ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Μηχανική μάθηση

Ενότητα 7: Metrics of Performance

Ιωάννης Τσαμαρδίνος

Τμήμα Επιστήμης Υπολογιστών

Given a Classification Model

1. We are given a classification model (e.g., an SVM model)
2. that is binary (2-class problem)
3. and a test set where the true classes are known

4. Estimate its performance

Confusion Matrix

Decision / Truth	Predicted Positive	Predicted Negative
Positive Example	True Positive	False Negative
Negative Example	False Positive	True Negative

- Positive/Negative refers to the **prediction**
- True/False refers to correctness

- False Positive / Type I error / False Alarm / hit
- False Negative / Type II error / miss

Confusion Matrix

Decision / Truth	Predicted Positive	Predicted Negative
Positive Example	True Positive	False Negative
Negative Example	False Positive	True Negative

- True Positive Rate (TPR) / Sensitivity :
  - Fraction of positives correctly classified
  - P(Predicted Positive | Pos)
  - Estimated as TP/Pos = TP/ (TP+FN)

Confusion Matrix

Decision / Truth	Predicted Positive	Predicted Negative
Positive Example	True Positive	False Negative
Negative Example	False Positive	True Negative

- **Specificity :**
  - fraction of negatives correctly classified
  - Estimate of P(Predicted Negative | Neg)
  - Estimated as TN/#Neg = TN/ (FP+TN)

Confusion Matrix

Decision \ Truth	Predicted Positive	Predicted Negative
Positive Example	True Positive	False Negative
Negative Example	False Positive	True Negative

- False Positive Rate (TPR) :
  - fraction of negatives incorrectly classified
  - Estimate of P(Predicted Positive | Neg)
  - Estimated as FP/#Neg = FP/ (FP+TN)
  - = 1 - specificity

Cost Matrix

Decision / Truth	Predicted Positive	Predicted Negative
Positive Example	0	c_1
Negative Example	c_2	0

- If you misclassify a positive then you incur cost of c_2
- If you misclassify a negative then you incur cost of c_1
- Typically, the diagonal entries are zero

Minimize Cost

- Ideally, we would like to minimize cost
- What is the cost?
- We get a TP and cost 0 with probability
  - $P(\text{ Predict Positive }, \text{Pos}) = P(\text{Predict Positive} \mid \text{Pos}) \cdot P(\text{Pos}) = sen \cdot P(\text{Pos})$
- We get a FP and cost of c_1 with probability
  - $P(\text{ Predict Positive }, \text{Neg}) = P(\text{Predict Positive} \mid \text{Neg}) \cdot P(\text{Neg})$
- etc.
- pos probability of being positive
- neg probability of being negative

$$c = c_1(1 - sen)\,pos + c_2(1 - spe)\,neg$$

Maximize Accuracy

- If $c_1 = c_2 = 1$ then the cost is the percentage of errors, i.e., P(error)
- P(error) = (1-sen) · *pos* + (1-spe) · *neg*
- Accuracy is the percentage of correct decisions
- Estimated as (TP + TN)/(TP+TN+FP+FN)
- Accuracy =
  - 1 – P(error)
  - *sen* · *pos* + *spe* · *neg*
  - *tpr* · *pos* + (1-fpr) · *neg*

Changing the Cost or the Prior Distr.

- Conditioned on the class, the probability of correct or incorrect classification is independent of the class distribution
- Why?
- So, the sensitivity, specificity, true positive rate, and false positive rate are independent of the class distribution

Example

- A physician gathers 200 cancerous tissues and 200 controls (non-cancerous) (case-control study)
- The tissue's gene-expression profile is produced with micro-arrays
- The data are split to 50-50% train and test sets with the same class distribution
- An SVM model is trained on the training data for diagnosing cancer based on the gene-expression values
- The confusion matrix on the **test** data is calculated

Decision Truth	Predicted Positive	Predicted Negative
Positive Example	90	10
Negative Example	30	70

Example

- $sen_1 = 0.9$
- $spe_1 = 0.7$
- Accuracy $A_1 = 160/200 = 80\%$

Decision / Truth	Predicted Positive	Predicted Negative
Positive Example	90	10
Negative Example	30	70

Example

- In the general population, under similar conditions, only 1/1000 such tissues is cancerous
- What is the accuracy on the general population?

