University of Colorado
Boulder

CSCI 4502/5502
Data Mining

Fall 2019
Lecture 13 (Oct 15)

Course Project (1)

- Interesting project ideas, wide variety
- Subtasks, pipelines
  - data collection, understand your data, preprocessing, management, mining, evaluation, visualization
- Analytical thinking
  - potential applications, subset selection, spatial-temporal, general trends vs. anomaly, efficiency & scalability, challenges, lessons, …

Course Project (2)

- Continue working on your project
  - project checkpoint: Week 12 (11/12, 11/14)
  - fall break: Week 14 (11/26, 11/28)
  - project final report: Week 16 (12/10, 12/12)
- Individual contributions
- Team work
- Communication among groups
- Have fun!

Reminders

- Homework 4
  - posted at moodle
  - due at 9:30am, Thursday, Oct 17
- Midterm schedule
  - Week 8 & 9: classification, clustering, outlier
  - Week 10: midterm review, sample exam questions
  - Thursday, Oct 31: in-class midterm exam

Review

- Chapter 8: Classification
  - Basic concepts
  - Decision tree induction
  - Bayesian classification
  - Rule-based classification
  - Model evaluation and selection
  - Improve classification accuracy
  - Summary

Chapter 9: Classification: Advanced Methods

Chap 9: Advanced Classification

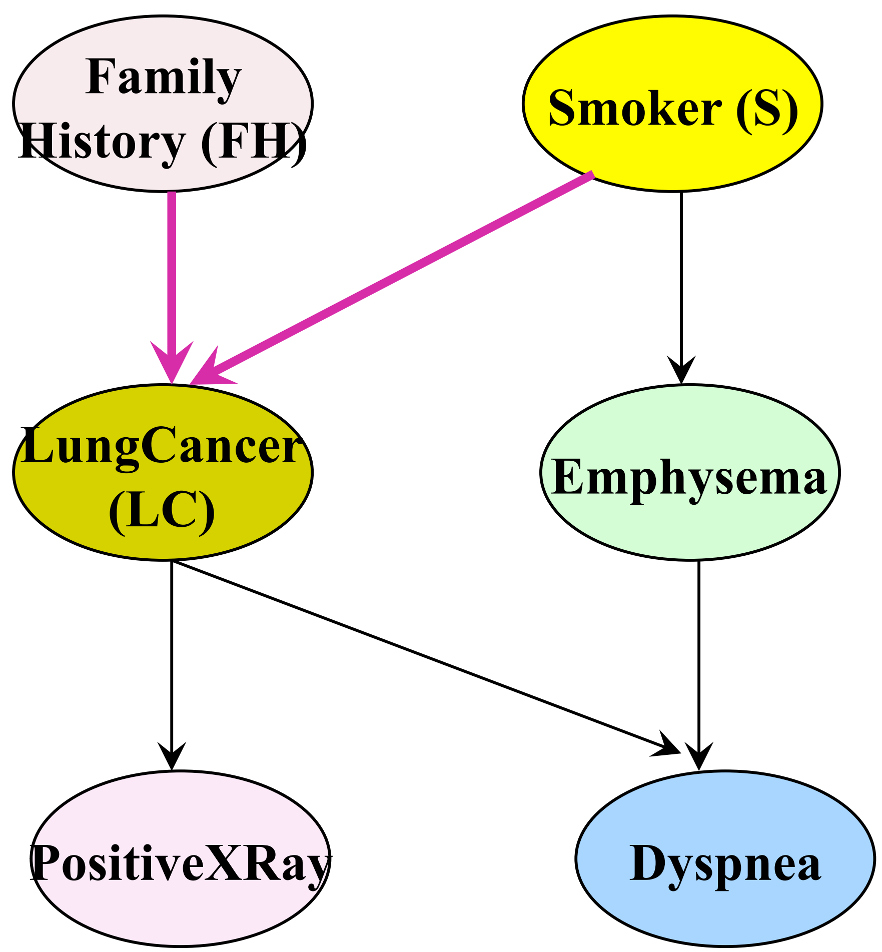
- Bayesian belief networks
- Backpropagation
- Support vector machines
- Classification using frequent patterns
- Lazy learners (or learning from your neighbors)
- Other classification methods
- Additional topics regarding classification
- Summary

Naïve Bayesian Classifier

- Advantage
  - easy to compute
  - good results in most cases
- Disadvantage
  - assumption: class conditional independence
  - dependencies exist in practice
    - e.g., hospital patients: age, family history, fever, cough, lunch cancer, diabetes, etc.

