



# Conclusions

---

Larry Holder  
CptS 570 – Machine Learning  
School of Electrical Engineering and Computer Science  
Washington State University



# Outline

---

- Overview of machine learning
- Fundamental research issues
- Grand challenge problems



# Overview of Machine Learning

---

- Supervised learning
- Evaluation of learning methods
- Learning theory
- Unsupervised learning
- Other learning methods
- Applications
- Related fields



# Supervised Learning

---

- Traditional methods
  - Version space
  - Candidate elimination algorithm
  - Decision tree induction
  - Neural networks
  - Bayesian learning
  - Instance-based learning



# Supervised Learning

---

- Advanced methods
  - Kernel methods
    - Support vector machines
  - Ensembles
    - Bagging
    - Boosting
  - Learning rule sets
    - Relational learning
    - Inductive logic programming (ILP)
    - Graph-based learning



# Evaluation of Learning Methods

---

- True error vs. sample error
- Bounding true error
- Comparison of hypotheses
- Comparison of learners
- Significance testing
- ROC curves



# Learning Theory

---

- Bayes optimal learning
- Sample complexity
- PAC learning framework
- VC dimension



# Unsupervised Learning

---

- Non-linear regression
- Pattern discovery
- Clustering
- Grammar (language) learning
- EM algorithm



# Other Learning Methods

---

- Genetic algorithms
- Analytical learning
- Reinforcement learning
- Integrated learning



# Applications

---

- Classification and prediction
  - Chemical properties
  - Biometrics
  - Object recognition
  - Organizational and behavioral patterns
- Skill acquisition
  - Robot navigation
  - Control and optimization
  - Heuristic search



# Related Fields

---

- Statistics
- Pattern recognition
- Control theory
- Cognitive science
- Psychology
- Neurophysiology



# Fundamental Research Issues

---

- General learning methods
- Limits of general methods
- Theory and principles guiding development of domain-specific learning algorithms
- Multi-relational learning
- Learning in dynamic environments
- Incorporation of domain-specific background knowledge
- Ethical responsibility and privacy



# Grand Challenge Problems

---

- “What are the Grand Challenges for Data Mining,” *SIGKDD Explorations*, 8(2):70-77, 2006.
- KDD 2006 conference panel
- G. Piatetsky-Shapiro, C. Djeraba, L. Getoor, R. Grossman, R. Feldman, M. Zaki
- GC problems define directions for the field and motivate and excite researchers
- E.g., Netflix Prize



# Good Grand Challenge Problems

---

- Problem is hard – very difficult to solve given the current state of the art
- Based on a large, publicly available data set
- There is a specific goal – it is clear when the problem is solved
- Problem is interesting to researchers and understandable to the public; preferably stated in one sentence
- There is significant public benefit if it is solved



# Grand Challenge Problem (1)

---

- Automatically annotate 1000 hours of digital video in 1 hour
- E.g., “basketball game”, “Michael Jordan”
- General approach
  - Automatically extract primitive features
  - Manually annotate subset of videos
  - Learn to predict annotations based on features
  - Use learned classifiers to annotate subsequent videos



# Grand Challenge Problem (2)

---

- Functional annotation of the proteome, the set of proteins in the cell
  - What is the function of a protein (e.g., insulin production, metabolism)?
  - What other proteins does it interact with?
- 100,000+ proteins, some with multiple functions
- Approach: Link mining, “guilt” by association



# Grand Challenge Problem (3)

---

- System capable of passing SAT reading comprehension test given access to the World-Wide Web
- Approach
  - Entity and relation extraction
  - Natural language understanding
  - Relational rule learning
  - Reasoning
- Automated student



# Conclusions

---

- Machine learning seeks to give computers the ability to improve their performance based on experience
- Many mature methods available and some theoretical results
- Basis of multi-billion dollar data mining industry
- Much research left to be done