- $A_2 = sen \cdot pos + spe \cdot neg$
  - $= 0.9 \cdot 1/1000 + 0.7 \cdot 999/1000$
  - $= 0.7002$

- What is the accuracy of the model: always predict negative?

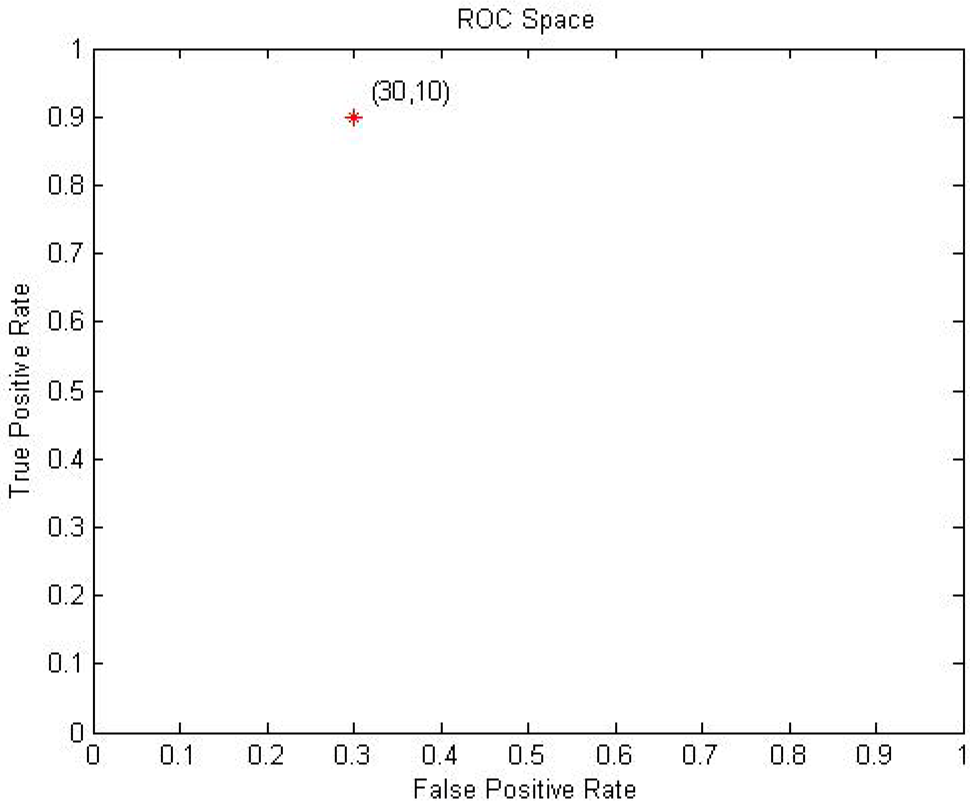
Evaluation Using Cost and Accuracy

- Best-accuracy model depends on the class distribution
  - no problem when training data and general population data follow the same distribution
- Best-cost model depends both on the class distribution and the cost matrix
  - the cost matrix is often unknown to the analyst
- Need to evaluate a model independent of both cost and class distribution. How?

ROC Space

- Plot the TPR versus the FPR or
- sensitivity versus (1-specificity)

ROC Space


ROC Space
(30,10)
True Positive Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

