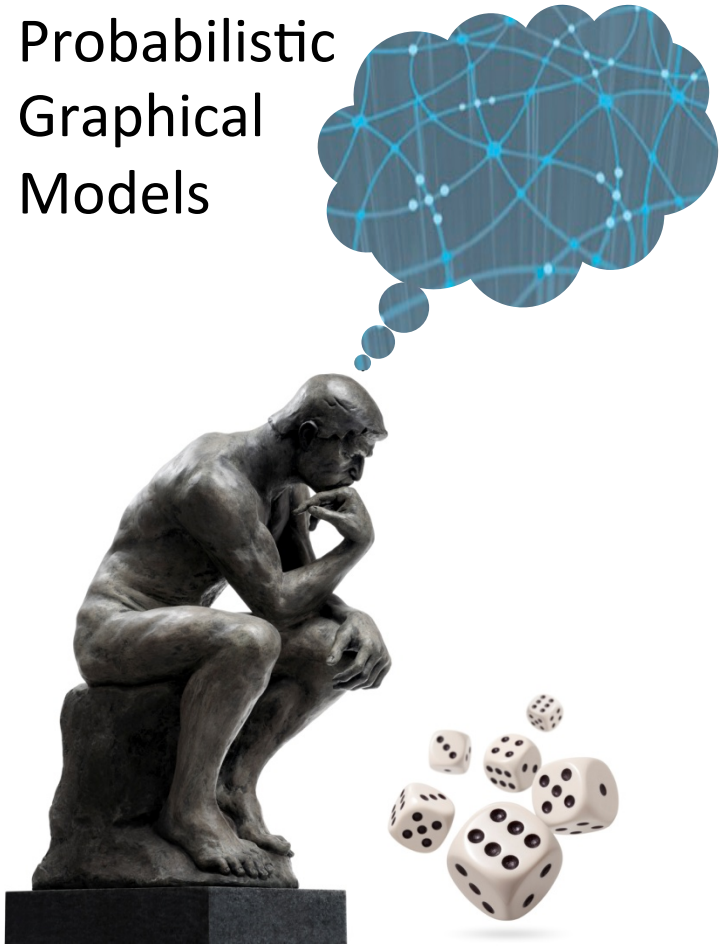
Probabilistic Graphical Models



Learning

Summary

Methods, Parameters, and Evaluation

Learning from 10K Feet

- Hypothesis (model) space
- Objective function
- Optimization algorithm

Hypothesis Space

- What are we searching for
  - Parameters
  - Structure
- Imposing constraints
  - For computational efficiency
  - To reduce model capacity
  - To incorporate prior knowledge

Objective Function

- Penalized likelihood
  - $\ell((G,\theta_G):D) + R(G,\theta_G)$
  - Parameter prior (MRFs - L_2 or L_1) (BNs - Dirichlet)
  - Structure complexity penalty
- Bayesian score (integrating parameters)
  - log P(G | D)
    = log P(D | G) + log P(G) + Const

marginal likelihood; graph prior

Optimization Algorithm

- Continuous
  - Closed form - BNs with multinomial
  - Gradient ascent < MRF missing data
  - EM - learning with missing data
- Discrete
  - Max spanning tree
  - Hill-climbing add, delete, reverse
- Discrete + continuous - computationally expensive

Hyperparameters

- Model hyperparameters
  - Equivalent sample size for parameter prior
  - Regularization strength for L1 or L2
  - Stopping criterion for EM
  - Strength of structure penalty
  - Set of features
  - # of values of latent variable
- Optimize on validation set

training set

train on training set
evaluate on validation set

cross-validation

Model Evaluation Criteria

- Log-likelihood on test set
- Task-specific objective
- "Match" with prior knowledge

segmentation accuracy
speech recognition WER

Troubleshooting: Underfitting

- Training & test performance both low
- Solutions
  - Decrease regularization
  - Reduce structure penalties
  - Add features via error analysis

Troubleshooting: Overfitting

- Training performance high, test performance low
- Solutions:
  - Increase regularization
  - Impose capacity constraints
  - Reduce feature set

Troubleshooting: Optimization

- Optimization may not be converging to good / global optimum
  - Can happen even if problem is convex
- Compare different learning rates, different random initializations

Troubleshooting: Objective Mismatch

$$\text{Objective}(M_1) \gg \text{Objective}(M_2)$$

$$\text{Performance}(M_1) \ll \text{Performance}(M_2)$$

- Need to redesign objective to match desired performance criterion

Typical Learning Loop

- Design model "template"
- Select hyperparameters via CV on training set
- Train on training set with chosen hyperparams
- Evaluate performance on held-out set
- Error analysis & model redesign
- Report results on separate test set