότερα δεδομένα τόσο για εκπαίδευση όσο και για έλεγχο.

Μέθοδοι Εκπαίδευσης Ταξινομητών

- **Μέθοδος Leave-One-Out** (κυρίως για μικρά datasets)
 - ❖ Επίλεξε ένα από τα N δείγματα. Εκπαίδευσε τον ταξινομητή χρησιμοποιώντας τα υπόλοιπα $N-1$ δείγματα. Έλεγε την απόδοση του ταξινομητή στο επιλεγμένο δείγμα. Στην περίπτωση λανθασμένης ταξινόμησης αύξησε τον μετρητή λαθών κατά ένα.
 - ❖ Επανάλαβε το παραπάνω εξαιρώντας ένα διαφορετικό δείγμα κάθε φορά.
 - ❖ Υπολόγισε την πιθανότητα λάθους ως το ποσοστό των λαθών επί του συνόλου των επαναλήψεων.
- ❖ **Πλεονεκτήματα**
 - Χρήση όλων των διαθέσιμων δεδομένων για εκπαίδευση και έλεγχο.
 - Ανεξαρτησία δεδομένων ελέγχου από αυτά της εκπαίδευσης.
- ❖ **Μειονεκτήματα**
 - Υψηλές υπολογιστικές απαιτήσεις.

Μέθοδοι Εκπαίδευσης Ταξινομητών

Παράδειγμα χρήσης δεδομένων τεκμηρίωσης

- **Μέθοδος k -fold cross validation (k -πλής διασταυρούμενης επικύρωσης)**
 - ❖ Χρησιμοποιείται για την ακριβέστερη αποτίμηση της γενικευτικής ικανότητας μοντέλων μηχανικής μάθησης.
 - ❖ Εκτελούμε μια τυχαία μετάθεση όλων των δεδομένων, χωρίζουμε το σύνολο των δεδομένων σε k ίσα μέρη (folds) και δημιουργούμε, για κάθε ένα από k πειράματα, τρία σύνολα δεδομένων με κυκλικό τρόπο ως εξής:
 - Το **σύνολο εκπαίδευσης** (training set) που χρησιμοποιείται για την εύρεση των παραμέτρων του ταξινομητή. Αφήνοντας ένα μέρος για το σύνολο ελέγχου, αυτό τυπικά περιλαμβάνει το 60 – 80% των υπόλοιπων ($k-1$) μερών (ή τα $(k-1)$ μέρη αν δεν υπάρχουν δεδομένα τεκμηρίωσης).
 - Το **σύνολο τεκμηρίωσης** (validation set) που χρησιμοποιείται για την εύρεση βέλτιστων τιμών για τις μεταπαραμέτρους του αλγορίθμου εκπαίδευσης (μπορεί να χρησιμοποιηθεί με οποιαδήποτε μέθοδο). Αυτό περιλαμβάνει τα αδιάθετα των $(k-1)$ μερών.
 - Το **σύνολο ελέγχου** (test set) για την αποτίμηση της γενίκευσης. Αυτό τυπικά περιλαμβάνει ένα μέρος, διαφορετικό για κάθε πείραμα.

Μετρικές αξιολόγησης ταξινομητών

ο **Συνολική ακρίβεια** (Overall Accuracy – OA)

- ❖ Είναι το άθροισμα των διαγώνιων στοιχείων του ΠΣ προς το πλήθος δεδομένων N.

$$OA = \frac{\sum_{i=1}^M A(i,i)}{N}$$

- ❖ Στο παράδειγμα, $OA = \frac{98+135+198+40+150}{700} = 88,7\%$

Παρατήρηση: Η ιδιαιτερότητα της ω_4 (η οποία συγγέεται με την ω_1 και την ω_3) δεν μπορεί να εκφραστεί από έναν μόνο αριθμό που προκύπτει από τον συγκερασμό διαφόρων στοιχείων του πίνακα σύγχυσης.