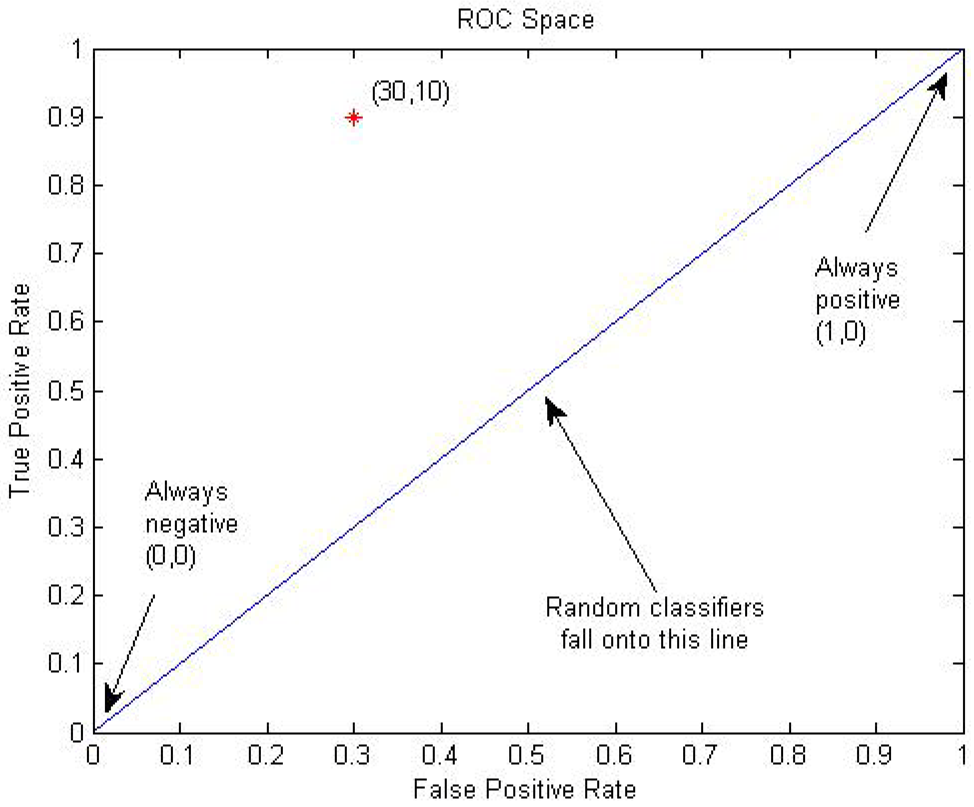
Trivial Classifiers

- Always predict positive
  - sensitivity = 1
  - specificity = 0
- Always predict negative
  - sensitivity = 0
  - specificity = 1
- Randomly predict positive with probability p
  - sensitivity = ?
  - specificity = ?

Trivial Classifiers

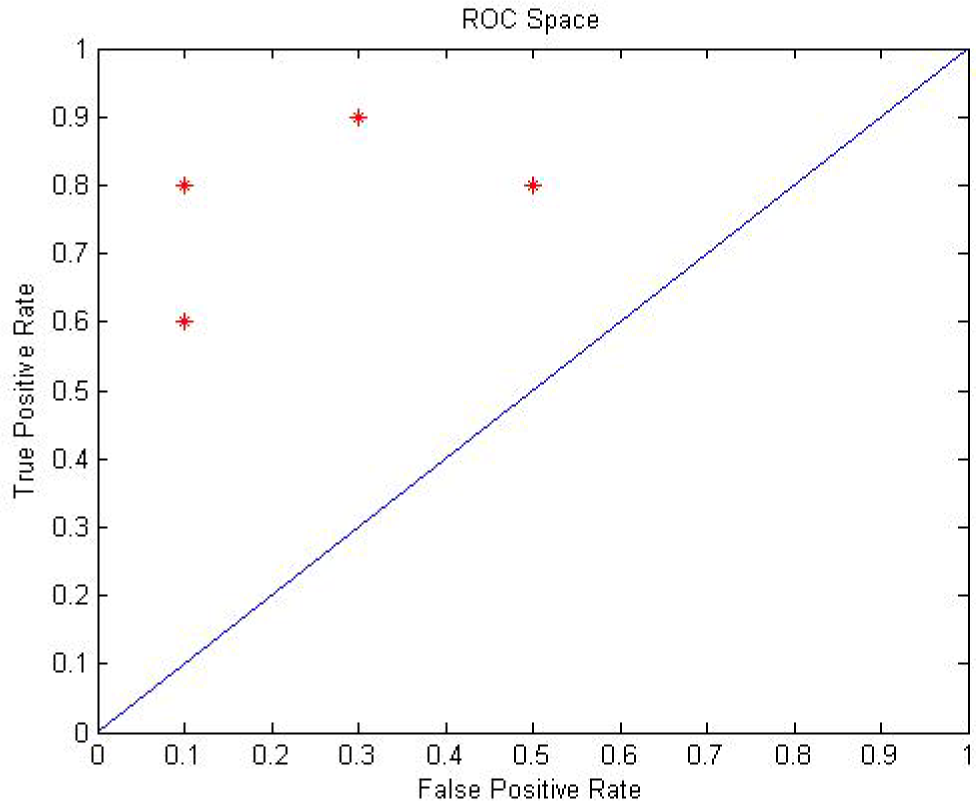
- Always predict positive
  - sensitivity = 1
  - specificity = 0
- Always predict negative
  - sensitivity = 0
  - specificity = 1
- Randomly predict positive with probability p
  - sensitivity = p
  - specificity = $1-p$

ROC Space



ROC Space
(30,10)
Always positive (1,0)
Always negative (0,0)
Random classifiers fall onto this line
True Positive Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Which is Better?



