THE UNIVERSITY OF TENNESSEE KNOXVILLE

AICIP RESEARCH

COSC 522 – Machine Learning

Lecture 10 – Performance Evaluation

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
https://www.eecs.utk.edu/people/hairong-qi/
Email: hqi@utk.edu

Course Website: http://web.eecs.utk.edu/~hqi/cosc522/

Roadmap

- Supervised learning
  - Maximum Posterior Probability (MPP): For a given x, if $P(w_1|x) > P(w_2|x)$, then x belongs to class 1, otherwise 2.
    - Parametric Learning
      - Case 1: Minimum Euclidean Distance (Linear Machine), $\Sigma_i = \sigma^2 I$
      - Case 2: Minimum Mahalanobis Distance (Linear Machine), $\Sigma_i = \Sigma$
      - Case 3: Quadratic classifier, Σ_i = arbitrary
      - Estimate Gaussian parameters using MLE
    - Nonparametric Learning
      - Parzon window (fixed window size)
      - K-Nearest Neighbor (variable window size)
- Unsupervised learning
  - Non-probabilistic approaches
    - kmeans, wta
  - Hierarchical approaches
    - Agglomerative clustering
- Supporting preprocessing techniques
  - Dimensionality Reduction
    - Supervised linear (FLD)
    - Unsupervised linear (PCA)
    - Unsupervised nonlinear (t-SNE)
- Supporting postprocessing techniques
  - Classifier Fusion
  - Performance Evaluation
- Optimization techniques
  - Gradient Descent (GD)

Questions

AICIP RESEARCH

- What are TP, TN, FP, FN?
- What are sensitivity and specificity? What are their relationship to TP, TN, FP, and FN?
- What are precision and recall? What are their relationship to TP, TN, FP, and FN?
- What is confusion matrix? How can you derive TP, TN, FP, and FN from a confusion matrix?
- What is ROC curve?
- What does each axis mean? What are the relationship of the axes?
- How to draw an ROC curve?
- Which curve is optimal on an ROC plot?
- What is cross-validation? Why do we need it?

terminology

ROC

Cross validation

Performance Metrics

- Often used in automatic target recognition and medical diagnosis
- True positive
  - The object is there and our classifier says it is there
- True negative
  - The object is not there and our classifier says it is not there
- False negative (false misses)
  - The object is there and our classifier says it is not there
- False positive (false hits)
  - The object is not there and our classifier says it is there

Performance Metrics (cont'd)

AICIP RESEARCH

- Sensitivity
  - Probability of a true-positive = TP/(TP+FN)
- Specificity
  - Probability of a true-negative = TN/(TN+FP)
- The probability of a correct decision = (TP+TN)/N, where N is the total number of samples

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Performance Metrics (cont'd)

- Precision = TP/(TP+FP)
  - What proportion of positive identifications was actually correct?
- Recall = TP/(TP+FN)
  - What proportion of actual positives was correctly identified?
- Accuracy = (TP+TN)/(TP+TN+FP+FN)

Performance Metrics (cont'd)

- Confusion matrix
- Example: 3-class classification problem (AAV, DW, HMV)

	AAV	DW	HMV
AAV	894	329	143
DW	99	411	274
HMV	98	42	713

The real class is DW, the classifier says it' s HMV

Questions

AICIP RESEARCH

- What are TP, TN, FP, FN?
- What are sensitivity and specificity? What are their relationship to TP, TN, FP, and FN?
- What are precision and recall? What are their relationship to TP, TN, FP, and FN?
- What is confusion matrix? How can you derive TP, TN, FP, and FN from a confusion matrix?
- What is ROC curve?
- What does each axis mean? What are the relationship of the axes?
- How to draw an ROC curve?
- Which curve is optimal on an ROC plot?
- What is cross-validation? Why do we need it?

