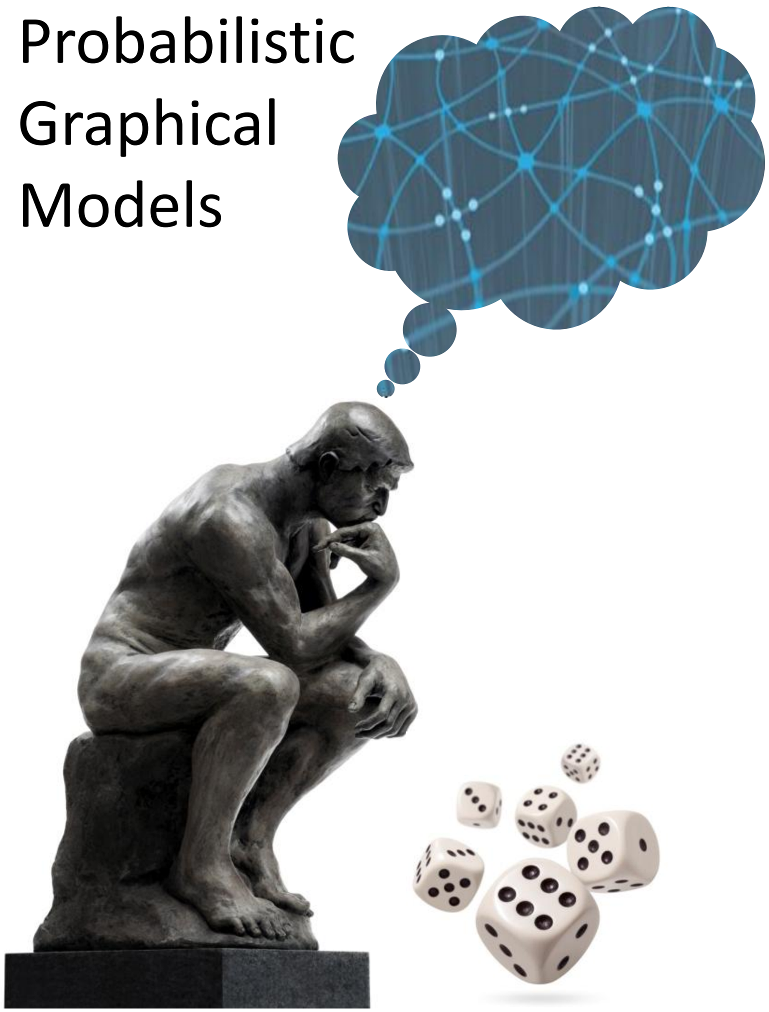
Probabilistic
Graphical
Models



Learning

Parameter Estimation

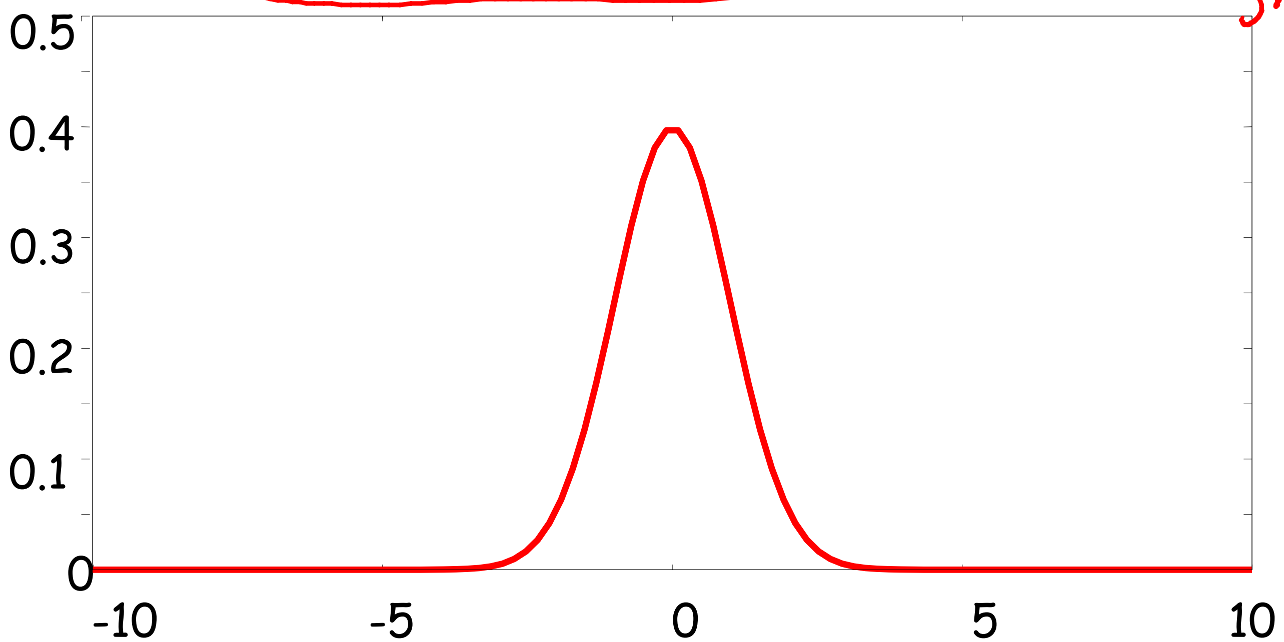
MAP Estimation for MRFs, CRFs

Gaussian Parameter Prior

$$P(\boldsymbol{\theta} : \sigma^2) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{\theta_i^2}{2\sigma^2} \right\}$$