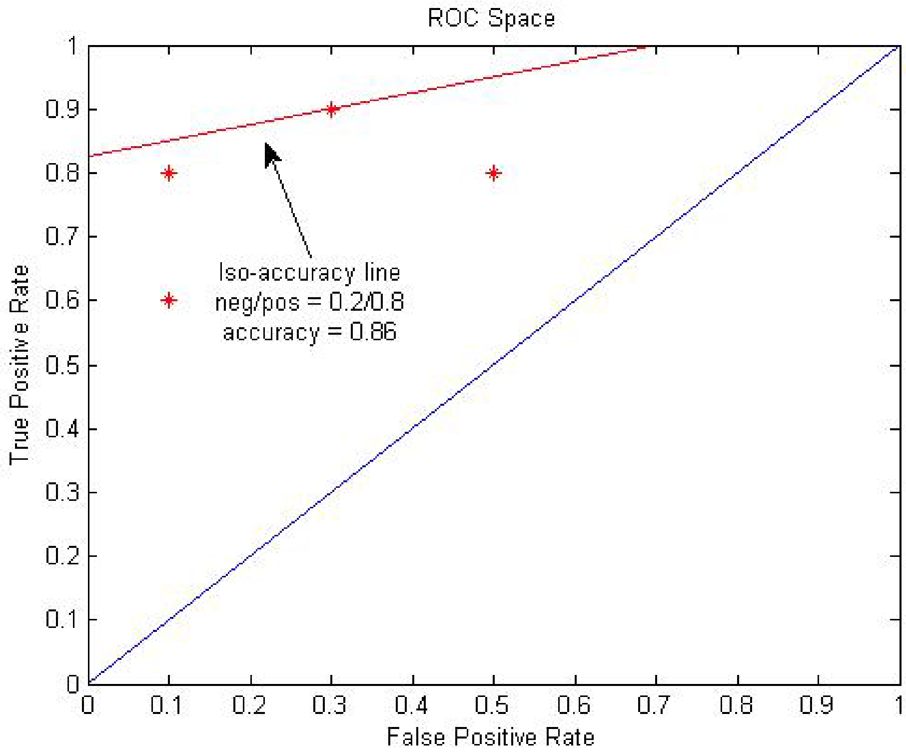
Iso-accuracy lines

- Assume models M_1 , M_2 have the same accuracy a
- $a = tpr_1 \cdot pos + (1\text{-}fpr_1) \cdot neg$, i.e.,

$$tpr_1 = \frac{neg}{pos} fpr_1 + \frac{a - neg}{pos}$$

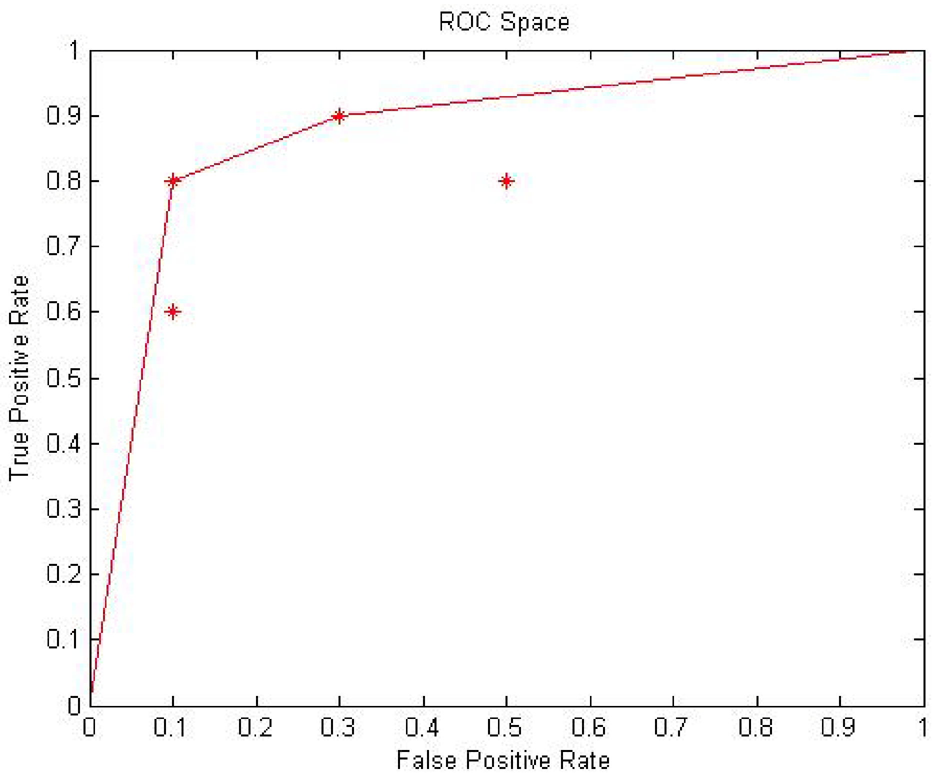
- and similarly for M_2
- When a is the same, all such models fall onto the same line with slope neg/pos and intercept (a-neg)/pos
- Such lines are called iso-accuracy lines

