14. Summary and Perspectives

- Formulations of Learning Problem
- Hypothesis Representations
- Search Methods
- Perspectives: Learning as ...
- What is state of the art?
- Where Is this Headed?

Formulations of Learning Problem

- Improve performance P at task T from experience E
- Learn function $f(x)$ given $\{\langle x_i, f(x_i)\rangle\}$
  - Boolean-valued, discrete-valued, real-valued, or vector-valued f
  - From discrete or real-valued inputs x
  - Noisy data, or noise-free (x or $f(x)$)
  - From randomly generated examples, or from active queries
- Learn from delayed reward (e.g., reinforcement learning)
- Learn from $\{\langle x_i, f(x_i)\rangle\}$ plus prior knowledge
- Learn from unclassified examples (clustering)

Formulation influences choice of hypothesis space, algorithm, and constrains its sample complexity

Hypothesis Representations

- Symbolic/discrete
  - Propositional (variable-free)
  - First order (variables)
- Decision trees
- Neural networks (and other parameterized expressions)
- Explicit probabilities (e.g., $P(a_i = v_j|h)$)
- Instances (e.g., nearest neighbor)

Search Methods/Hypothesis Space Structure

- Specific-to-general, general-to-specific (e.g., VS)
- Simple-to-complex (e.g., C4.5)
- Gradient descent (e.g., Backpropagation)
- Greedy divide and conquer (e.g., C4.5)
- Greedy covering (e.g., FOIL)
- Randomized search (e.g., GA's)
- Invert deduction (e.g., CIGOL)

Perspectives: Learning as ...

- Search in hypothesis space
- Efficiently approximating Bayes optimal classifications
- An optimization problem (to optimize what?)
- Combining inductive and deductive reasoning
- Statistical estimation
- Futile without inductive bias

Frameworks for Characterizing Learners

- Inductive bias ($data + bias \vdash classifications$)
- Equivalent MAP learners (assumed priors $P(h)$ and $P(D|h)$)
- Sample complexity (and computational complexity)
- Mistake bounds

What is state of the art?

Practical approaches:

- Backpropagation, regression, decision tree and rule learning, kNN, naive Bayes, ..., are robust, useful, commercialized
- More specialized algorithms can outperform these
- Best algorithm? Depends on target fn., data
- Cross validation important for choosing learning parameters

Many successful applications:

- Find database regularities: credit risk evaluation
- Classify text: NewsWeeder
- Classify images: face recognition
- Robot control: autonomous driving
- Game playing: Backgammon

Where Is this Headed?

Today: tip of the iceberg

- First-generation algorithms: neural nets, decision trees, regression ...
- Applied to well-formated data
- Budding industry

Opportunity for tomorrow: enormous impact

- Learn across multiple mixed-media databases, including the web and newsfeeds
- Learn by active experimentation
- Architectures for learning, perception, action
- Cumulative, lifelong learning
- Synergy between studies of biological and artificial learning
- Learning embedded in programing languages?