



# Μεθοδολογία Έρευνας και Στατιστική Ανάλυση με Χρήση Ανοικτού Λογισμικού

Καμπύλη λειτουργικού χαρακτηριστικού δέκτη

Receiver Operating Characteristic (ROC)

# Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

Με τη μέθοδο της λογιστικής παλινδρόμησης δημιουργούνται **εξισώσεις που**, αξιοποιώντας άλλα διαθέσιμα μετρήσιμα χαρακτηριστικά, **υπολογίζουν την πιθανότητα για ένα αντικείμενο να έχει ή να μην έχει μία ιδιότητα**. Ωστόσο, ο ερευνητής, αν θέλει να χρησιμοποιήσει την εξίσωση αυτή ως εργαλείο ταξινόμησης αγνώστων αντικειμένων, θα πρέπει να δώσει μία απάντηση στην ερώτηση:

“Πάνω από ποια τιμή εκτιμώμενης πιθανότητας, δέχομαι ως σωστή την πρόβλεψη;”

Το όριο αυτό, είναι δυνατό να οριστεί από τον ερευνητή σε οποιαδήποτε τιμή μεταξύ 0 και 1, με τις διάφορες επιλογές να αντιστοιχούν σε διαφορετικούς πίνακες συμπτώσεων με διαφορετικές τιμές ευαισθησίας και ειδικότητας.

Το σύνολο των ζευγών (ευαισθησία, 1 – ειδικότητα) είναι δυνατόν να τοποθετηθούν σε ένα διάγραμμα που θα διευκολύνει την απόφαση του ερευνητή. Το διάγραμμα αυτό είναι γνωστό ως **Καμπύλη λειτουργικού χαρακτηριστικού δέκτη** - receiver operating characteristic curve ή πιο απλά **καμπύλη ROC**.

# Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

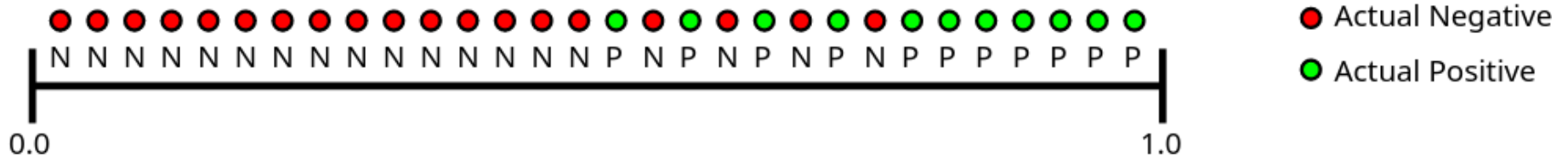
Η ονομασία “Καμπύλη ROC” χρονολογείται από τον Δεύτερο Παγκόσμιο Πόλεμο και τους ηλεκτρολόγους μηχανικούς που υποστήριζαν τη λειτουργία των ραντάρ.

Όταν η ενίσχυση του σήματος στο ραντάρ ήταν μηδενική, τότε δεν ανιχνευόταν κανένα εχθρικό αεροπλάνο. Με την αύξηση της ενίσχυσης, ανιχνευόταν όλο και περισσότερα αντικείμενα, αυξάνοντας ωστόσο και την ποσότητα του θορύβου η οποία οδηγούσε σε εσφαλμένα θετικά αποτελέσματα (False Positive), δηλαδή αναγνώριση αντικειμένων που μπορεί να φαίνονταν σαν αεροπλάνα αλλά δεν ήταν, όπως σύννεφα βροχής ή κοπάδια πουλιών. Από κάποια τιμή, η περαιτέρω αύξηση της ενίσχυσης ήταν αντιπαραγωγική, καθώς ο θόρυβος (False Positive) αρχίζει να υπερτερεί των σημάτων (True Positive).