Iso-accuracy lines



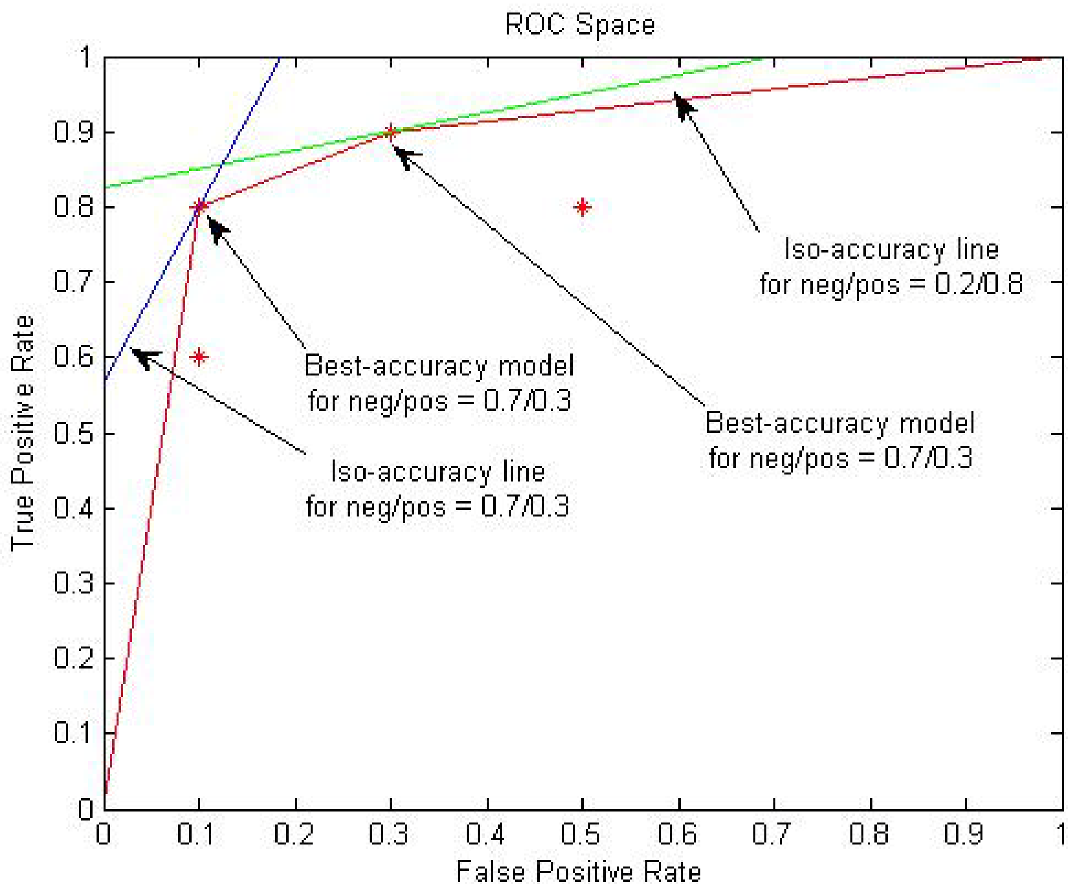
- All points on the iso-accuracy line have the same accuracy but for different sensitivity and specificity

ROC Convex Hull



- Points on the convex hull achieve the best accuracy for some class distribution, sensitivity and specificity