Bayesian Belief Networks

- Subset of variables conditionally independent
- Causal model: a directed, acyclic graph
  - node, link, parent, CPTs



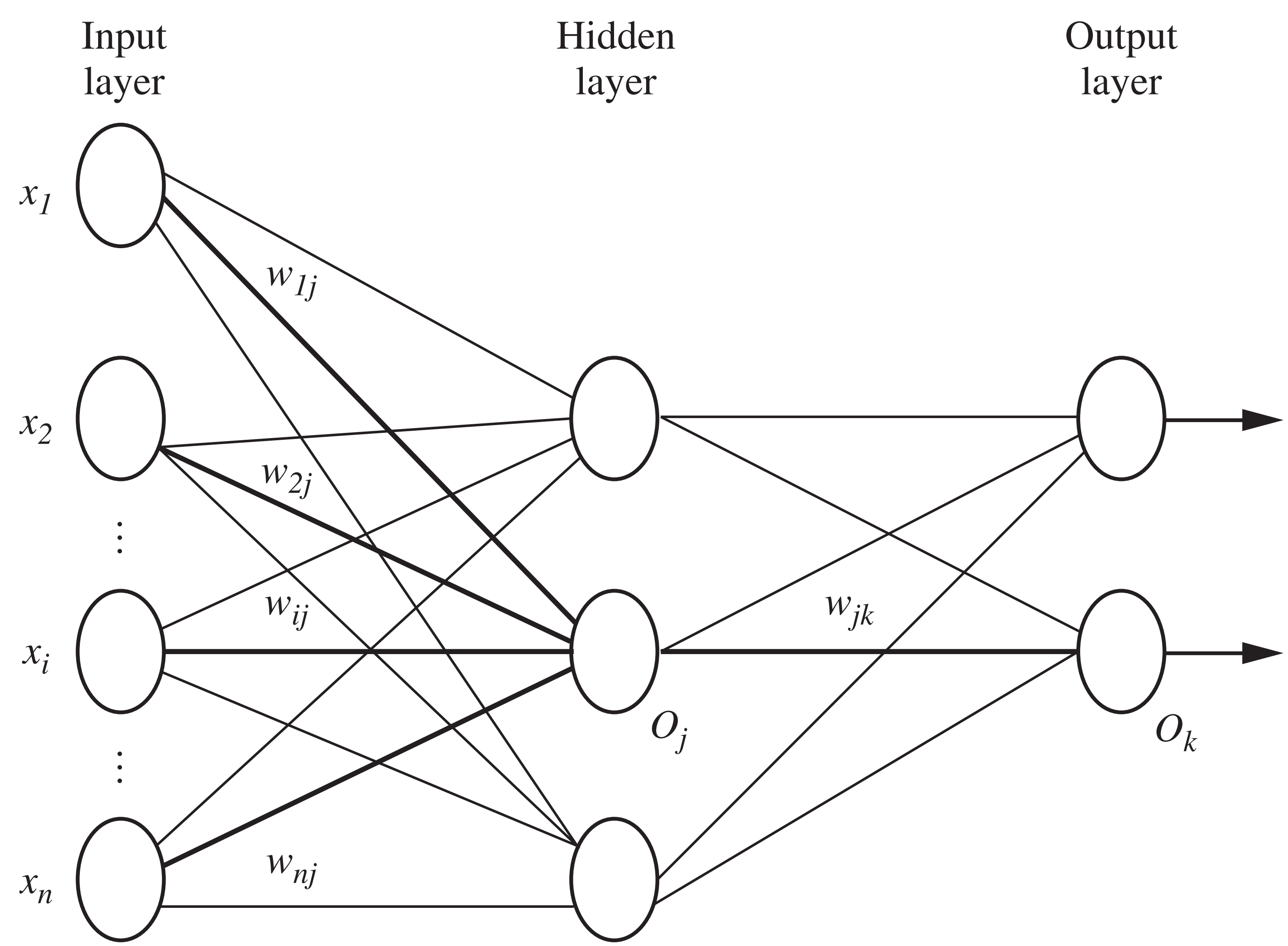
CPT: Conditional Probability Table
for variable LungCancer:

