MACHINE LEARNING

Introduction

1

Learning Agents

- AI is the enterprise of design and analysis of intelligent agents.
- Intelligent behavior requires knowledge (e.g., model of the environment)
- Explicitly specifying the knowledge needed for specific tasks is hard, and often infeasible
- How to acquire knowledge?
  – Learning

2

Learning Agents

- Learning modifies the agent's decision mechanisms to improve performance
- Environment changes over time – adapt to changes
- Learning is essential for unknown environments,
  - i.e., when designer lacks omniscience

Why study machine learning

- Applications
  - Medical diagnosis/image analysis (e.g. pneumonia, pap smears)
  - DNA sequence identification
  - Scientific discovery
  - Spam Filtering, fraud detection (e.g. credit cards, phone calls)
  - Search and recommendation (e.g. google, amazon)
  - Automatic speech recognition & speaker verification
  - Locating/tracking/identifying objects in images & video (e.g. faces)
  - …

Data Mining

- Huge amounts of data available
  - Sources: business, science, medicine, economics, geography, environment, sports, …
- Data is a potentially valuable resource
- Raw data is useless: need techniques to automatically extract information from it
  - Data: recorded facts
  - Information: patterns underlying the data
- Machine learning techniques: automatically finding patterns in data

5

Examples: The weather problem

- Conditions for playing a certain game

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	Normal	False	Yes
…	…	…	…	…

- Classification rule:

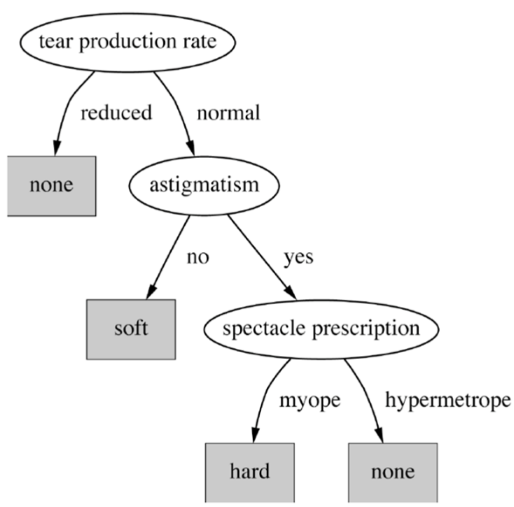
```
If outlook = sunny and humidity = high then play = no
If outlook = rainy and windy = true then play = no
If outlook = overcast then play = yes
If humidity = normal then play = yes
If none of the above then play = yes
```

6

The contact lenses data

Age	Spectacle prescription	Astigmatism	Tear production rate	Recommended lenses
Young	Myope	No	Reduced	None
Young	Myope	No	Normal	Soft
Young	Myope	Yes	Reduced	None
Young	Myope	Yes	Normal	Hard
Young	Hypermetrope	No	Reduced	None
Young	Hypermetrope	No	Normal	Soft
Young	Hypermetrope	Yes	Reduced	None
Young	Hypermetrope	Yes	Normal	hard
Pre-presbyopic	Myope	No	Reduced	None
Pre-presbyopic	Myope	No	Normal	Soft
Pre-presbyopic	Myope	Yes	Reduced	None
Pre-presbyopic	Myope	Yes	Normal	Hard
Pre-presbyopic	Hypermetrope	No	Reduced	None
Pre-presbyopic	Hypermetrope	No	Normal	Soft
Pre-presbyopic	Hypermetrope	Yes	Reduced	None
Pre-presbyopic	Hypermetrope	Yes	Normal	None
Presbyopic	Myope	No	Reduced	None
Presbyopic	Myope	No	Normal	None
Presbyopic	Myope	Yes	Reduced	None
Presbyopic	Myope	Yes	Normal	Hard
Presbyopic	Hypermetrope	No	Reduced	None
Presbyopic	Hypermetrope	No	Normal	Soft
Presbyopic	Hypermetrope	Yes	Reduced	None
Presbyopic	Hypermetrope	Yes	Normal	None

