Wrap-Up and Review

Louis Oliphant
oliphant@cs.wisc.edu
cs540 section 2

Announcements

- HW5 Graded
  - Solution and grades will be up on the web
- Projects not yet graded
- Team Reviews due today
- Final Dec 19th, 10:05am CS&S 1325

What Have We Learned (first half)

- Search
  - Different Global and Local Search Techniques
    - DFS, BFS, Greedy Search, A* Search
    - Hill Climbing, Simulated Annealing, Genetic Algorithms
  - Searching With an Opponent
    - Game Playing – Mini-Max with alpha beta pruning
    - Game Theory
- Logic
  - Propositional Logic
  - First Order Logic

What we have learned (second half)

- Learning
  - Induction of models
  - Inference with models
  - Lots of models
    - K-NN, Decision Trees, Neural Nets, Naïve Bayes, Bayesian Networks, Ensembles, Support Vector Machines, Inductive Logic Programming, etc.
  - Each model is trying to capture a function: f(x) that is a mapping from feature space to a classification for every possible input
  - Model Bias
- Evaluation of Models: How well have you learned the function
  - Accuracy, precision, recall, confusion matrix, etc.

Review

- For each model know:
  - How to create the model (Induction)
  - How to label unseen examples (Inference)
  - Learning Bias
- Types of Learning
  - Unsupervised
  - Reinforment learning
  - Supervised
- Converting to Fixed Length Feature Vectors

Review

- K-NN
  - Euclidean Distance
  - Weighted Features
  - Choosing K
- Decision Trees
  - Information, Information Gain
  - Pruning
- Ensembles
  - Boosting
  - Occam's Razor

Review

- Types of Data
  - Noisy data
  - Missing Values
  - Continuous Features
  - Skewed Data
  - Irrelevant Features
- Feature Subset Selection
  - Forward chaining
  - Backward chaining

Review

- Methodology
  - Accuracy
  - Learning curves
  - Precision, Recall
  - N-Fold cross validation
  - Confusion Matrix
  - Train/Tune/Test set splits
  - Laplacian Priors

Review

- Perceptrons
  - Step Function, Sigmoid Function
  - Perceptron Training Rule, Delta rule
  - Threshold
  - Gradient Descent
  - Perceptrons and Logic
- Artificial Neural Networks (ANNs)
  - BackPropagation Algorithm
  - Non-boolean features
  - More than two classes
  - Overfitting Problems
- K-Means Clustering

Review

- Basic Probability
  - Joint probability, Full joint probability
  - Conditional probability
  - Marginalization
  - Bayes Rule
  - Chain Rule
  - Independence, Conditional Independence
- Naïve Bayes

Review

- Bayesian Networks
  - Exact Methods of Inference
    - Inference by Enumeration
    - Variable Elimination
  - Approximate Methods of Inference
    - Direct Sampling
    - Rejection Sampling
    - Likelihood Weighting
    - Markov Chain Monte Carlo
  - Induction
    - Parameter Learning
      - Maximum Likelihood (ML)
      - Maximum A-Posteriori (MAP)
    - Topology Learning

Review

- Inductive Logic Programming
  - More than fixed length feature vectors
  - Covering Algorithms in general
  - Top Down Approaches
    - FOIL
    - PROGOL
      - Seeds and Bottom Clauses
  - Bottom Up Approaches
    - GOLEM
- Support Vector Machines
  - Support Vector
  - Maximizing the margin
  - Kernels (mapping to higher dimensions)

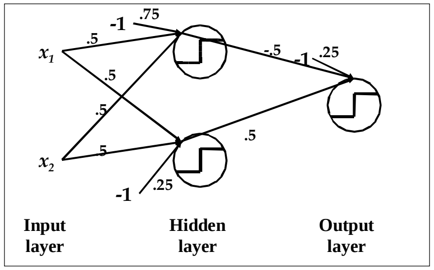
Information and Gain Example

F1	F2	F3	Class
0	1	0	A
1	2	0	A
1	0	1	B
0	1	1	B
2	1	1	A
1	0	2	C
2	2	0	A
2	2	1	C

Information is: $\boxed{\sum_{i=1}^{n} -P(v_i)\log_2(P(v_i))}$

Calculate Information of the above dataset.
Calculate Information Gain for each F1.

Back-Propagation Example



- Calculate Errors:
  - Output Units:
    - $\boxed{\delta_k \leftarrow o_k(1-o_k)(t_k-o_k)}$
  - Hidden Units:
    - $\boxed{\delta_h \leftarrow o_h(1-o_h)_{k\ outputs}(w_{kh}\delta_k)}$
- Update each network weight:

$$\boxed{\begin{array}{l} w_{ji} \leftarrow w_{ji}+\Delta w_{ji} \\ \text{where } \Delta w_{ji} = \alpha\delta_{hj}x_{ji} \end{array}}$$

- Sigmoid Function:

$$\boxed{\frac{1}{(1+e^{-(\vec{w}\cdot\vec{x})})}}$$

- Update weights using back-prop and the example: 1,0,1 and the learning rate 0.2

Probability Example

- Full Joint Probability is

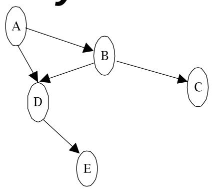
	A		~A	
	B	~B	B	~B
C	0.15	0.2	0.1	0.025
~C	0.3	0.15	0.05	0.025

P(A,B,C)=?
P(A|B,C)=?
P(A)=?
P(A,B)=?
P(A|B)=?
is A independent of B?

Bayes Rule:
P(A|B)=P(B|A)P(A)/P(B)

Chain Rule:
P(A,B,C,...,Z)=
P(A|B,C,...,Z)P(B|C,...,Z)...P(Z)

Bayesian Network Example



P(D|A,E)=
$\Sigma_B\Sigma_C P(A)P(B|A)P(C|B)P(D|A,B)P(E|D)=$
$P(A)P(E|D)\Sigma_B P(B|A)P(D|A,B)\Sigma_C P(C|B)=$
Continued on board...