# Machine Learning
# CSE - 465

## Lecture - 10

# Lecture 10
# Classifier Performance Evaluation

# Outline

- Model evaluation & selection
- Evaluating classifier accuracy
- Metrics for evaluating classifier performance
  - Confusion matrix
- Plotting an ROC curve

# Model Evaluation & Selection

- How can we measure accuracy?
  - Use **test set** of class-labeled tuples instead of training set when assessing accuracy
- Methods for estimating a classifier's accuracy:
  - Holdout method, random sub-sampling
  - Cross-validation
  - Bootstrap
- Comparing classifiers:
  - Confidence intervals
  - Cost-benefit analysis and ROC Curves

# Evaluating Classifier Accuracy

- **Holdout method**
  - Given data is randomly partitioned into two independent sets
    - Training set (e.g., 2/3) for model construction
    - Test set (e.g., 1/3) for accuracy estimation
  - Random sampling: a variation of holdout
    - Repeat holdout k times, accuracy = avg. of the accuracies obtained

# Evaluating Classifier Accuracy

- **Cross-validation**
  - In k-fold cross-validation, the initial data are randomly partitioned into k mutually exclusive subsets or "folds," $D_1$, $D_2$,..., $D_k$, each of approximately equal size. Training and testing is performed k times. In iteration i, partition $D_i$ is reserved as the test set, and the remaining partitions are collectively used to train the model. That is, in the first iteration, subsets $D_2$,..., $D_k$ collectively serve as the training set to obtain a first model, which is tested on $D_1$; the second iteration is trained on subsets $D_1$, $D_3$,..., $D_k$ and tested on $D_2$; and so on.

# Metrics for Evaluating Classifier Performance

- True positives (TP)
  - These refer to the positive tuples that were correctly labeled by the classifier. Let TP be the number of true positives.

- True negatives(TN)
  - These are the negative tuples that were correctly labeled by the classifier. Let TN be the number of true negatives.

# Metrics for Evaluating Classifier Performance

- False positives (FP)
  - These are the negative tuples that were incorrectly labeled as positive (e.g., tuples of class buys computer = no for which the classifier predicted buys computer = yes). Let FP be the number of false positives.

- False negatives (FN)
  - These are the positive tuples that were mislabeled as negative (e.g., tuples of class buys computer = yes for which the classifier predicted buys computer = no). Let FN be the number of false negatives.

# Metrics for Evaluating Classifier Performance

- Confusion Matrix
  - A confusion matrix (also known as an error matrix) is a table that is often used to describe the performance of a classification model (or classifier) on a set of test data for which the true values are known. It allows the visualization of the performance of an algorithm
  - A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix.

# Metrics for Evaluating Classifier Performance

- Confusion Matrix
  - The confusion matrix shows the ways in which your classification model is confused when it makes predictions
  - It gives us insight not only into the errors being made by a classifier but more importantly the types of errors that are being made
  - TP and TN tell us when the classifier is getting things right, while FP and FN tell us when the classifier is getting things wrong

# Metrics for Evaluating Classifier Performance

**Confusion Matrix**

| | | Predicted class | | |
|---|---|---|---|---|
| | | *yes* | *no* | Total |
| **Actual class** | *yes* | *TP* | *FN* | *P* |
| | *no* | *FP* | *TN* | *N* |
| | Total | *P'* | *N'* | *P + N* |

| *Classes* | *buys_computer = yes* | *buys_computer = no* | *Total* |
|---|---|---|---|
| *buys_computer = yes* | **6954** | **46** | 7000 |
| *buys_computer = no* | **412** | **2588** | 3000 |
| Total | 7366 | 2634 | 10,000 |

# Metrics for Evaluating Classifier Performance

| *Measure* | *Formula* |
|---|---|
| accuracy, recognition rate | $\frac{TP+TN}{P+N}$ |
| error rate, misclassification rate | $\frac{FP+FN}{P+N}$ |
| sensitivity, true positive rate, recall | $\frac{TP}{P}$ |
| specificity, true negative rate | $\frac{TN}{N}$ |
| precision | $\frac{TP}{TP+FP}$ |
| $F$, $F_1$, $F$-score, harmonic mean of precision and recall | $\frac{2 \times precision \times recall}{precision + recall}$ |
| $F_\beta$, where $\beta$ is a non-negative real number | $\frac{(1+\beta^2) \times precision \times recall}{\beta^2 \times precision + recall}$ |

# Plotting an ROC curve

| Tuple # | Class | Prob. | TP | FP | TN | FN | TPR | FPR |
|---|---|---|---|---|---|---|---|---|
| 1 | *P* | 0.90 | 1 | 0 | 5 | 4 | 0.2 | 0 |
| 2 | *P* | 0.80 | 2 | 0 | 5 | 3 | 0.4 | 0 |
| 3 | *N* | 0.70 | 2 | 1 | 4 | 3 | 0.4 | 0.2 |
| 4 | *P* | 0.60 | 3 | 1 | 4 | 2 | 0.6 | 0.2 |
| 5 | *P* | 0.55 | 4 | 1 | 4 | 1 | 0.8 | 0.2 |
| 6 | *N* | 0.54 | 4 | 2 | 3 | 1 | 0.8 | 0.4 |
| 7 | *N* | 0.53 | 4 | 3 | 2 | 1 | 0.8 | 0.6 |
| 8 | *N* | 0.51 | 4 | 4 | 1 | 1 | 0.8 | 0.8 |
| 9 | *P* | 0.50 | 5 | 4 | 0 | 1 | 1.0 | 0.8 |
| 10 | *N* | 0.40 | 5 | 5 | 0 | 0 | 1.0 | 1.0 |

# Plotting an ROC curve

- Receiver Operating Characteristics Curve
- The figure shows the probability value (column 3) returned by a probabilistic classifier for each of the 10 tuples in a test set, sorted by decreasing probability order. Column 1 is merely a tuple identification number, which aids in our explanation. Column 2 is the actual class label of the tuple. There are five positive tuples and five negative tuples, thus P = 5 and N = 5. As we examine the known class label of each tuple, we can determine the values of the remaining columns, TP, FP, TN, FN, TPR, and FPR.

