

## ΜΕΤΡΙΚΕΣ ΑΞΙΟΛΟΓΗΣΗΣ

Τα συστήματα που θα δημιουργηθούν με την εργασία θα πρέπει αξιολογηθούν. Αυτό γίνεται μέσω κάποιων μεγεθών που ονομάζονται *μετρικές αξιολόγησης*.

Η πιο απλή και αντιπροσωπευτική μετρική που μπορούμε να χρησιμοποιήσουμε είναι η *Γενική Ορθότητα Πρόβλεψης (General Predictive Accuracy)* που είναι απλά το ποσοστό των στιγμιότυπων του συνόλου ελέγχου που ταξινομήθηκαν στην σωστή κλάση.

Μια ομάδα επιπλέον συνήθων μετρικών είναι οι παρακάτω:

- Ορθότητα (Accuracy)  $acc = \frac{TP + TN}{TP + FP + FN + TN}$
- Ακρίβεια (Precision),  $prec = \frac{TP}{TP + FP}$
- Ευαισθησία (Sensitivity)  $sen = \frac{TP}{TP + FN}$
- Εξειδίκευση (Specificity)  $spec = \frac{TN}{TN + FP}$

όπου οι παράμετροι TP (True Positive), FN (False Negative), FP (False Positive), και TN (True Negative) υπολογίζονται ως εξής:

TP = όσα παραδείγματα ανήκουν στην κλάση (εξόδου) 1 και ταξινομήθηκαν στην 1  
FN = όσα παραδείγματα ανήκουν στην κλάση (εξόδου) 1, αλλά ταξινομήθηκαν στην 2  
FP = όσα παραδείγματα ανήκουν στην κλάση (εξόδου) 2, αλλά ταξινομήθηκαν στην 1  
TN = όσα παραδείγματα ανήκουν στην κλάση (εξόδου) 2 και ταξινομήθηκαν στην 2

Οι παραπάνω ορισμοί τους αφορούν δυαδικές εξόδους. Στην περίπτωση που έχουμε περισσότερες από δύο κλάσεις εξόδου, τότε οι παραπάνω παράμετροι υπολογίζονται χωριστά για κάθε κλάση  $i$ , ως εξής:

$TP_i$  = όσα ανήκουν στην κλάση (εξόδου)  $i$  και ταξινομήθηκαν στην  $i$   
 $FN_i$  = όσα ανήκουν στην κλάση (εξόδου)  $i$ , αλλά δεν ταξινομήθηκαν σ' αυτήν  
 $FP_i$  = όσα δεν ανήκουν στην κλάση (εξόδου)  $i$ , αλλά ταξινομήθηκαν στην  $i$   
 $TN_i$  = όσα δεν ανήκουν στην κλάση (εξόδου)  $i$  και δεν ταξινομήθηκαν σ' αυτήν

Ουσιαστικά για κάθε κλάση βλέπουμε το πρόβλημα ως δυαδικό, όπου δηλαδή η πρώτη έξοδος είναι η ίδια η κλάση και η δεύτερη έξοδος όλες οι υπόλοιπες. Προσέξτε ότι αντίθετα με την περίπτωση πραγματικά δυαδικού προβλήματος, εάν το σύστημα προβλέψει σωστά ότι ένα στιγμιότυπο δεν ανήκει στην κλάση που ελέγχουμε (περίπτωση True Negative), αυτό δεν σημαίνει απαραίτητα ότι το σύστημα το πρόβλεψε και στην σωστή. Για αυτόν τον λόγο η παράμετρος TN και όσες μετρικές την χρησιμοποιούν (Accuracy, Specificity) χάνουν την αξιοπιστία τους.

Για αυτόν τον λόγο σε προβλήματα με περισσότερες από δύο εξόδους, εκτός από την Γενική Ορθότητα Πρόβλεψης δίνουμε βάση στις μετρικές Ακρίβεια (Precision, που ορίσαμε πριν) και Ανάκληση (Recall που αντιστοιχεί στην Ευαισθησία-Sensitivity που ορίσαμε πριν). Άλλη χρήσιμη μετρική είναι η F-Measure που συνδυάζει τις Recall και Precision όπως φαίνεται στους τύπους:

$$recall = \frac{TP}{TP + FN}, \quad precision = \frac{TP}{TP + FP}, \quad F\_measure = \frac{2 \times precision \times recall}{precision + recall}$$

Τέλος έχοντας υπολογίσει τις παραπάνω μετρικές για κάθε κλάση, μπορούμε να πάρουμε τον μέσο όρο. Συνήθως χρησιμοποιούμε βάρη στον υπολογισμό του MO ώστε να λάβουμε υπόψη την συχνότητα της κάθε κλάσης στο σύνολο ελέγχου.

Για παράδειγμα αν έχουμε m κλάσεις και το σύνολο ελέγχου έχει k στιγμιότυπα με  $k_i$  τα στιγμιότυπα της κλάσης i τότε μπορούμε να υπολογίσουμε τον μέσο όρο της Ανάκλησης ως εξής:

$$Weight\_Avg\_Recall = \sum_{i=1}^m \frac{k_i}{k} \times Recall_i$$

Οι παραπάνω μετρικές χρησιμοποιούνται ευρέως από συστήματα μηχανικής μάθησης και εξόρυξης δεδομένων όπως το WEKA.

Γενικά για την ερμηνεία των μετρικών ανατρέξτε στα εξής:

[http://en.wikipedia.org/wiki/Accuracy\\_and\\_precision](http://en.wikipedia.org/wiki/Accuracy_and_precision)

[http://en.wikipedia.org/wiki/Precision\\_and\\_recall](http://en.wikipedia.org/wiki/Precision_and_recall)

[http://en.wikipedia.org/wiki/Specificity\\_\(tests\)](http://en.wikipedia.org/wiki/Specificity_(tests))