CS 551: Pattern Recognition

Spring 2009

Description: This course concentrates on statistical pattern recognition techniques. We will talk about Bayesian decision theory, parametric and non-parametric density estimation, feature reduction and selection, supervised and unsupervised learning, classifiers and discriminant functions, clustering and ensemble methods. We will also introduce structural and syntactic pattern recognition at the end of the semester.

Instructor: Selim Aksoy (Office: EA 423, Email: saksoy@cs.bilkent.edu.tr)

Schedule: Mon 8:40–10:30, Wed 10:40–11:30 (EA 502)

Web page: http://www.cs.bilkent.edu.tr/∼saksoy/courses/cs551/index.html

Prerequisites: Probability theory, statistics, linear algebra

Texts:

- R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, 2nd edition, John Wiley & Sons, Inc., 2000. (required)
- S. Theodoridis, K. Koutroumbas, *Pattern Recognition*, 3rd edition, Academic Press, 2006.
- C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- A. Webb, *Statistical Pattern Recognition*, 2nd edition, John Wiley & Sons, Inc., 2002.
- T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer, 2003.
- K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1990.
- R. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches*, John Wiley & Sons, Inc., 1992.
- A. K. Jain, R. C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, 1988.

Grading (tentative):

Homework:	50%
Midterm exam:	20%
Final exam:	25%
Class participation:	5%

Exams: There will be one midterm exam and one final exam.

Assignments: There will be regular homework assignments that will involve both programming and essay type questions.

Lecture Schedule:

- Introduction to Pattern Recognition (DHS Ch 1, Appendix A.1–A.2, A.4–A.5)
  - Pattern recognition systems
  - The design cycle
  - An example
- Bayesian Decision Theory (DHS Ch 2.1–2.9; skip 2.3.1, 2.3.2)
  - Modeling using continuous and discrete features
  - Discriminant functions
  - The Gaussian density
  - Error estimation
- Parametric Models (DHS Ch Ch 3.1–3.5, 3.9, 10.2–10.4, 3.10, 2.11; skip 10.4.3, 10.4.4)
  - Maximum-likelihood estimation
  - Bayesian estimation
  - Expectation-Maximization and mixture density estimation
  - Hidden Markov Models
  - Bayesian Belief Networks
- Non-parametric Methods (DHS Ch 4.1–4.4; skip 4.3.5, 4.3.6)
  - Density estimation
  - Histogram-based estimation
  - Parzen windows estimation
  - Nearest neighbor estimation
- Feature Reduction and Selection (DHS Ch 3.7–3.8, 10.13–10.14)
  - Problems of dimensionality
  - Component analysis
    - Principal components analysis (PCA)
    - Linear discriminant analysis (LDA)
  - Feature selection
- Non-Bayesian Classifiers (DHS Ch 4.5–4.6, 5.1–5.3, 5.11, 6.1–6.3, 8.1–8.3; skip 4.5.1–4.5.3, 4.5.5)
  - k-nearest neighbor classifier
  - Linear discriminant functions
  - Support vector machines
  - Neural networks
  - Decision trees
- Unsupervised Learning and Clustering (DHS Ch 10.1, 10.6–10.7, 10.9–10.10, 10.12)
  - Criterion functions for clustering
  - k-means clustering
  - Hierarchical clustering
  - Graph-theoretic clustering
  - Cluster validity
- Algorithm-Independent Learning Issues (DHS Ch 9.1–9.2, 9.5–9.7; skip 9.2.2)
  - No Free Lunch Theorem
  - Resampling for classifier design
  - Comparing classifiers
  - Combining classifiers
- Structural and Syntactic Pattern Recognition (DHS Ch 8.5–8.6)
  - Recognition with strings
  - Grammatical methods
  - Graph-theoretic methods