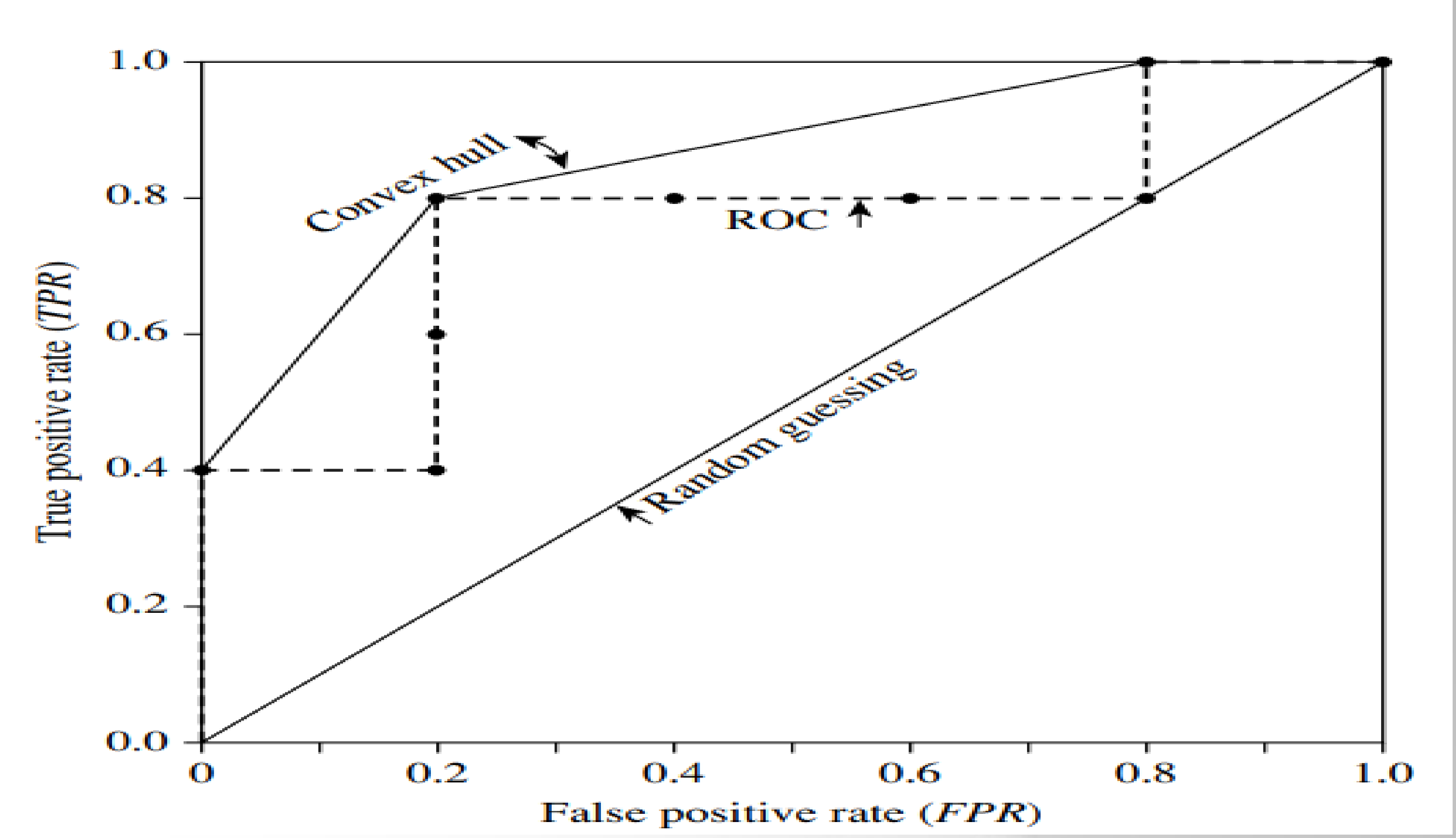
# Plotting an ROC curve

- We start with tuple 1, which has the highest probability score, and take that score as our threshold, that is, t = 0.9. Thus, the classifier considers tuple 1 to be positive, and all the other tuples are considered negative. Since the actual class label of tuple 1 is positive, we have a true positive, hence TP = 1 and FP = 0.
- Among the remaining nine tuples, which are all classified as negative, five actually are negative (thus, TN = 5). The remaining four are all actually positive, thus, FN = 4. We can therefore compute TPR = TP/P = 1/5 = 0.2, while FPR = 0. Thus, we have the point (0.2,0) for the ROC curve.

# Plotting an ROC curve

- Next, threshold t is set to 0.8, the probability value for tuple 2, so this tuple is now also considered positive, while tuples 3 through 10 are considered negative. The actual class label of tuple 2 is positive, thus now TP = 2. The rest of the row can easily be computed, resulting in the point (0.4,0).
- Next, we examine the class label of tuple 3 and let t be 0.7, the probability value returned by the classifier for that tuple. Thus, tuple 3 is considered positive, yet its actual label is negative, and so it is a false positive. Thus, TP stays the same and FP increments so that FP = 1. The rest of the values in the row can also be easily computed, yielding the point (0.4,0.2).

# Plotting an ROC curve



# Thank You