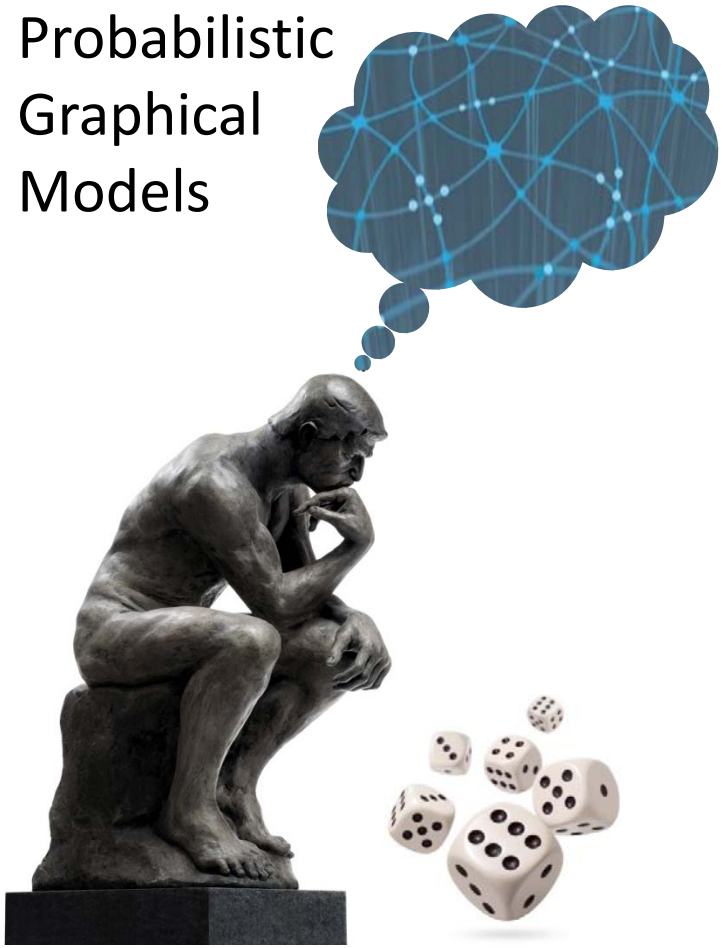
Probabilistic Graphical Models

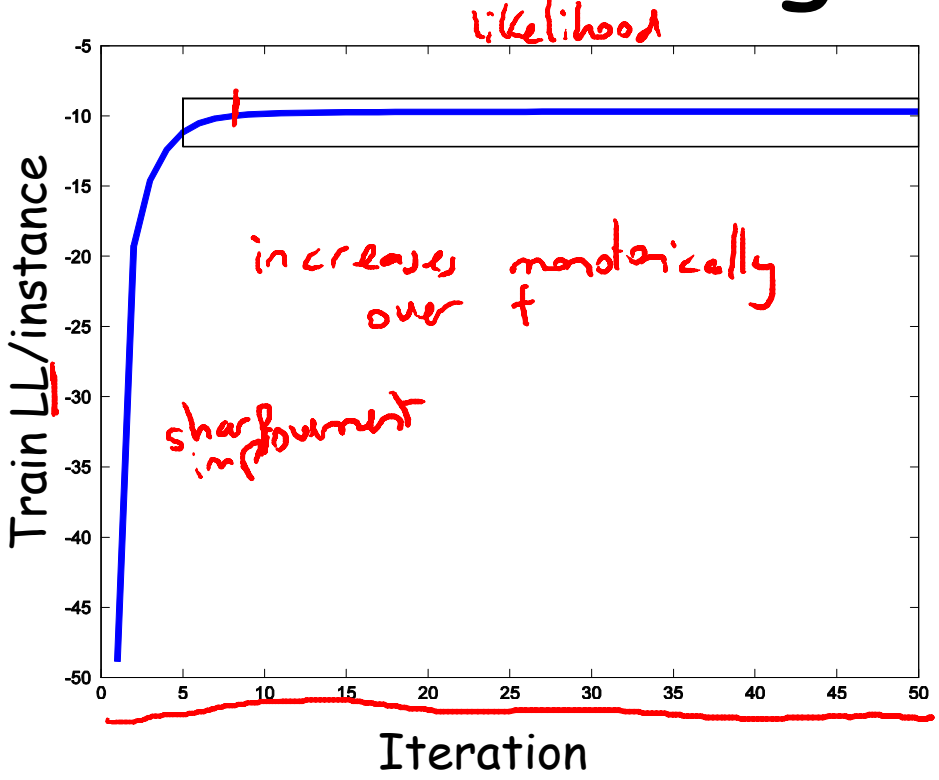


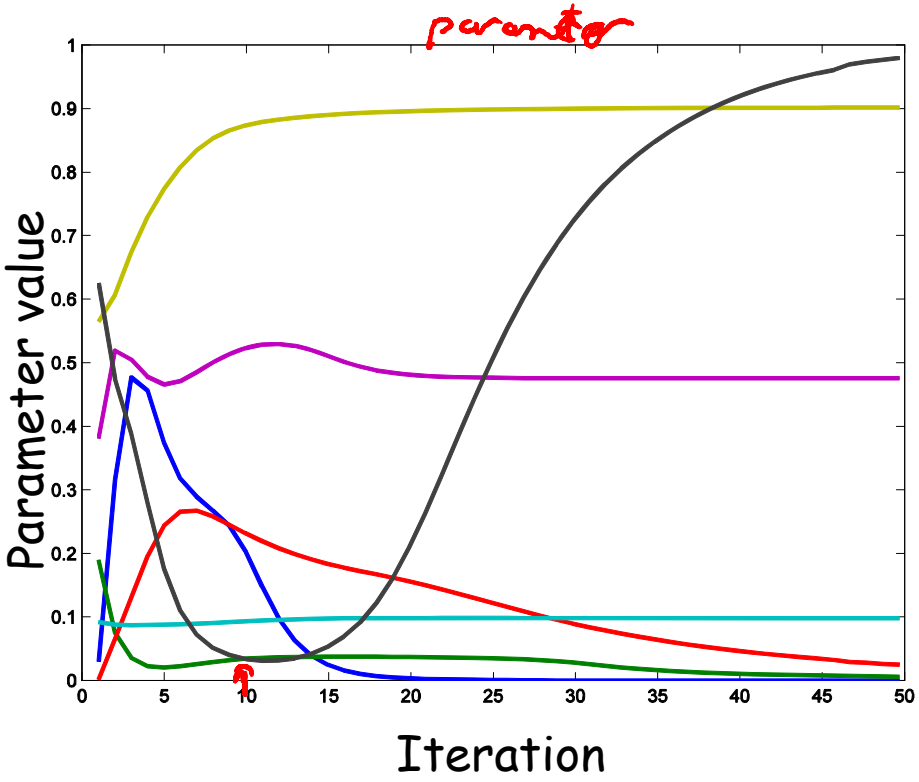
Learning

Incomplete Data