Η καμπύλη ROC επινοήθηκε ως ένα απλό στη δημιουργία του διάγραμμα που θα βοηθούσε το χειριστή να αποφασίσει ποιο επίπεδο ενίσχυσης θα προσέφερε ικανοποιητική ευαισθησία (True Positive Rate, TPR) και σχετικά μεγάλη ειδικότητα ή ισοδύναμα μικρή πιθανότητα εσφαλμένης ανίχνευσης (False Positive Rate  $FPR = 1 - \text{ειδικότητα}$ ).

Πηγή: Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, 52(2), 121–128. <https://doi.org/10.1177/070674370705200210>

# Καμπύλη ROC



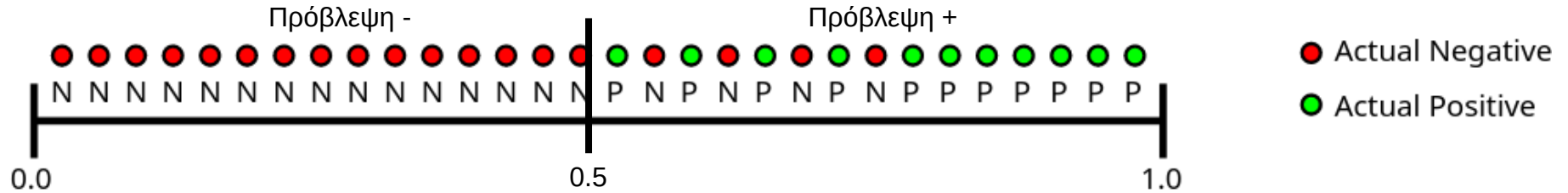
Στην προσπάθεια προσαρμογής ενός μοντέλου πρόβλεψης spam email, έγινε χρήση δείγματος 30 email τα οποία **ήταν spam**, ή **δεν ήταν**.

Μοντέλο πρόβλεψης αξιοποίησε άλλα χαρακτηριστικά των email (αποστολέας, πλήθος παραληπτών, περιεχόμενο κλπ) και απέδωσε μία πιθανότητα μεταξύ 0 και 1 σε κάθε ένα από αυτά να είναι spam. Τα 30 email τοποθετήθηκαν στην σειρά ως προς την εκτιμώμενη πιθανότητα. Το ερώτημα που τώρα πρέπει να απαντηθεί είναι:

**Πάνω από ποια τιμή εκτιμώμενης πιθανότητας, το μοντέλο έχει ικανοποιητική ευαισθησία και αρκετά μεγάλη ειδικότητα;**



# Καμπύλη ROC



Επιλέγοντας ως όριο θετικότητας την πιθανότητα 0,5, ο πίνακας συμπτώσεων της πρόβλεψης του μοντέλου με την πραγματική κατάσταση των email είναι:

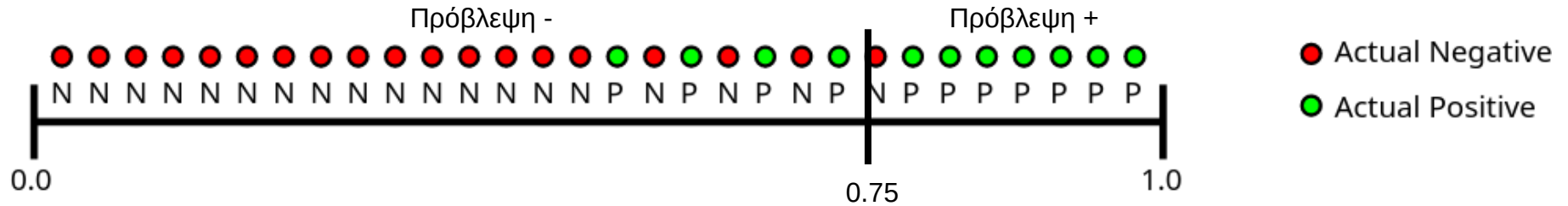
Πίνακας: Όριο αποδοχής 0,5				
		Πρόβλεψη		
		Όχι	Ναι	Σύνολο
Το email είναι spam	Όχι	$\alpha = 15$	$\beta = 4$	$O_1 = 19$
	Ναι	$\gamma = 0$	$\delta = 11$	$O_2 = 11$
Σύνολο		$P_1 = 15$	$P_2 = 15$	$\Sigma = 30$