	(FH, S)	(FH, ~S)	(~FH, S)	(~FH, ~S)
LC	0.8	0.5	0.7	0.1
~LC	0.2	0.5	0.3	0.9

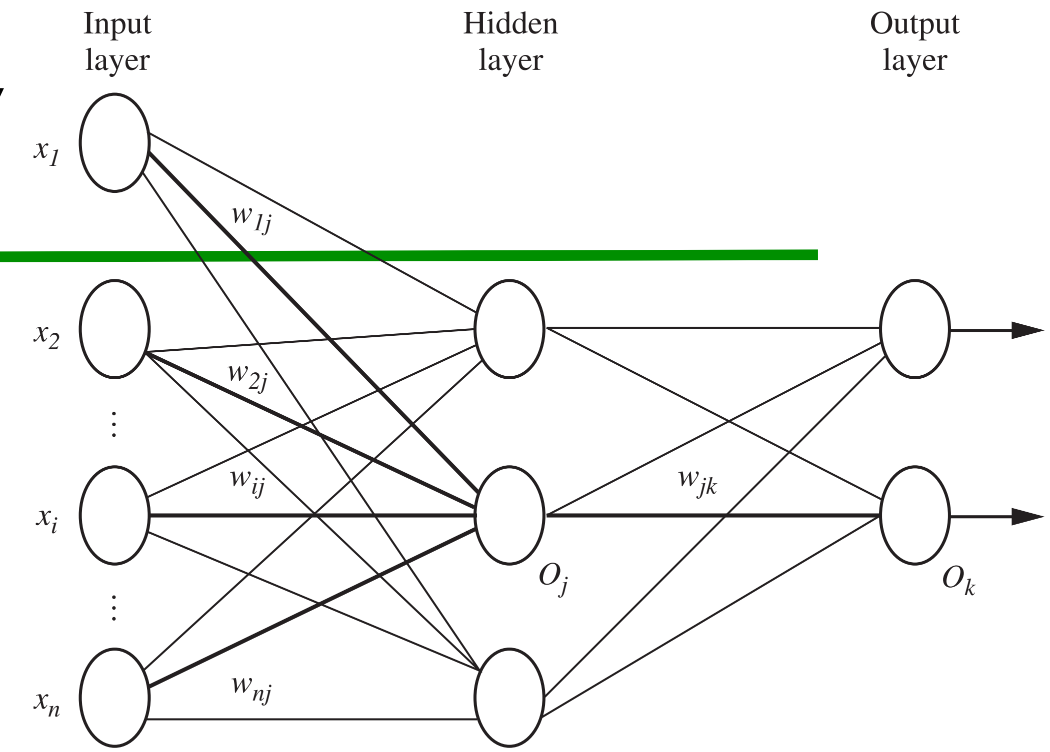
$$P(x_1, \cdots, x_n) = \prod_{i=1}^{n} P(x_i | Parents(Y_i))$$

Neural Network

- A set of connected input/out units
- Weighted connections
- Multi-layer
- Feed-forward
- Fully connected
- Backpropagation
- Adjust weights



