

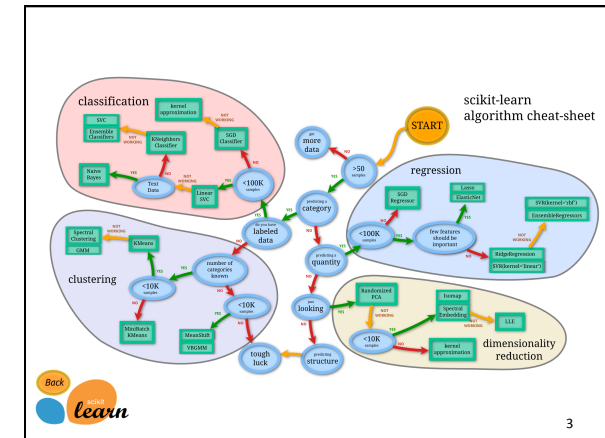
## INFO411 – Review

3/10/2017

## Final Exam

- 40% of the total
- 3-hours
- Approved calculators allowed (but No need)
- Two sections
- What is assessed:
  - Conceptual understanding
  - Ability to apply knowledge to problem solving

2



3

## INFO 411 – Overview

- Principles
- Key algorithms
- Differences & Connections
- Subtleties

4

## Dimension Reduction

- Feature Selection vs. Feature Extraction
- Selection: approaches
  - Wrapper / Filter / Embedded
- Extraction / projection
  - PCA / KPCA
  - Random projection
  - Isomap etc.

5

## Clustering

- Unsupervised learning
- Approaches
- K-means, E-M
  - Initializations
  - Connections
- Subtleties:
  - $k$ , initialization, other parameters, distance metrics

6

## Classification

- Approaches
  - Parametric vs non-parametric
- Key algorithms
- Connections & extensions
  - $k$ -NN, Bayes
  - One-class classification (outlier detection)
  - Class imbalance

7

## Online Learning

- Principles
- Algorithms
  - Variations to basic competitive learning
- Applications
  - Vector quantization
  - Video foreground-background segmentation
  - Anomaly detection

8

## Performance Evaluation of Classifiers

- Criteria for algorithm performance Training/Testing Schemes
- Confusion Matrix
- K-Fold CV Paired  $t$ -Test
- Comparing more than two classifiers
- Shortcomings of the basic scalar evaluation metrics
- ROC Curves and their construction
- Area Under (the) Curve (AUC)
- Problems with AUC

9

## Evaluation and model selection

- Bias vs Variance
- Metrics for performance evaluation
- Random splits vs LOOC vs  $k$ -fold CV
- Criteria for model selection:
  - Classification
  - Clustering

10

## Classifier Combination

- Voting
- Bagging
- Boosting:
  - AdaBoost
- Stacking
- Fine-tuning ensembles
- Mixture of experts
- Cascading
- What is gained?

11

## Swarm Optimizers

- Particle swarm optimization
  - Principles
  - Drawbacks
  - Possible improvements
  - Applications
- Ant colony: basic ideas

12