Από τον πίνακα υπολογίζουμε, πως στην επιλογή ορίου (cut – off score) 0,5 αντιστοιχεί:

Ευαισθησία (sensitivity) =  $\delta / O_2 = 11 / 11 = 100\%$

Ειδικότητα (specificity) =  $\alpha / O_1 = 15 / 19 = 78,9\%$

# Καμπύλη ROC



Επιλέγοντας ως όριο θετικότητας την πιθανότητα 0,75, ο πίνακας συμπτώσεων της πρόβλεψης του μοντέλου με την πραγματική κατάσταση των email είναι:

Πίνακας: Όριο αποδοχής 0,75

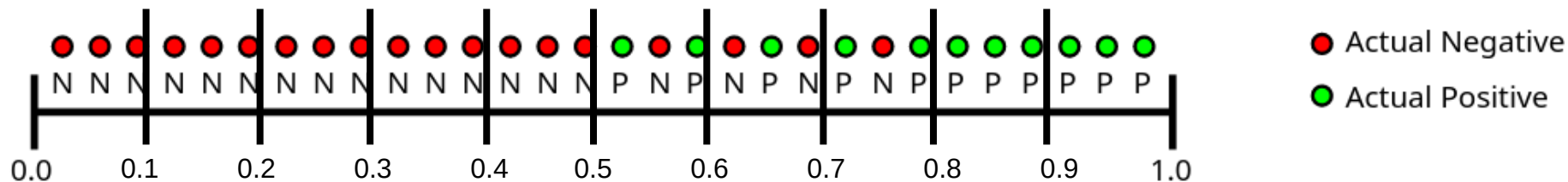
		Πρόβλεψη		
		Όχι	Ναι	Σύνολο
Το email είναι spam	Όχι	$\alpha = 18$	$\beta = 1$	$O_1 = 19$
	Ναι	$\gamma = 4$	$\delta = 7$	$O_2 = 11$
Σύνολο		$P_1 = 22$	$P_2 = 8$	$\Sigma = 30$

Από τον πίνακα υπολογίζουμε, πως στην επιλογή ορίου (cut – off score) 0,5 αντιστοιχεί:

Ευαισθησία (sensitivity) =  $\delta / O_2 = 7 / 11 = 63,6\%$

Ειδικότητα (specificity) =  $\alpha / O_1 = 18 / 19 = 94,7\%$

# Καμπύλη ROC



Η διαδικασία που ακολουθήσαμε για cut-off score  $\alpha = 0,25, 0,5$  και  $0,75$  είναι δυνατόν να επαναληφθεί για κάθε τιμή μεταξύ 0 και 1. Στον πίνακα παρουσιάζονται οι τιμές για  $\alpha = 0,1, 0,2, \dots, 1$ .

$\alpha$	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
1 – ειδικότητα (FPR)	100%	84,2%	68,4%	52,6%	36,8%	21,1%	15,8%	5,3%	0,0%	0,0%	0,0%
ευαισθησία (TPR)	100%	100%	100%	100%	100%	100%	81,8%	72,7%	54,5%	27,3%	0,0%



$\alpha$	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
1 – ειδικότητα (FPR)	100%	84,2%	68,4%	52,6%	36,8%	21,1%	15,8%	5,3%	0,0%	0,0%	0,0%
ευαισθησία (TPR)	100%	100%	100%	100%	100%	100%	81,8%	72,7%	54,5%	27,3%	0,0%

