STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY

CPE/EE 695: Applied Machine Learning

Lecture 3 – 2: Evaluating Classifiers

Dr. Shucheng Yu, Associate Professor
Department of Electrical and Computer Engineering
Stevens Institute of Technology

Classification measures

- Accuracy is only one measure (error = 1-accuracy).

  Accuracy is not always suitable

  - In text mining, we may only be interested in the documents of a particular topic, which are only a small portion of a big document collection.
  - In classification involving skewed or highly imbalanced data, e.g., network intrusion and financial fraud detections, we are interested only in the minority class.
    - High accuracy does not mean any intrusion is detected.
    - E.g., 1% intrusion. Achieve 99% accuracy by doing nothing.
- The class of interest is commonly called the **positive class**, and the rest **negative classes**.

Performance Measures of Classifiers

Confusion Matrix

Sometimes we need both a high accuracy rate and a low error rate. ***Confusion matrix*** is an important tool to measure both accuracy rates and error rates.

	Classified Positive	Classified Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

True Positive (TP): the number of true positive examples
True Negative (TN): the number of true negative examples
False Positive (FP): the number of false positive examples (type I error)
False Negative (FN): the number of false negative examples (type II error)

Performance Measures of Classifiers

Precision, Recall and F_1-Score

We are often interested in two measures on relevance:

- ***precision*** $p = \frac{TP}{TP+FP}$, is the number of true positive divided by all classified as positive.
- ***recall*** *(a.k.a.,* ***sensitivity****)* $r = \frac{TP}{TP+FN}$, is the number of true positive divided by all positive.
- $F_1 = \frac{2pr}{p+r}$, is a measure of both p and r. F_1-score is large only if both p and r are large.

	Classified Positive	Classified Negative
Actual Positive	1	99
Actual Negative	0	200

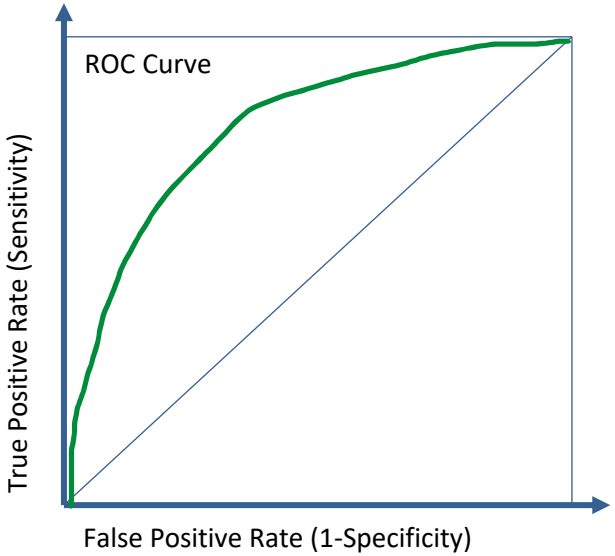
"**Precision** can be seen as a measure of quality and **recall** as a measure of quantity.

- Higher precision means that an algorithm returns *more relevant results* than irrelevant ones,
- high recall means that an algorithm returns *most of the relevant results*." (Wikipedia)

Performance Measures of Classifiers

Receiver Operating Characteristic (ROC) Curve

Another common measure is the ***Receiver Operating Characteristic*** (ROC) curve, which shows the tradeoff between **True Positive Rate** (i.e., ***sensitivity***) and **False Positive Rate** (i.e., ***1 - specificity***).



- The closer to the top-left corner of the ROC space, the better the classifier is.
- The closer to the 45-degree diagonal of the ROC space, the worse the classifier is.
- The area under the curve (AUC) is usually used to measure above properties.
- ROC/AUC are usually used to measure ***how well a binary classifier distinguish the two classes***.

STEVENS

INSTITUTE *of* TECHNOLOGY

THE INNOVATION UNIVERSITY®

stevens.edu