terminology

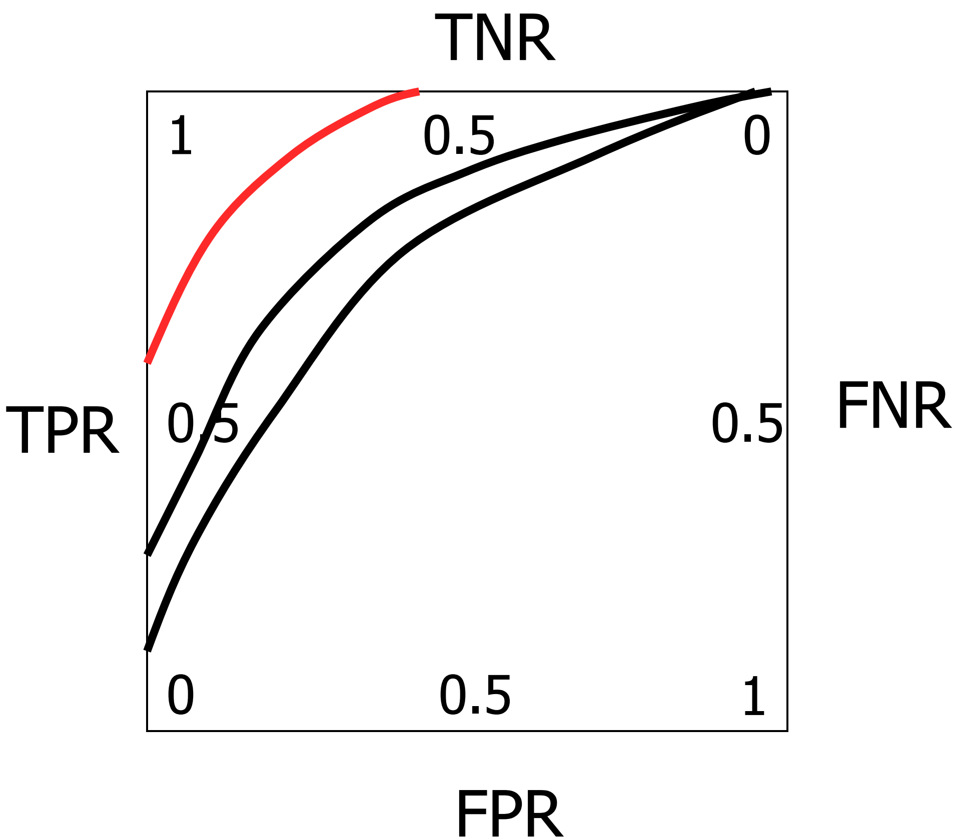
ROC

Cross validation

Parameters vs. Performance

- Once we have designed our classifier, we invariably have some parameters we'd like to adjust. E.g.,
  - Prior probability
- The optimal classifier is one with sensitivity as close to 100% as possible, and at the same time with specificity as close to 100% as possible

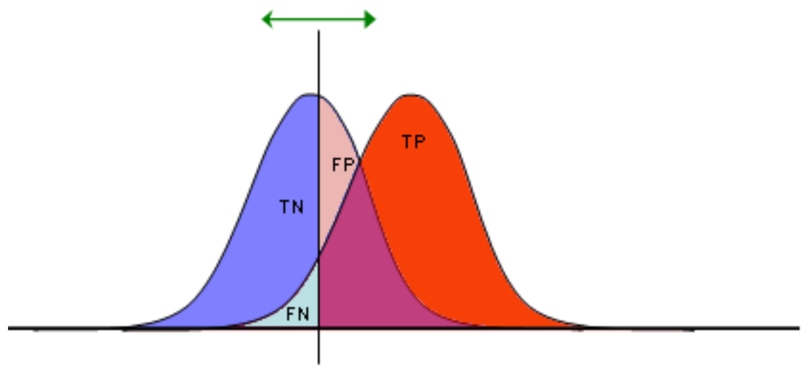
ROC – Receiver Operating Characteristic



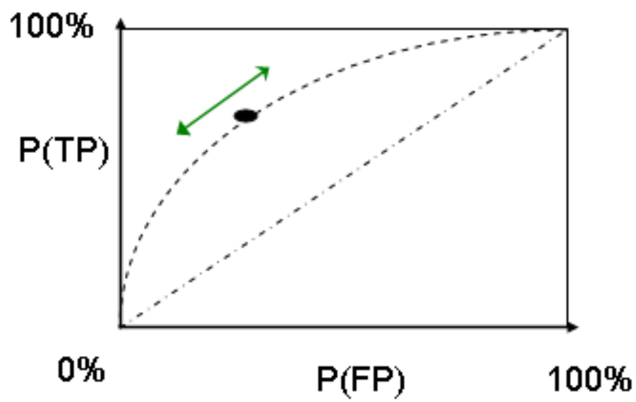
- Each curve represents the performance of a particular classifier as some parameter is varied over its range
- Of the three curves, the one with the sharpest bend, which passes closest to the upper left corner, is the best
- Calculate the area above the curve, the one with the smallest area is the best

- TPR: TP out of the total actual positives (Sensitivity or Recall)
- FPR: FP out of the total actual negatives (1-Specificity)

ROC (cont'd)



TP	FP
FN	TN
1	1



http://en.wikipedia.org/wiki/Receiver_operating_characteristic

Questions

AICIP RESEARCH

- What are TP, TN, FP, FN?
- What are sensitivity and specificity? What are their relationship to TP, TN, FP, and FN?
- What are precision and recall? What are their relationship to TP, TN, FP, and FN?
- What is confusion matrix? How can you derive TP, TN, FP, and FN from a confusion matrix?
- What is ROC curve?
- What does each axis mean? What are the relationship of the axes?
- How to draw an ROC curve?
- Which curve is optimal on an ROC plot?
- What is cross-validation? Why do we need it?

terminology

ROC

Cross validation

Determine ROC from Test Data

- Apparent error rate vs. true error rate
  - Apparent error rate: counting the number of elements in the training set that are misclassified. However, this error rate leads to an optimistic result since the classifier has been designed to minimize the number of misclassifications of the training set
- Solutions???

Solution 1 – Separating the training set and the test set

- Divide the training set into two parts (randomly) and build the classifier using half of the data, then test the classifier on the other half. This other half is called the validation set.
- Clarification: training set, test set, validation set

Solution 2 – The Leave-One-Out Approach

- Assume there are n points in the training set.
- Remove point 1 from the set and design the classifier (determine the pdf) using the other n-1 points. Then test the classifier on point 1.
- Repeat for all points.
- The resulting error rate can be shown to be an almost unbiased estimate of the expected true error rate
- This requires we design n classifiers. However, we only need to do it once.

m-Fold Cross Validation

AICIP RESEARCH

- A generalization to both solution 1 and 2.
- The training set is randomly divided into m disjoint sets of equal size. The classifier is trained m times, each time with a different set held out as a validation set
- For example, when m = 3,
  - m1+m2 as training, test on m3
  - m1+m3 as training, test on m2
  - m2+m3 as training, test on m1
- When m=2, it is solution 1
- When m=n, it is solution 2 (n is the number of samples in the original training set)