A decision tree for this problem



Predicting CPU performance

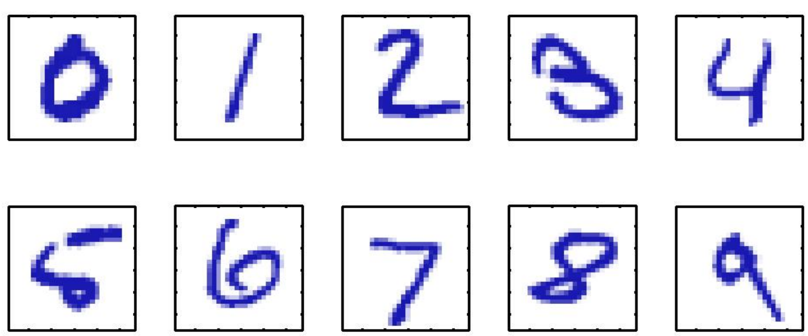
- Example: 209 different computer configurations

	Cycle time (ns)	Main memory (Kb)		Cache (Kb)	Channels		Performance
	MYCT	MMIN	MMAX	CACH	CHMIN	CHMAX	PRP
1	125	256	6000	256	16	128	198
2	29	8000	32000	32	8	32	269
…							
208	480	512	8000	32	0	0	67
209	480	1000	4000	0	0	0	45

- Linear regression function

```
PRP = -55.9 + 0.0489 MYCT + 0.0153 MMIN + 0.0056 MMAX
        + 0.6410 CACH - 0.2700 CHMIN + 1.480 CHMAX
```


Handwritten Digit Recognition



Supervised Learning Problem

Machine learning approach to Handwritten Digit Recognition

- Input to learning: a set of labeled instances -> dataset
- Instance
  - Individual, independent example of target concept
  - Characterized by a predetermined set of attributes/features
  - Represented as a feature vector **x**=(X1, X2, …)
- A training example is a pair (**x**, t), t: *target value* (category of the digit)
- *Training dataset* $\{(x_i,t_i)\}$

11

Supervised Learning (Cont.)

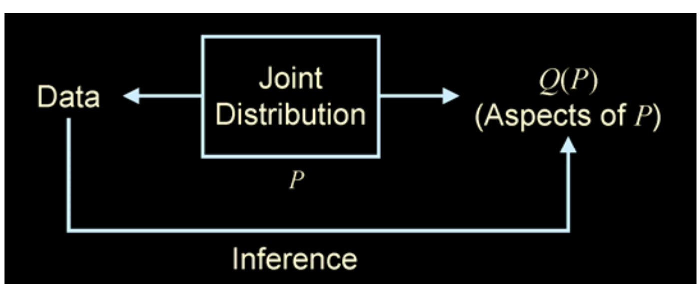
- Learning/training phase: find a model h(**x**)
- Choose model/hypothesis space
  - Linear models
  - Decision trees
  - Support vector machines
  - Neural networks
  -
- Problem: find the best model/hypothesis *h(**x**)*
  - given the training dataset
- Use learned h(**x**) to categorize/predict new instances

12

Canonical Learning Problems

- *Supervised Learning*: given examples of inputs and corresponding desired outputs, predict outputs on future inputs.
  - *Classification*: target has finite domain - categories
  - *Regression*: target has continuous domain
- *Unsupervised Learning*: given only inputs, automatically discover representations, features, structure, etc.
  - *Clustering*: group similar instances, e.g. automatically group (unlabeled) handwritten digits
- *Reinforcement Learning*: occasional rewards or punishments
  - how an agent ought to take actions in an environment so as to maximize some notion of long-term reward in sequential decision problems, e.g., learn to play chess without human instruction

Mainstream Machine Learning



Learning as Bayesian Inference

- Formulate the learning task as a process of probabilistic inference
  - Inference step: determine P(x,t) from data
  - Decision step: for given x, determine optimal t.
- *Bayesian Decision Theory*
  - A fundamental statistical approach to the problem of pattern recognition and machine learning
- Bayesian framework provides a sound probabilistic basis for understanding many learning algorithms and designing new algorithms

ML Models

- Supervised learning
  - Naive Bayes Classifier
  - Nearest neighbor methods
  - Linear models
  - Decision trees
  - (Deep) Neural networks
  - Support vector machines
  - Ensemble learning

ML Models

- Probabilistic graphical models: Bayesian networks, Markov random fields
- Unsupervised learning:
  - Clustering: mixture models, K-means, Hierarchical Clustering
  - PCA, ICA
- Sequential data: HMMs, Recurrent Neural Networks
- Markov decision process and Reinforcement Learning

17