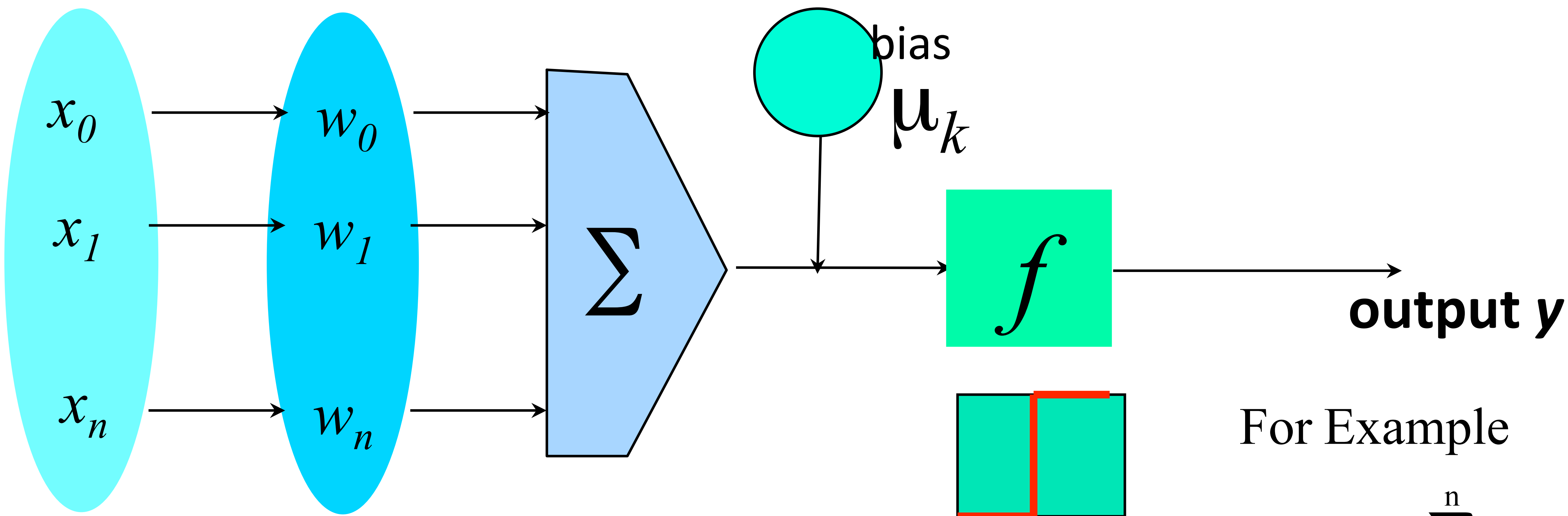
Network Topology



- # hidden layers
- # units in each layer
- Input values
  - continuous: normalize to [0.0, 1.0]
  - discrete: one unit per attribute value, initially 0
- Output: 1 unit per class if more than two classes
- Trial-and-error
  - different topology, different initial weights

Neuron

- A hidden/output layer unit



Input vector x **weight vector w** **weighted sum** **Activation function**

For Example

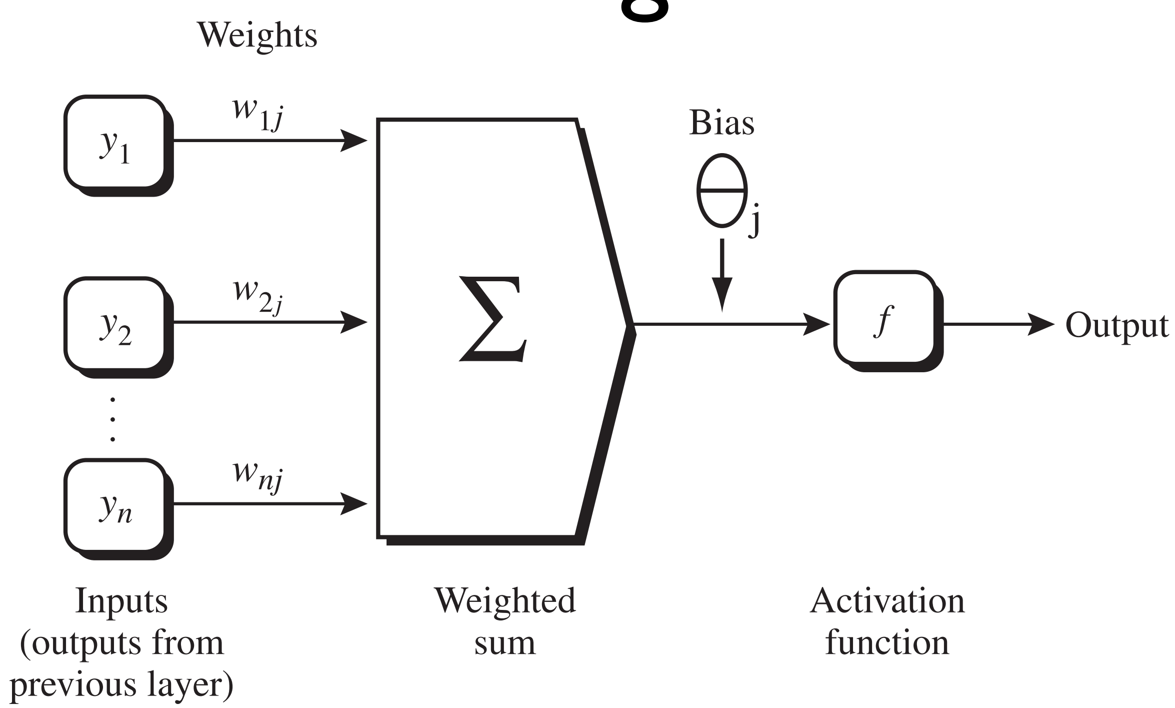
$$y = \text{sign}\left(\sum_{i=0}^{n} w_i x_i - \mu_k\right)$$

Backpropagation

- Initialize weights, biases: small random numbers
- Propagate the inputs forward
- Backpropagate the error
- Terminating condition