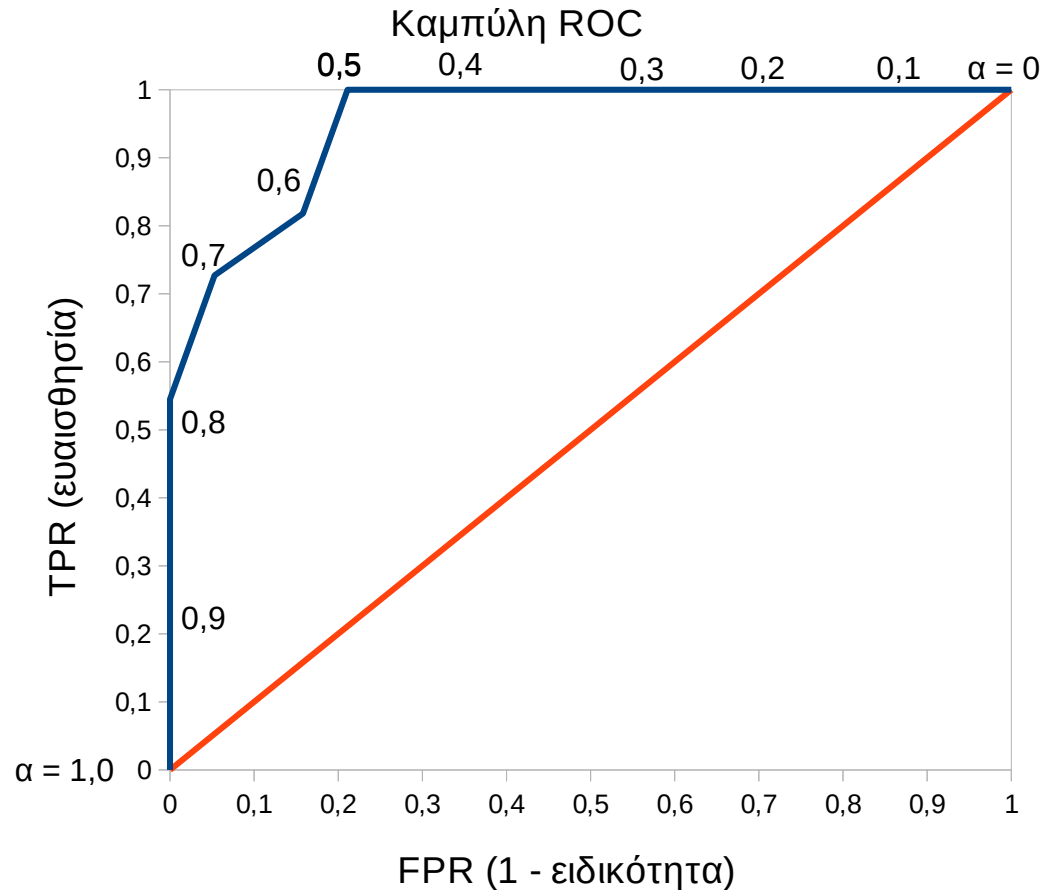
Τα ζεύγη των τιμών

$$(FPR, TPR) = (1 - \text{ειδικότητα}, \text{ευαισθησία}),$$

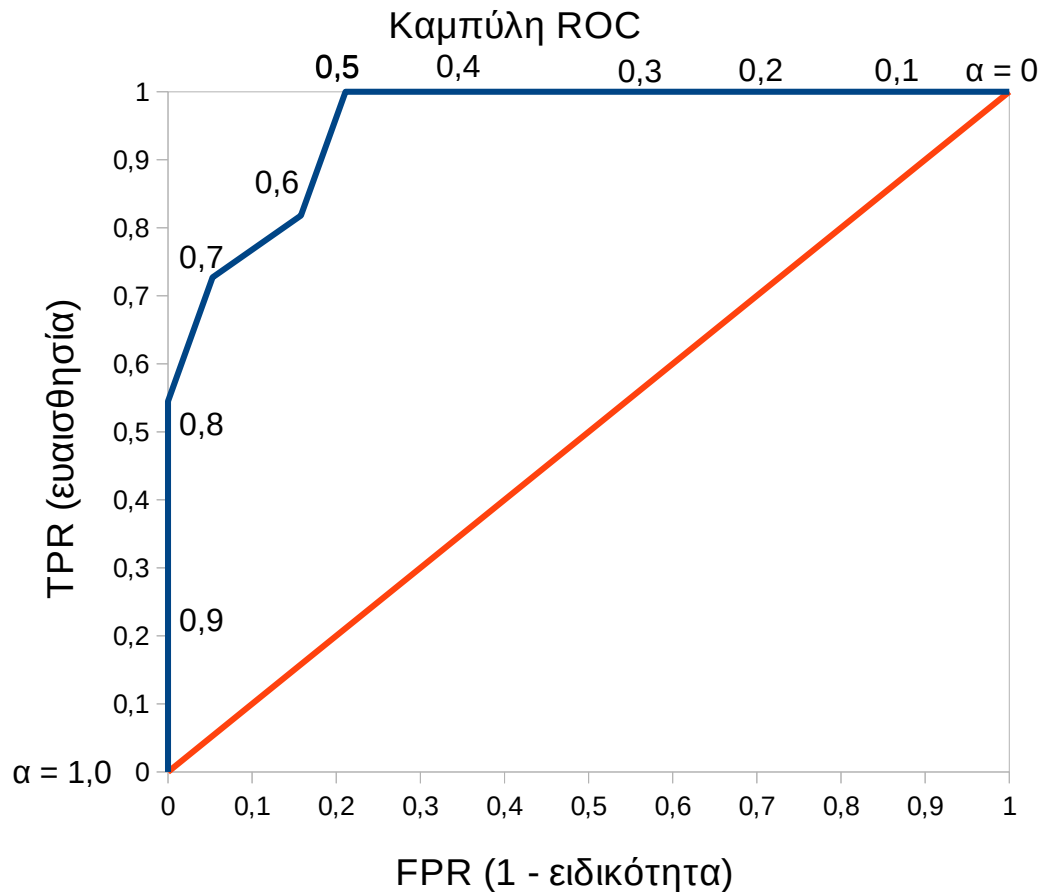
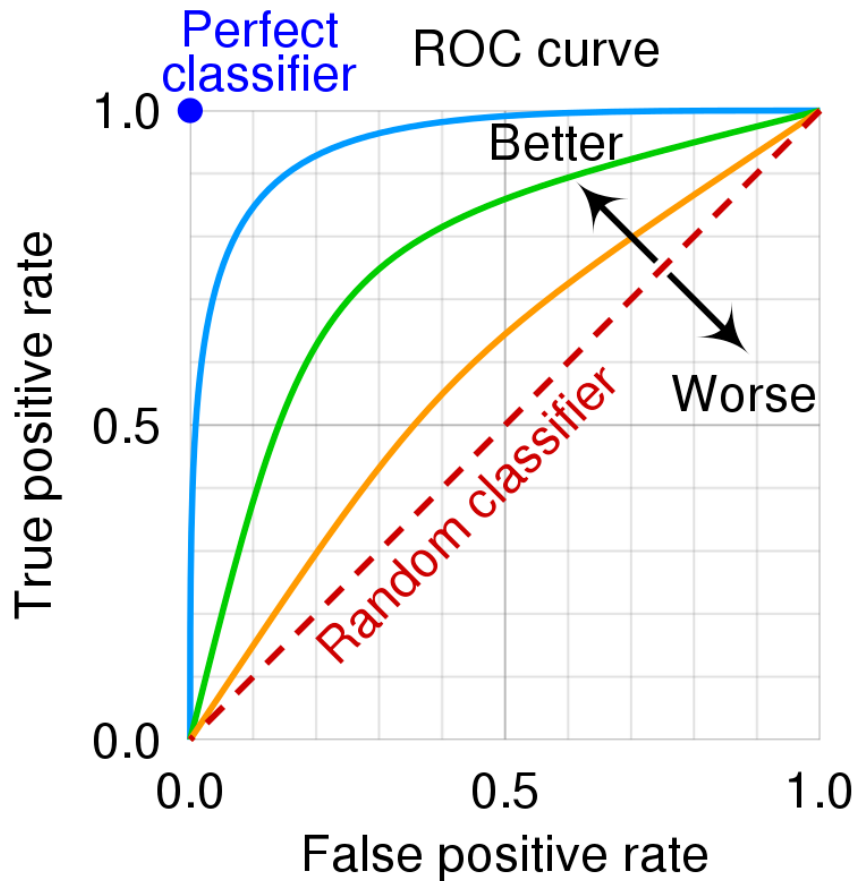
αν αναπαρασταθούν σε διαγραμματική μορφή, συνιστούν την καμπύλη λειτουργικού χαρακτηριστικού δέκτη (καμπύλη ROC). Ο ερευνητής μπορεί να επιλέξει το όριο  $\alpha$  με το οποίο επιτυγχάνεται αξιοπρεπής ευαισθησία και ειδικότητα στο μοντέλο πρόγνωσης που έχει αναπτύξει.