EM in Practice

EM Convergence in Practice

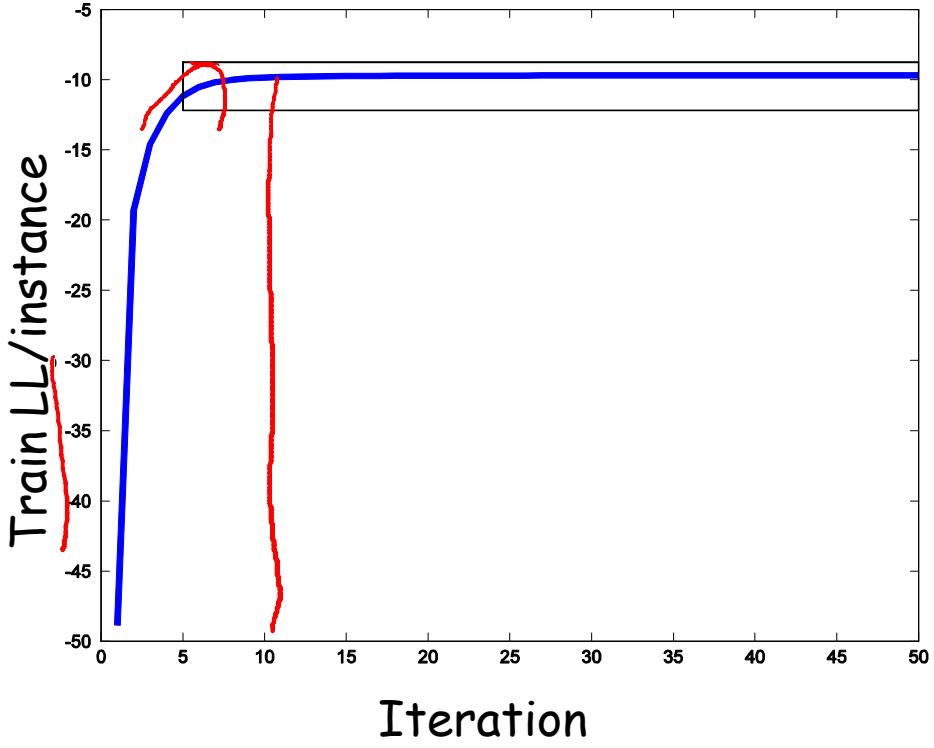


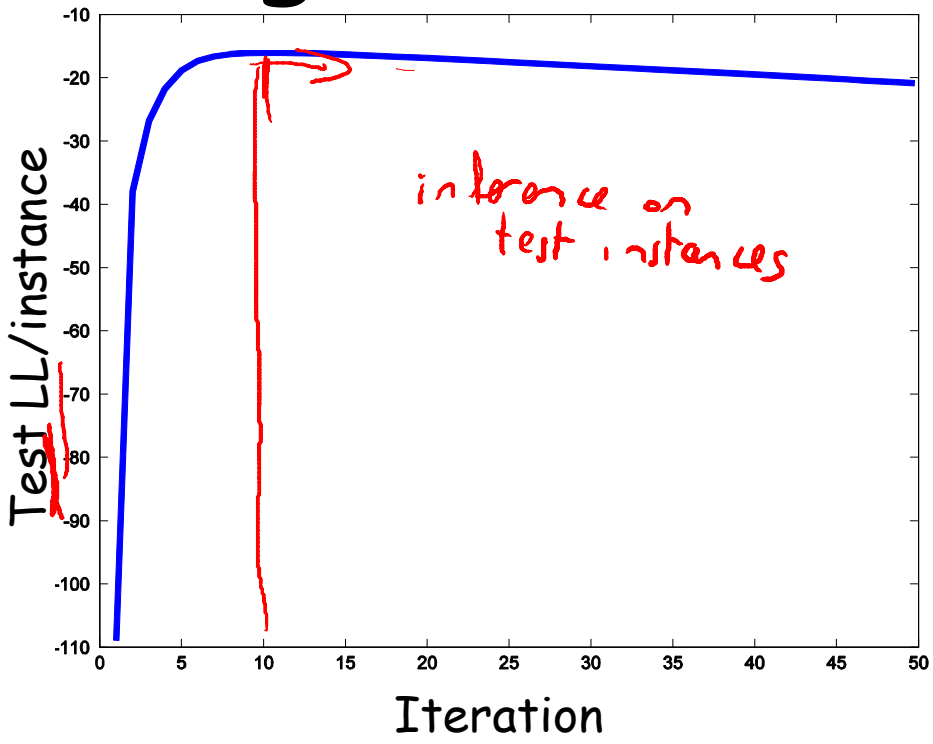


G. Elidan

Daphne Koller

Overfitting (numerical)