	Εκτ- ω_1	Εκτ- ω_2	Εκτ- ω_3	Εκτ- ω_4	Εκτ- ω_5	Συν.
ω_1	98	0	2	0	0	100
ω_2	0	135	0	10	5	150
ω_3	0	1	198	0	1	200
ω_4	35	0	25	40	0	100
ω_5	0	0	0	0	150	150
Συν.	133	136	225	50	156	700

Μετρικές αξιολόγησης ταξινομητών

ο **Ανάκληση** (Recall)

- ❖ Είναι το ποσοστό των διανυσμάτων που προέρχονται από την κλάση ω_i και ταξινομούνται (σωστά) στην κλάση αυτή:

$$R_i = \frac{A(i,i)}{\sum_{j=1}^N A(i,j)}$$

- ❖ Σχετίζεται με το ερώτημα: "Δοθέντος ενός δείγματος που προέρχεται από την κλάση ω_i , πόσο πιθανό είναι να ταξινομηθεί σωστά από τον ταξινομητή;"

ο **Ακρίβεια** (Precision)

- ❖ Είναι το ποσοστό των διανυσμάτων που ταξινομούνται στην κλάση ω_i και πράγματι ανήκουν στην κλάση αυτή:

$$P_i = \frac{A(i,i)}{\sum_{j=1}^N A(j,i)}$$

- ❖ Σχετίζεται με το ερώτημα: "Δοθέντος ενός δείγματος που ταξινομήθηκε στην κλάση ω_i , πόσο πιθανό είναι η ταξινόμηση αυτή να είναι σωστή;"

Μετρικές αξιολόγησης ταξινομητών

○ Precision/Recall

	Εκτ- ω_1	Εκτ- ω_2	Εκτ- ω_3	Εκτ- ω_4	Εκτ- ω_5	Συν.	<i>R</i>
ω_1	98	0	2	0	0	100	98/100
ω_2	0	135	0	10	5	150	135/150
ω_3	0	1	198	0	1	200	198/200
ω_4	35	0	25	40	0	100	40/100
ω_5	0	0	0	0	150	150	150/150
Συν.	133	136	225	50	156	700	
<i>P</i>	98/133	135/136	198/225	40/50	150/156		

Μετρικές αξιολόγησης ταξινομητών

○ Precision/Recall – Η περίπτωση των δύο κλάσεων

- ❖ Πολλές φορές οι παραπάνω έννοιες χρησιμοποιούνται για την περίπτωση δύο κλάσεων όπου (συνήθως) η μία (η κλάση ενδιαφέροντος) είναι πολύ μικρότερη της άλλης (πρόβλημα μη-ισοζυγισμένων κλάσεων – **class imbalance problem**).
- ❖ Εδώ εστιάζουμε στα *R* και *P* της κλάσης ενδιαφέροντος.
- ❖ Η πιθανότητα σφάλματος σε τέτοιες περιπτώσεις **δεν είναι καλό μέτρο της απόδοσης** του ταξινομητή.
- ❖ Ο πίνακας σύγχυσης τώρα γίνεται:

	Εκτ- ω_1	Εκτ- ω_2	<i>R</i>
ω_1	Αληθώς θετικά (true positives - TP)	Ψευδώς αρνητικά (false negatives - FN)	TP/(TP+FN)
ω_2	Ψευδώς θετικά (false positives - FP)	Αληθώς αρνητικά (true negatives - TN)	
<i>P</i>	TP/(TP+FP)		

Μετρικές αξιολόγησης ταξινομητών

○ **Precision/Recall** – Η περίπτωση των δύο κλάσεων

- ❖ **Παράδειγμα:** Το πρόβλημα της ταξινόμησης ανθρώπων σε "θετικούς" (κλάση ω_1) και "αρνητικούς" (κλάση ω_2) σε μια σπάνια ασθένεια. Το **δείγμα** μας αποτελείται από $N = 100.000$ ανθρώπους από τους οποίους οι 10 μόνο είναι θετικοί στην ασθένεια (δηλαδή ανήκουν στην ω_1).
- ❖ **Σενάριο 1:** Έστω ταξινομητής που ταξινομεί όλους τους ανθρώπους στην κλάση ω_2 . Ο πίνακας σύγχυσης σε αυτήν την περίπτωση θα είναι:

	Εκτ- ω_1	Εκτ- ω_2	R
ω_1	TP=0	FN=10	TP/(TP+FN)=0
ω_2	FP=0	TN=99990	
P	TP/(TP+FP)=0		

- ❖ Η ΟΑ για έναν ταξινομητή που ταξινομεί όλους τους ανθρώπους ως "αρνητικούς" είναι 99,99% ! Ωστόσο, ο ταξινομητής αυτός δεν θα ανιχνεύσει ποτέ έναν ασθενή.
- ❖ Στην περίπτωση αυτή έχουμε: $P=0$ και $R=0$.