Η καμπύλη ROC, συνήθως απεικονίζεται μαζί με την ευθεία  $y = x$ , η οποία αντιστοιχεί στην τυχαία ταξινόμηση, όπου  $FPR = TPR$ . Ως εκ τούτου, επιθυμία είναι η επιλογή cut-off score  $\alpha$  που να εξασφαλίζει ζεύγος  $(FPR, TPR)$  όσο το δυνατόν πιο μακριά από αυτή.

Στην συγκεκριμένη περίπτωση η επιλογή  $\alpha = 0,5$  ή  $\alpha = 0,6$  αντιστοιχεί σε μοντέλο με ικανοποιητική επίδοση στην πρόβλεψη των spam email.



# Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)



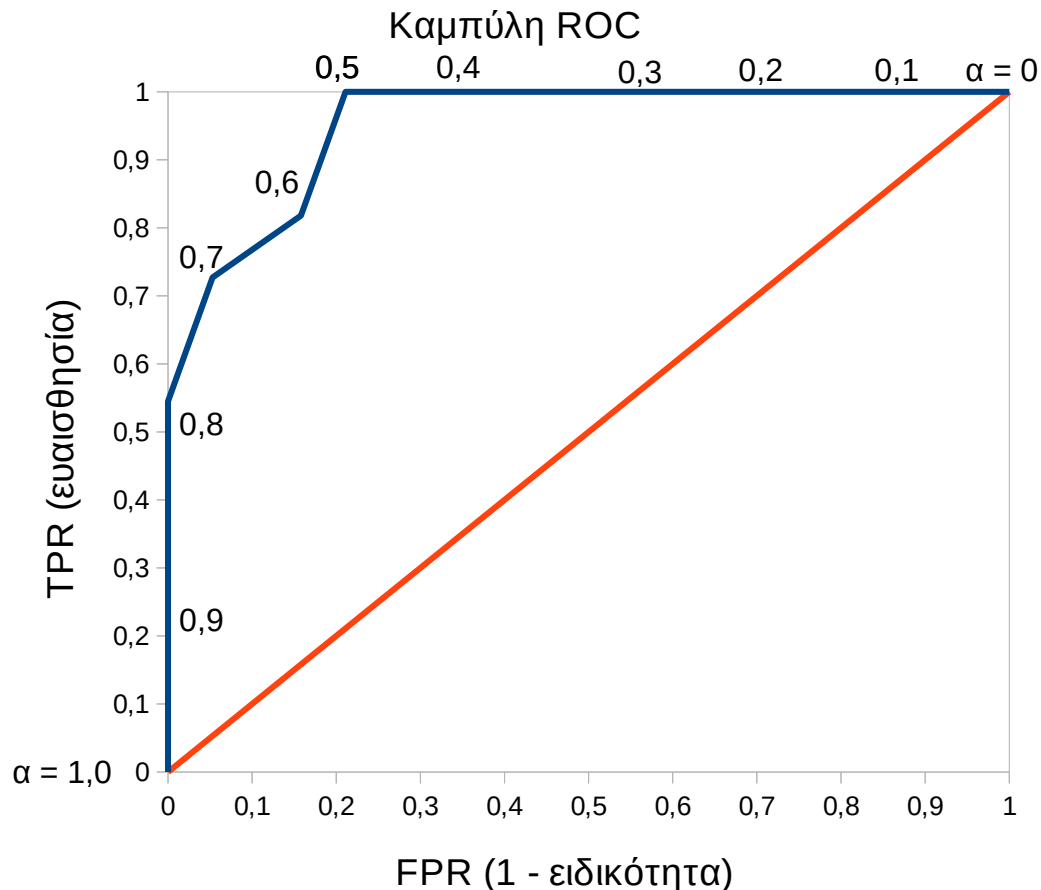
# Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