$$I_j = \sum_i w_{ij} O_i + \theta_j$$

$$O_j = \frac{1}{1 + e^{-I_j}}$$



$$Err_j = O_j(1 - O_j)(T_j - O_j)$$

$$Err_j = O_j(1 - O_j) \sum_k Err_k W_{jk}$$

$$w_{ij} = w_{ij} + (l) Err_j O_i$$

$$\theta_j = \theta_j + (l) Err_j$$

Neural Network as Classifier

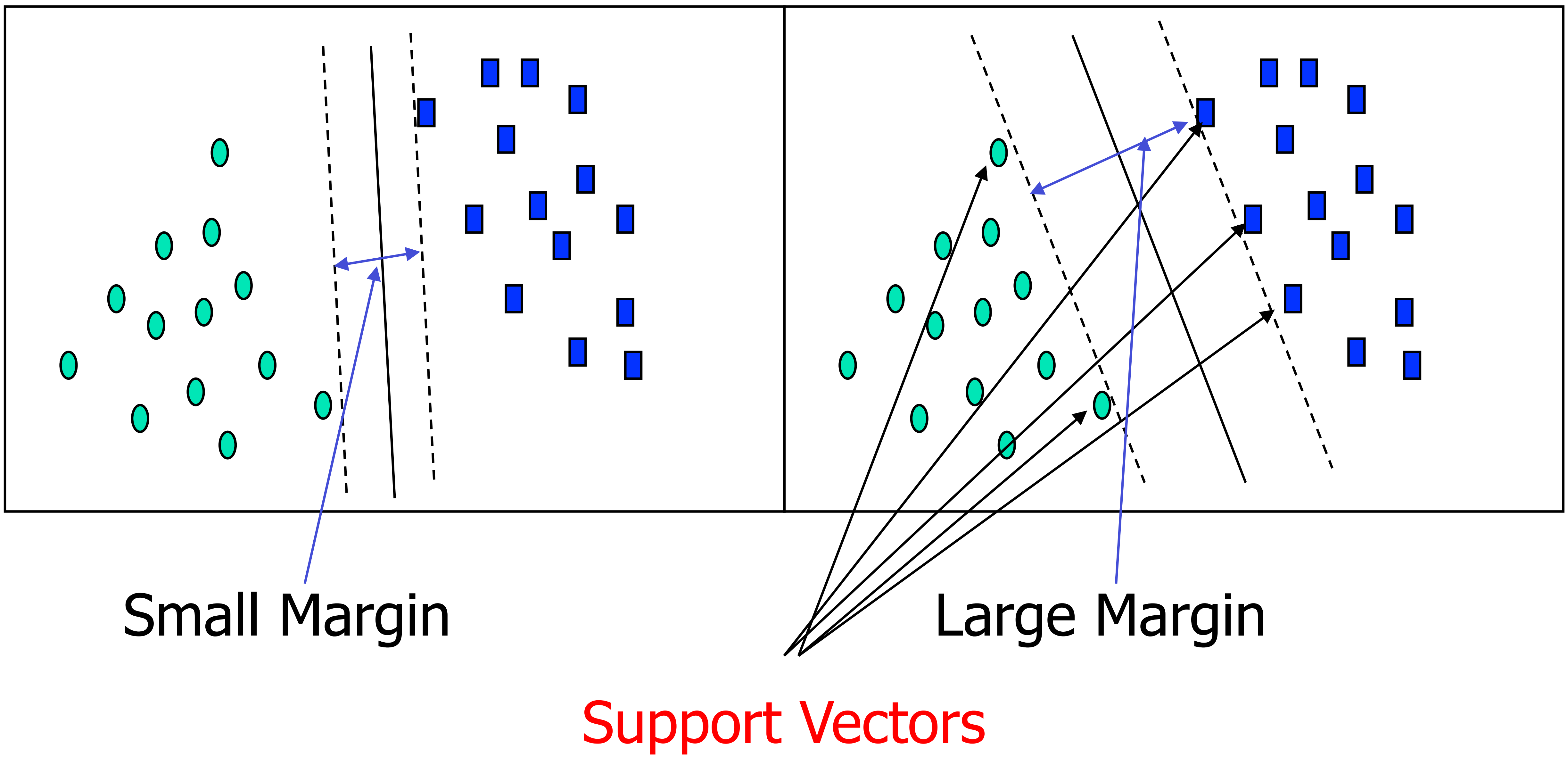
- Weakness
  - long training time; parameters determined empirically; poor interpretability
- Strength
  - high tolerance to noisy data; can classify untrained patterns; well-suited for continuous-valued inputs & outputs; success on a wide array of real-world data; inherently parallel; rule extraction

Chap 9: Advanced Classification

- Bayesian belief networks
- Backpropagation
- Support vector machines
- Classification using frequent patterns
- Lazy learners (or learning from your neighbors)
- Other classification methods
- Additional topics regarding classification
- Summary