Μετρικές αξιολόγησης ταξινομητών

○ **Precision/Recall** – Η περίπτωση των δύο κλάσεων

- ❖ **Παράδειγμα:** Το πρόβλημα της ταξινόμησης ανθρώπων σε "θετικούς" (κλάση ω_1) και "αρνητικούς" (κλάση ω_2) σε μια σπάνια ασθένεια. Το **δείγμα** μας αποτελείται από $N = 100.000$ ανθρώπους από τους οποίους οι 10 μόνο είναι θετικοί στην ασθένεια (δηλαδή ανήκουν στην ω_1).
- ❖ **Σενάριο 2:** Έστω ταξινομητής που ταξινομεί όλους τους ανθρώπους στην ω_2 εκτός έναν της ω_1 που τον ταξινομεί σωστά. Ο πίνακας σύγχυσης τώρα θα είναι:

	Εκτ- ω_1	Εκτ- ω_2	R
ω_1	TP=1	FN=9	TP/(TP+FN)=1/10
ω_2	FP=0	TN=99990	
P	TP/(TP+FP)=1		

- ❖ Η ΟΑ για τον ταξινομητή αυτό θα είναι $99.991/100.000 = 99,99\%$!
- ❖ Στην περίπτωση αυτή: $P=1$ και $R=1/10$.

Μετρικές αξιολόγησης ταξινομητών

- **Precision/Recall** – Η περίπτωση των δύο κλάσεων
 - ❖ **Παράδειγμα:** Το πρόβλημα της ταξινόμησης ανθρώπων σε "θετικούς" (κλάση ω_1) και "αρνητικούς" (κλάση ω_2) σε μια σπάνια ασθένεια. Το **δείγμα** μας αποτελείται από $N = 100.000$ ανθρώπους από τους οποίους οι 10 μόνο είναι θετικοί στην ασθένεια (δηλαδή ανήκουν στην ω_1).
 - ❖ **Σενάριο 3:** Έστω ταξινομητής που ταξινομεί όλους τους ανθρώπους στην ω_1 . Ο πίνακας σύγχυσης σε αυτήν την περίπτωση θα είναι:

	Εκτ- ω_1	Εκτ- ω_2	R
ω_1	TP=10	FN=0	TP/(TP+FN)=1
ω_2	FP=99990	TN=0	
P	TP/(TP+FP)=1/10000		
 - ❖ Η ΟΑ για έναν ταξινομητή που ταξινομεί όλους τους ανθρώπους ως "θετικούς" είναι $10/100.000 = 0,01\%$!
 - ❖ Στην περίπτωση αυτή: $P=0.0001$ και $R=1$.

Μετρικές αξιολόγησης ταξινομητών

- **Precision/Recall** – Η περίπτωση των δύο κλάσεων
 - ❖ Συνήθως οι P και R δεν παίρνουν ταυτόχρονα μεγάλες τιμές. Το σε ποιά από τις δύο θα δώσουμε μεγαλύτερη βαρύτητα εξαρτάται από το εκάστοτε πρόβλημα.
 - ❖ Αν π.χ. στο παράδειγμα θέλουμε να ανιχνεύσουμε όλους τους ασθενείς που πράγματι έχουν την ασθένεια (έστω και αν κάποιους τους κατατάξουμε εσφαλμένα ως ασθενείς – FP), θέλουμε μεγάλη τιμή για το R.
 - ❖ Αν όμως οι επιπλέον εξετάσεις που απαιτούνται για την ανίχνευση της νόσου είναι πολύ ακριβές, θα θέλαμε επιπλέον και μια σχετικά μεγάλη τιμή για το P.
 - ❖ Συνδυασμός των P και R.
 - ❖ **F1-score ή F1-measure:**
 - Ο αρμονικός μέσος των P & R
 - $$F1\text{-score} = \frac{2}{\frac{1}{R} + \frac{1}{P}} = \frac{2 * R * P}{R + P}$$

	Εκτ- ω_1	Εκτ- ω_2	R
ω_1	Αληθώς θετικά (true positives - TP)	Ψευδώς αρνητικά (false negatives - FN)	TP/(TP+FN)
ω_2	Ψευδώς θετικά (false positives - FP)	Αληθώς αρνητικά (true negatives - TN)	
P	TP/(TP+FP)		