Το εμβαδόν της περιοχής κάτω από την ROC καμπύλη (Area Under Curve – AUC) είναι ένα βασικό μέτρο αξιολόγησης της μεθόδου ταξινόμησης.

Αν  $AUC < 0,5$  τότε η επίδοση του μοντέλου είναι χειρότερη από την τυχαία επιλογή και πρέπει να απορριφθεί εξ' ολοκλήρου.

Όσο πιο κοντά στο 1 βρίσκεται το AUC τόσο καλύτερη είναι η διαδικασία ταξινόμησης στο σύνολό της.

Επιπλέον, το AUC επιτρέπει τη σύγκριση δύο ή περισσότερων διαφορετικών μοντέλων που στοχεύουν στην ταξινόμηση των παρατηρήσεων.



# Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

## Παράδειγμα 1

Ένας τρόπος διάγνωσης του υποθυρεοειδισμού, είναι το επίπεδο της ορμόνης T4. Για τον έλεγχο της αξιοπιστίας της μέτρησης αυτής ελέγχθηκαν 125 ασθενείς από τους οποίους οι 32 ήταν βέβαιο ότι έπασχαν από υποθυρεοειδισμό και βρέθηκαν τα αποτελέσματα του πίνακα.

(α) Να συμπληρωθούν πίνακες αξιολόγησης της διαγνωστικής ικανότητας της ορμόνης T4 στα επίπεδα, των 5,0, 7,0 και 9,0 microliter/dl.

(β) Να συμπληρωθεί ο πίνακας και να γίνει η καμπύλη ROC

Όριο διάγνωσης T4	5,0	7,0	9,0
Ευαισθησία			
Ειδικότητα			
1 – Ειδικότητα			

Ορμόνη T4	Υποθυρεοειδισμός		Σύνολο
	Ναι	Όχι	
< 5,0	18	1	19
5,1 – 7,0	7	17	24
7,1 – 9,0	4	36	40
> 9,0	3	39	42
Σύνολο	32	93	125

# Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

Όριο 5,0	Υποθυρεοειδισμός		Σύνολο
	Ναι	Όχι	
Ορμόνη Τ4			
< 5,0	18	1	19
≥ 5,0	14	92	104
Σύνολο	32	93	125

Όριο 5,0: Ευαισθ. = 18/19, Ειδ. = 92/104

Όριο 7,0: Ευαισθ. = 25/43, Ειδ. = 75/82

Όριο 9,0: Ευαισθ. = 29/83, Ειδ. = 39/42

Όριο 7,0	Υποθυρεοειδισμός		Σύνολο
	Ναι	Όχι	
Ορμόνη Τ4			
< 7,0	25	18	43
≥ 7,0	7	75	82
Σύνολο	32	93	125

Όριο 9,0	Υποθυρεοειδισμός		Σύνολο
	Ναι	Όχι	
Ορμόνη Τ4			
< 9,0	29	54	83
≥ 9,0	3	39	42
Σύνολο	32	93	125

# Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

Όριο 5,0: Ευαισθ. =  $18/19 = 0,947$ , Ειδ. =  $92/104$

Όριο 7,0: Ευαισθ. =  $25/43$ , Ειδ. =  $75/82$

Όριο 9,0: Ευαισθ. =  $29/83$ , Ειδ. =  $39/42$

Όριο T4 για διάγνωση	5,0	7,0	9,0
Ευαισθησία	0,947	0,581	0,349
Ειδικότητα	0,885	0,915	0,929
1 – Ειδικότητα	0,115	0,085	0,071