SVM Example

- Linearly separable data

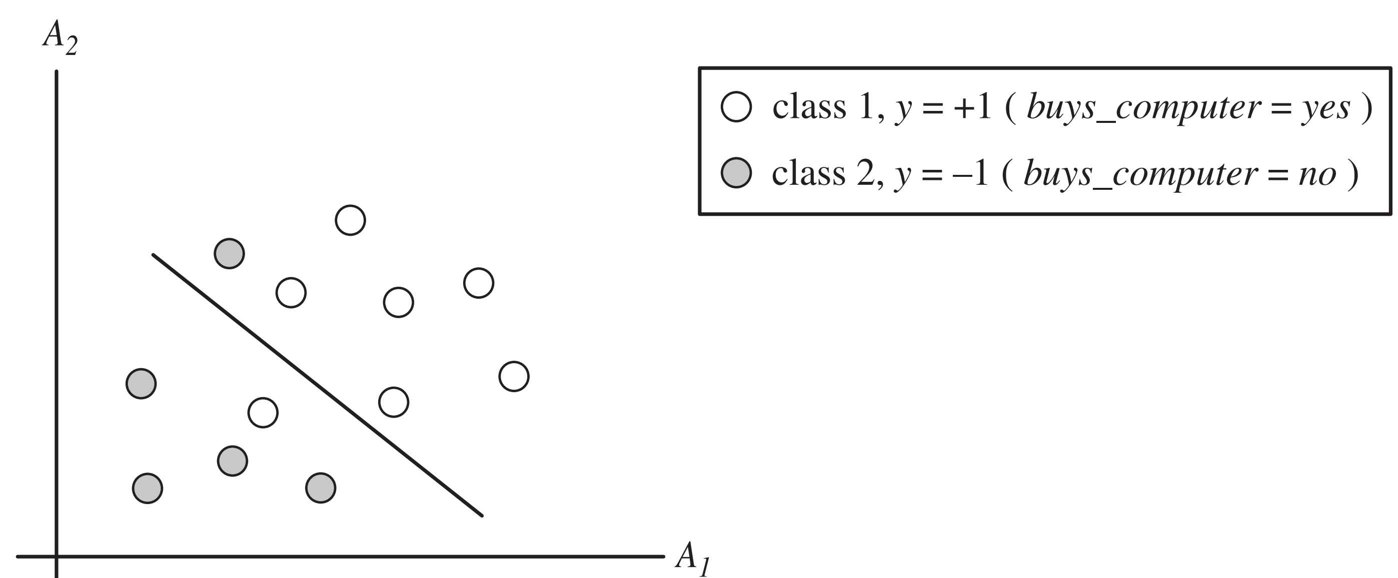


SVM: Linearly Separable Data

- D: (X_1, y_1), ..., $(X_{|D|}, y_{|D|})$, X_i is n-dimensional
- A separating hyperplane
  - $W * X + b = 0$
- Find maximum marginal hyperplane (MMH)
- Hyperplanes H_1 and H_2
  - sides of the margin
- Support vectors
  - training tuples fall on H_1 H_2

SVM: Linearly Inseparable

- Transform data into a higher dimension
- Search for optimal linear separating hyperplane in the new space
- Computing dot product on the transformed data mathematically equivalent to applying a kernel function to original data
  - $K(X_i, X_j)$
  - $= \varnothing(X_i) \cdot \varnothing(X_j)$



Support Vector Machines

- Classification for both linear & nonlinear data
- Transforms data to higher dimension using nonlinear mapping
- Searches for optimal linear separation hyperplane in the new dimension
- SVM finds this hyperplane using support vectors ("essential" training tuples) and margins (defined by the support vectors)