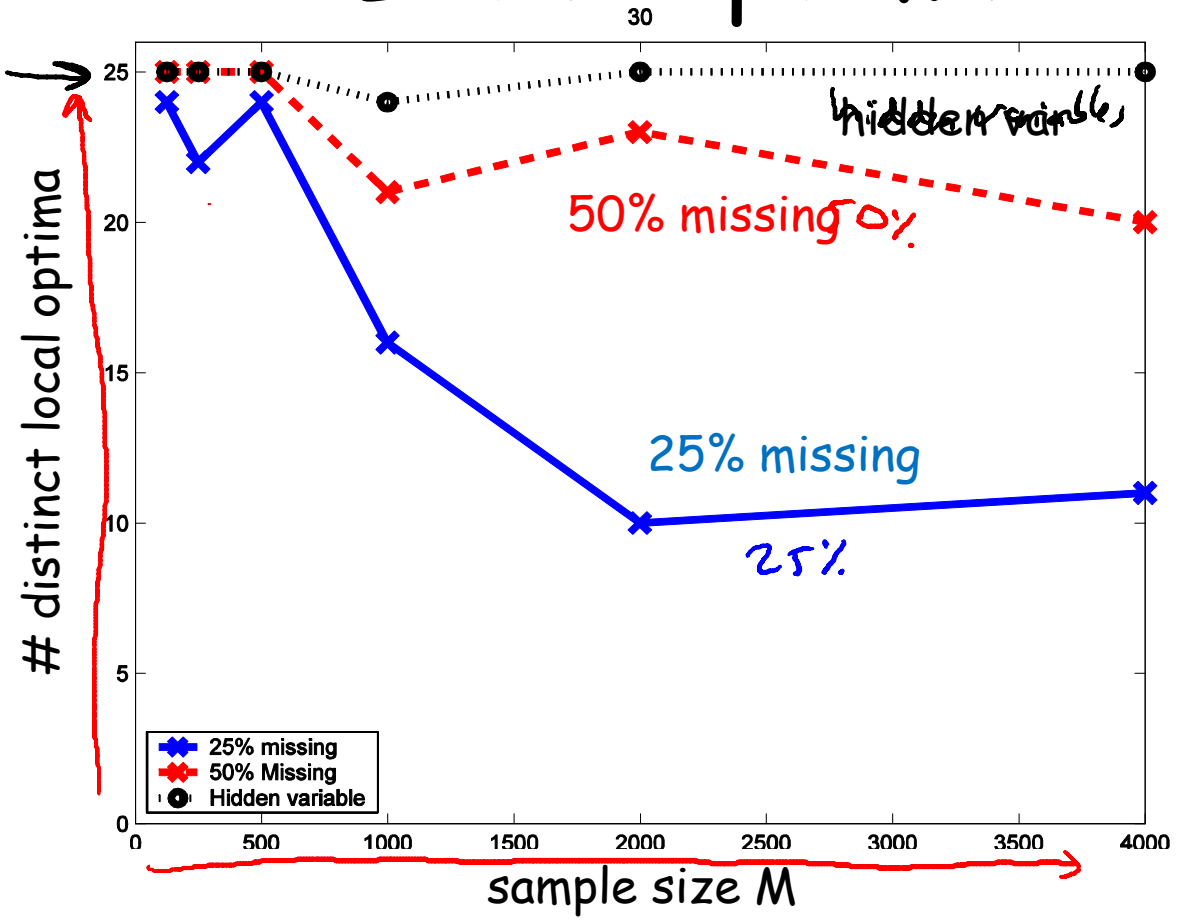


- Early stopping using cross validation
- Use MAP with parameter priors rather than MLE

G. Elidan

Daphne Koller

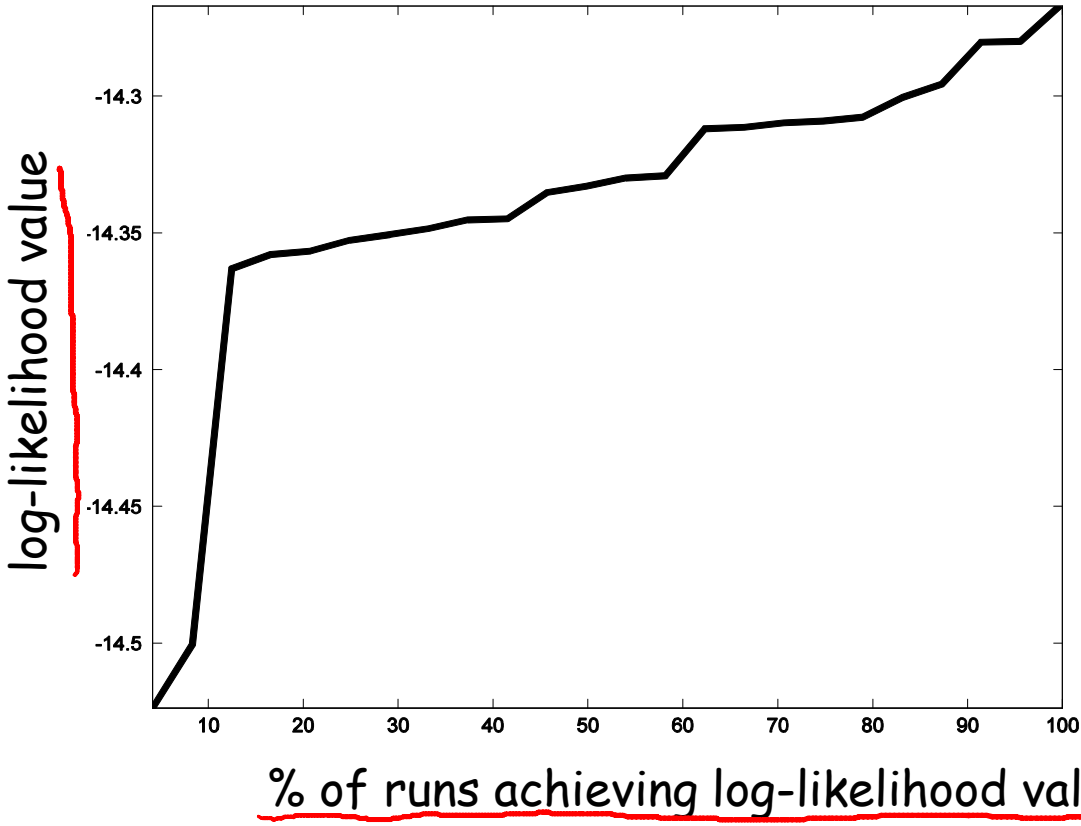
Local Optima



G. Elidan

Daphne Koller

Significance of Local Optima



G. Elidan

Daphne Koller

Initialization is Critical

- Multiple random restarts
- From prior knowledge
- From the output of a simpler algorithm

clustering (k-means, hierarchical agglomerative clustering)

Summary

- Convergence of likelihood ≠ convergence of parameters
- Running to convergence can lead to overfitting
- Local optima are unavoidable, and increase with the amount of missing data
- Local optima can be very different
- Initialization is critical