

# Analytical Thinking

ISTE-600 Week 14

Final Exam Review

# Week 01

- Thinking
  - Elements & Standards
- Analytical Thinking
- Synthetical Thinking
- Systemic Thinking
- Critical Thinking
- Creative Thinking

- Statistical Thinking
- Visual Thinking
- Computational Thinking
- ....

# Week 02: Intro to DM

- Why Mine Data?
- What is Data Mining?
- Data Mining Process
- Data Mining Techniques
  - Directed/Supervised/Predictive
  - Undirected/Unsupervised/Descriptive
- Weka

# Week 3: Data & Data Processing

- Lesson 1: Data
  - What is Data?
    - Types of Attributes
    - Discrete & Continuous
  - Data Quality
    - Noise; Outliers; Missing Values; Duplicate data
- Lesson 2: Data Preprocessing
  - Aggregation
  - Sampling
  - Dimension Reduction
  - Feature Selection
  - Discretization
- Lesson 3: Weka
  - Algorithms for Data Quality
  - Algorithms for Data Preprocessing

# Week 03: Data Exploration

- Summary Statistics
  - Frequency & Mode
  - Measure of Location
  - Measure of Spread
- Visualization
  - Representation; Arrangement; Selection
  - Histogram; Box Plot; Scatter/Contour Plots
- Online Analytical Processing (OLAP)
  - Data Cube
  - OLAP Operations
    - Slicing/Dicing
    - Roll-up & Drill-down

# Week 4/5. Classification: Basic Concepts & Techniques

## Lesson 1. Classification

- Basic Concepts
- Decision Tree
  - Hunt's Algorithm
  - Tree Induction Issues
    - Stopping Criteria
    - Best Split

## Lesson 2. Practical Issues of Classification

- Underfitting & Overfitting
- Confusion Matrix
- Cost of classification

## Lesson 3. Model Evaluation

- Metrics for Performance Evaluation
- Methods for Performance Evaluation
- Methods for Model Comparison

Weka: J48 (C4.5)

# Week 6/7: Alternative Classifications

1. Rule-Based Classifier
2. Nearest Neighbor Classifiers
3. Bayesian Classifiers
4. Artificial Neural Networks (ANN)
5. Support Vector Machines
6. Ensemble Methods

# Week 8/10: Association Rule Mining

## Lesson 1. Association Rules Basics

- Association Rules
- Frequent Itemsets
- Mining Association Rules: 2-step approach

## Lesson 2. Apriori Algorithm

- Apriori Principle
  - Support Counting
    - » Using a Hash Tree
  - Rule Generation



# Week 11/12 Clustering

## Lesson 1. Partitional Clustering

- Overview: Cluster Analysis
- Partitional Clustering
  - K-means Clustering

## Lesson 2. Hierarchical Clustering

- Agglomerative (bottom-up)
  - ◆ Single-linkage
  - ◆ Complete-linkage
  - ◆ Group Average
- Divisive (top-down)

## Lesson 3. Cluster Validity

- External Index:
  - ◆ Entropy
- Internal Index:
  - ◆ Sum of Squared Error (SSE)
- Relative Index
  - ◆ SSE or entropy

# Week 13 Anomaly Detection

- What are anomalies/outliers?
- Types of anomaly detection schemes
  - Graphical
    - Boxplot (1-D), Scatter plot (2-D)
    - Convex Hull Method
  - Statistical-based
    - Data distribution/Parameter of distribution
    - Number of expected outliers (confidence limit)
    - Likelihood Approach
  - Distance-based
    - Nearest-neighbor based
    - Density based
    - Clustering based