Chap 9: Advanced Classification

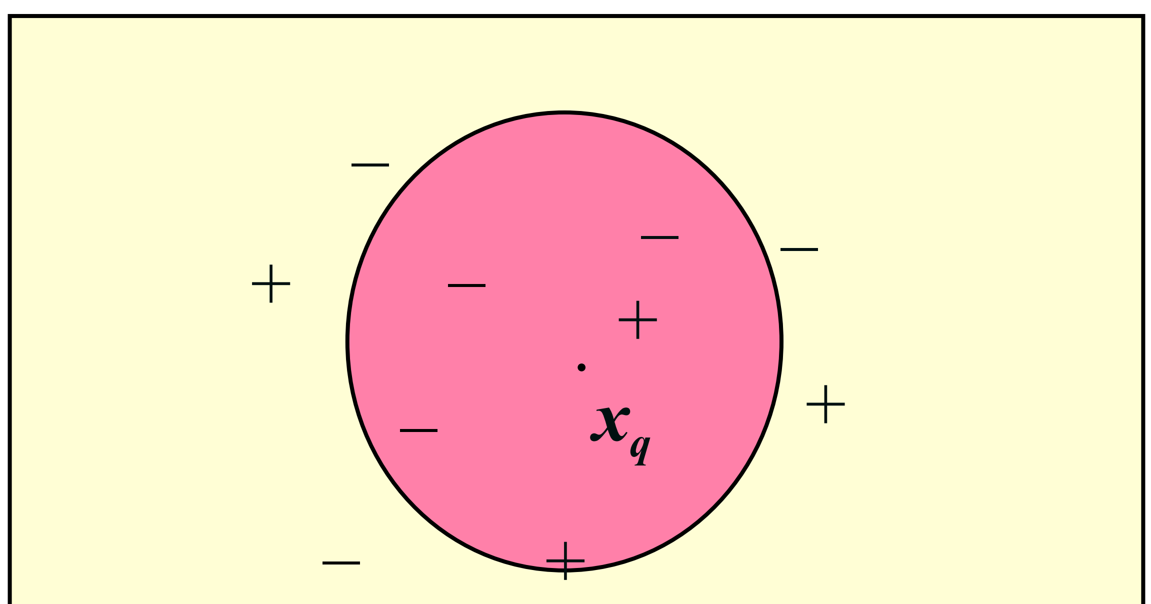
- Bayesian belief networks
- Backpropagation
- Support vector machines
- Classification using frequent patterns
- Lazy learners (or learning from your neighbors)
- Other classification methods
- Additional topics regarding classification
- Summary

Lazy vs. Eager Learning

- Eager learning: constructs classification model on training data before receiving test data
- Lazy learning: stores training data and delays processing until a new instance must be classified
- Lazy: less time in training, more time in predicting
- Lazy: require efficient storage techniques
- Lazy: richer hypothesis space using many local linear functions for implicit global approximation

k-Nearest-Neighbor Classifiers

- Find k-nearest-neighbors of a test tuple
- Discrete-valued: most common values
- Continuous-valued: average of k values
- Give greater weight to closer neighbors
- Indexing for nearest-neighbor search



Chap 9: Advanced Classification

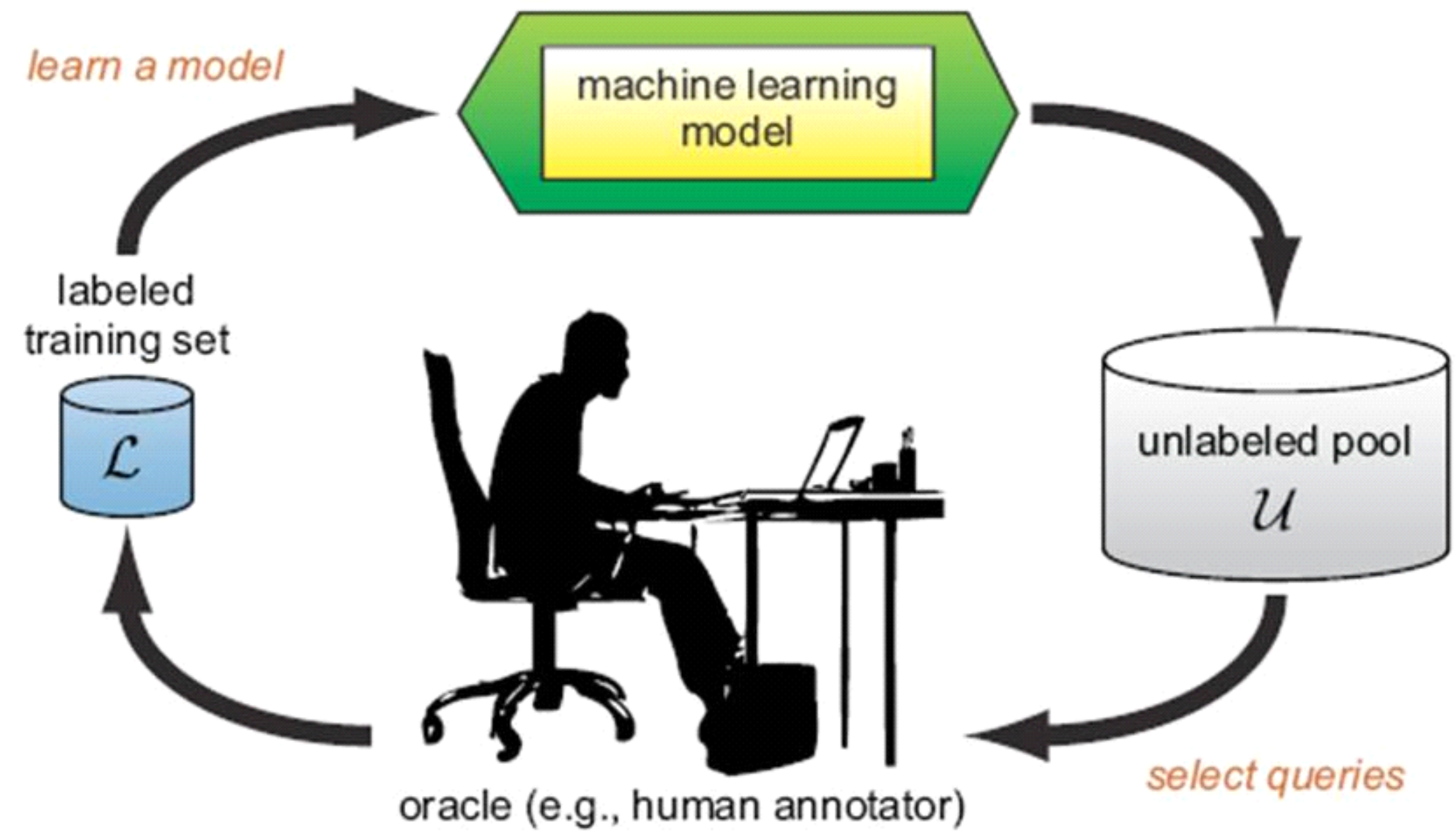
- Bayesian belief networks
- Backpropagation
- Support vector machines
- Classification using frequent patterns
- Lazy learners (or learning from your neighbors)
- Other classification methods
- Additional topics regarding classification
- Summary