Choosing the Best-Accuracy Model



Iso-Cost Lines

- Equivalently, we define iso-cost lines as the lines in ROC space producing the same cost for different sensitivity and specificity

$$c = c_1(1 - sen)\,pos + c_2(1 - spe)\,neg$$

- Or

$$tpr = \frac{neg}{pos}\cdot\frac{c_2}{c_1}\,fpr + \frac{c_1 \cdot pos - c}{c_1 \cdot pos}$$

- The slope depends on the class distribution and the misclassification costs
- If the class distribution changes, we can get the same cost by changing the misclassification costs

Rankers versus Classifiers

- IDEA! Instead of classifying to positive/negative give a score to of your confidence that an example is positive
- The larger the score, the larger should be the confidence of the model that the example is positive
- How can we have models (e.g., SBC?) that output a score instead of a binary classification?
- What do we gain? We can threshold the score and get different points on the ROC for the **same** model! Let's see how.

Thresholding a Ranker

Score	True Class	Prediction
10	+	
7	+	
5	-	
1	+	
-3	-	

Thresholding a Ranker

Score	True Class	Prediction
10	+	-
7	+	-
5	-	-
1	+	-
-3	-	-

Threshold = 11

T=11, (sen, 1-spe) = (0, 0)

	PP	PN
Pos	0	3
Neg	0	2

Thresholding a Ranker

Score	True Class	Prediction
10	+	+
7	+	-
5	-	-
1	+	-
-3	-	-

Threshold = 8

T=11, (sen, 1-spe) = (0, 0)

T=8, (sen, 1-spe) = (0.33, 0)

	PP	PN
Pos	1	2
Neg	0	2

Thresholding a Ranker

Score	True Class	Prediction
10	+	+
7	+	+
5	-	-
1	+	-
-3	-	-

Threshold = 6

T=11, (sen, 1-spe) = (0, 0)

T=8, (sen, 1-spe) = (0.33, 0)

T=6, (sen, 1-spe) = (0.66, 0)

	PP	PN
Pos	2	1
Neg	0	2

Thresholding a Ranker

Score	True Class	Prediction
10	+	+
7	+	+
5	-	+
1	+	-
-3	-	-

Threshold = 2

	PP	PN
Pos	2	1
Neg	1	1

T=11, (sen, 1-spe) = (0, 0)

T=8, (sen, 1-spe) = (0.33, 0)

T=6, (sen, 1-spe) = (0.66, 0)

T=2, (sen, 1-spe) = (0.66, 0.5)

Thresholding a Ranker

Score	True Class	Prediction
10	+	+
7	+	+
5	-	+
1	+	+
-3	-	-

	PP	PN
Pos	3	0
Neg	1	1

Threshold = 0

T=11, (sen, 1-spe) = (0, 0)

T=8, (sen, 1-spe) = (0.33, 0)

T=6, (sen, 1-spe) = (0.66, 0)

T=2, (sen, 1-spe) = (0.66, 0.5)

T=0, (sen, 1-spe) = (1, 0.5)

Thresholding a Ranker

Score	True Class	Prediction
10	+	+
7	+	+
5	-	+
1	+	+
-3	-	+

	PP	PN
Pos	3	0
Neg	2	0

Threshold = -4

T=11, (sen, 1-spe) = (0, 0)

T=8, (sen, 1-spe) = (0.33, 0)