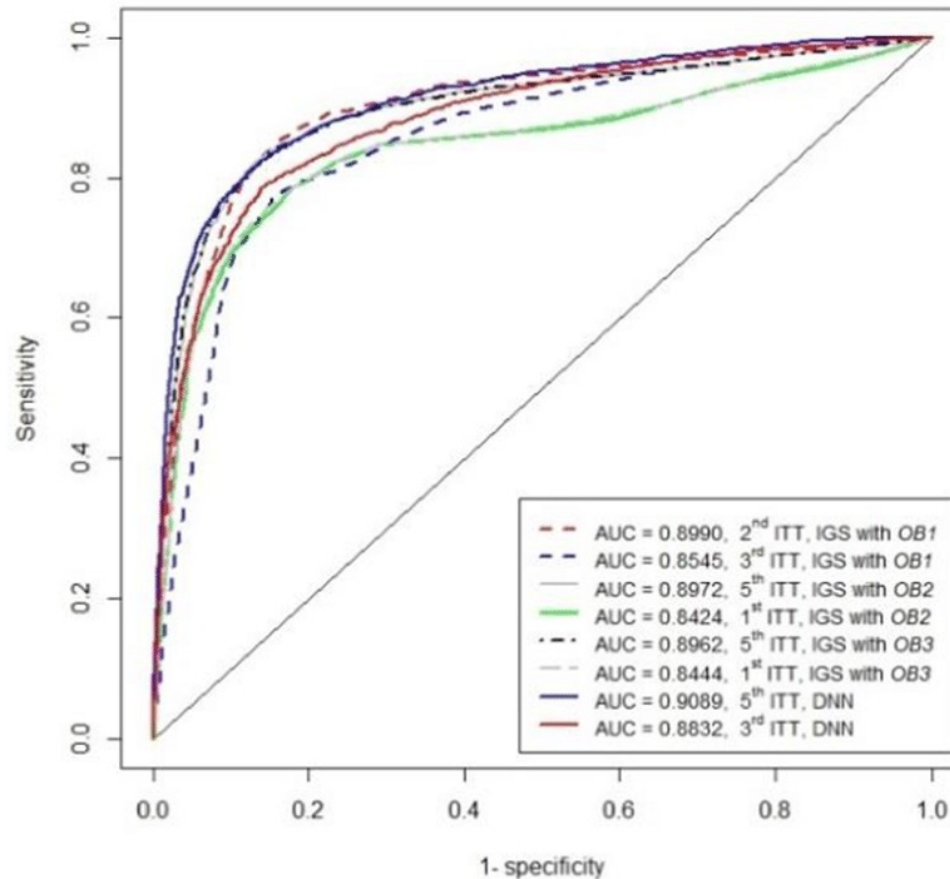
# Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

## Παράδειγμα 2

Σε πρόσφατη έρευνα, δεδομένα από εθνικό σύστημα υγείας της Ταϊβάν χρησιμοποιήθηκαν για την ανάπτυξη ενός συστήματος υποστήριξης κλινικών αποφάσεων με στόχο την πρόβλεψη του οξέος εμφράγματος του μυοκαρδίου.

Αναπτύχθηκαν 8 διαφορετικά μοντέλα πρόβλεψης, τα οποία αξιοποίησαν 75 διαφορετικά χαρακτηριστικά των ασθενών.

Τα 8 διαφορετικά μοντέλα πρόβλεψης συγκρίθηκαν ως προς την ποιότητα της πρόβλεψης που προσφέρουν με ένα πολλαπλό διάγραμμα ROC.



Πηγή: Wu, Fu-Hsing & Lai, Huey-Jen & Lin, Hsuan-Hung & Chan, Po-Chou & Tseng, Chien-Ming & Chang, Kun-Min & Chen, Yung-Fu & Lin, Chih-Sheng. (2022). Predictive models for detecting patients more likely to develop acute myocardial infarctions. The Journal of Supercomputing. 78. 1-29. 10.1007/s11227-021-03916-z.