- Multiclass classification
  - one-versus-all (OVA)
  - all-versus-all (AVA)

- Semi-supervised training
  - self-training
  - cotraining

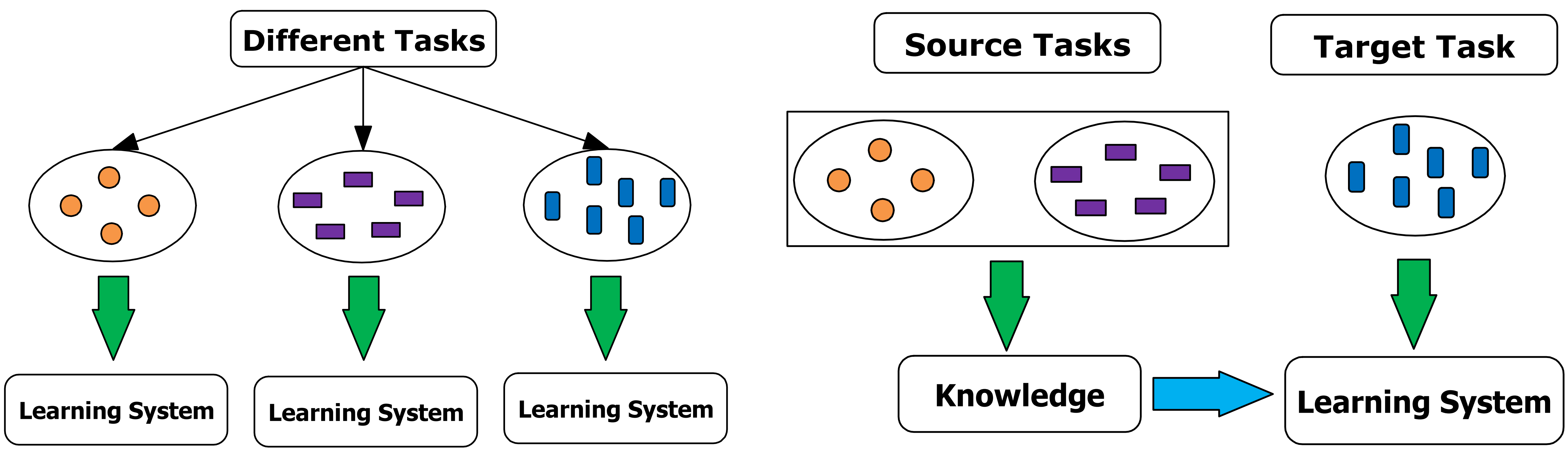
Active Learning

- Achieve high accuracy using as few labeled instances as possible
- Uncertainty sampling, vision space



Transfer Learning

- E.g., sentiment classification
  - camera reviews ==> TV reviews
- E.g., TrAdaBoost (Transfer AdaBoost)



Chap 9: Advanced Classification

- Bayesian belief networks
- Backpropagation
- Support vector machines
- Classification using frequent patterns
- Lazy learners (or learning from your neighbors)
- Other classification methods
- Additional topics regarding classification
- Summary