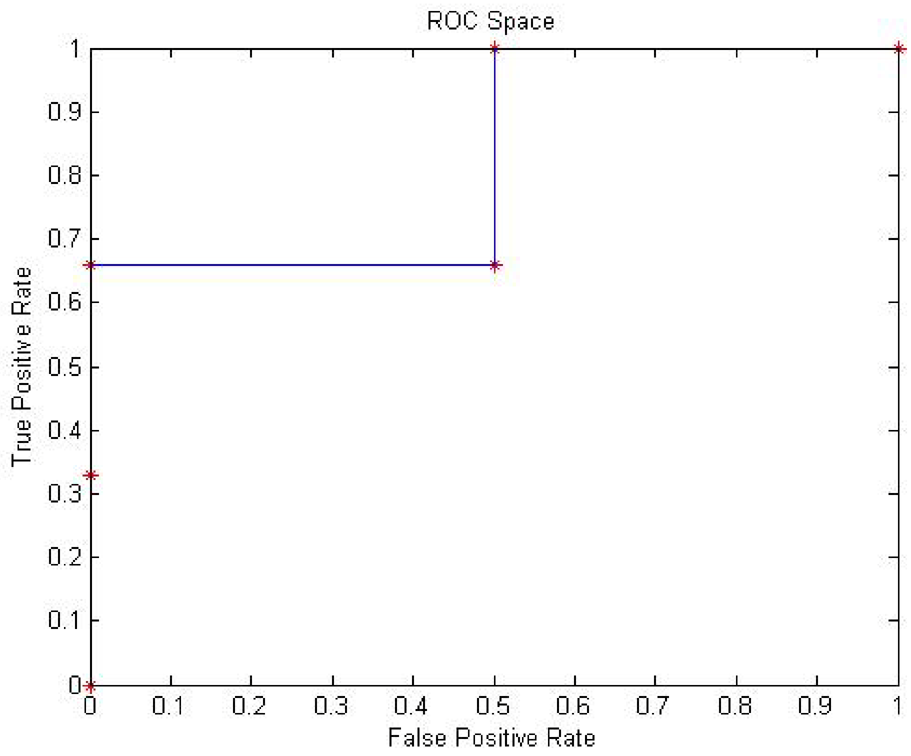
T=6, (sen, 1-spe) = (0.66, 0)

T=2, (sen, 1-spe) = (0.66, 0.5)

T=0, (sen, 1-spe) = (1, 0.5)

T=0, (sen, 1-spe) = (1, 1)

Points in ROC Space



- This is not a function!
- But. For the same FPR, we need to only consider the threshold/point with the largest sensitivity

ROC Curve

- The (best) points generated for all possible thresholds approximate the ROC Curve of the ranker (model)
- For an infinite number of points (i.e., assuming an infinitely large test set and a ranker that produces dense scores) we would get a curve

Generating the ROC Curve

- Naïve Method
  - For each possible threshold
    - Construct the confusion matrix
    - Generate a point in ROC space
    - For the same specificity / FPR keep only the point with the maximum sensitivity
  - Actually, need to try at most k+1 thresholds, where k is the number of test instances
  - Need to pay attention for equally scored instances

Generating the ROC Curve

- Sort all examples by decreasing score
- Say for some threshold k we have t true positives and n true negatives
  - $sen_k = t/pos$
  - $spe_k = n/pos$
- We try the next threshold moving one example down the list
- If the example is positive, the number of true positives will now be $t+1$
  - $sen_k = (t+1)/pos$
  - $spe_k = n/neg$
- The new point is $1/pos$ up from the previous
- Similarly, if the example is negative

Generating the ROC Curve

- Sort all examples by decreasing score
- First point p = (0,0)
- For each possible threshold *t*
  - if next example is positive move up by 1/pos
  - if next example is negative move right by 1/neg

- Pay attention to score ties:
  - Move diagonally

Judging a Model by Its ROC Curve

- What's the ROC of the optimal model?
- What's the ROC of the model providing random predictions?
- What's the ROC of a worse than random model?
  - How can we improve worse than random?

Comparing ROCs

- A very useful metric for evaluating models!
- Independent of the prior class distribution or the cost matrix

The Area Under the ROC Curve

- Called AUC
- All pos before neg: AUC = 1
- Random order : AUC = 0.5
  - the reverse does not hold!
- All neg before pos : AUC = 0

Statistical Interpretation of AUC

- equivalent to the probability that the model will rank a randomly chosen positive instance higher than a randomly chosen negative instance

Multi-Class Metrics

- Accuracy: straightforward to generalize
- AUC: difficult to generalize
  - Becomes Volume Under the ROC Surface
  - How to construct the surface has not been agreed upon
- Other metrics exist
  - Relative Classifier Information (relative reduction in entropy we achieve by using the model)
  - etc.

Τέλος Ενότητας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημειώματα

Σημείωμα αδειοδότησης (1)

Σημείωμα αδειοδότησης (2)

- Ως **Μη Εμπορική** ορίζεται η χρήση:
  - που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
  - που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
  - που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

- Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Κρήτης, Ιωάννης Τσαμαρδίνος 2015. «Μηχανική Μάθηση. Metrics of Performance». Έκδοση: 1.0. Ηράκλειο 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: https://opencourses.uoc.gr/courses/course/view.php?id=362.

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.