0 mean univariate Gaussian

hyperparameter

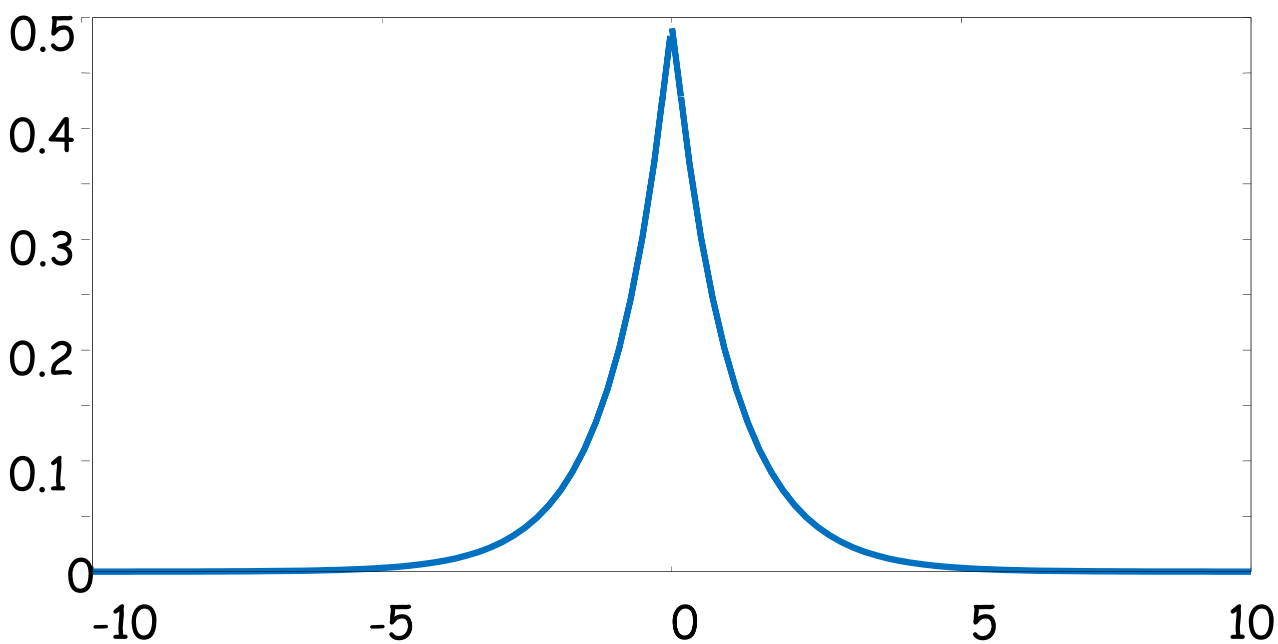


Daphne Koller

Laplacian Parameter Prior

hyperparameter

$$P(\boldsymbol{\theta} : \beta) = \prod_{i=1}^{k} \frac{1}{2\beta} \exp\left\{-\frac{|\theta_i|}{\beta}\right\}$$



MAP Estimation & Regularization

$$P(\boldsymbol{\theta} : \sigma^2) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{\theta_i^2}{2\sigma^2} \right\} \qquad P(\boldsymbol{\theta} : \beta) = \prod_{i=1}^{k} \frac{1}{2\beta} \exp\left\{ -\frac{|\theta_i|}{\beta} \right\}$$

likelihood

prior

$$\underline{\mathrm{argmax}_{\boldsymbol{\theta}} \boxed{P(\mathcal{D}, \boldsymbol{\theta})}} = \mathrm{argmax}_{\boldsymbol{\theta}} P(\mathcal{D} \mid \theta) P(\boldsymbol{\theta})$$

log likelihood

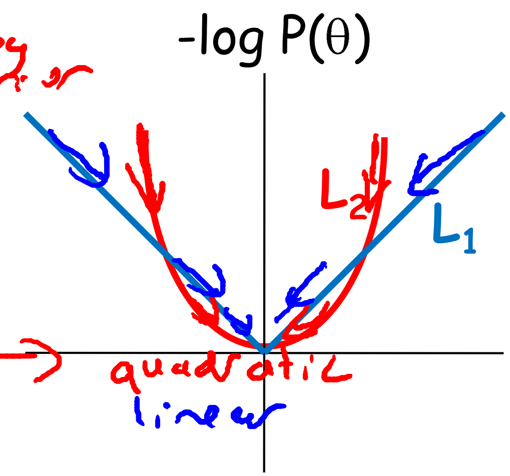
log prior

$$= \underline{\mathrm{argmax}_{\boldsymbol{\theta}}} \left(\underline{\ell(\boldsymbol{\theta} : \mathcal{D})} + \log P(\boldsymbol{\theta}) \right)$$

many $\theta_i \neq 0$

dense → L_2-regularization →

sparse L_1-regularization



Daphne Koller

Summary

- In undirected models, parameter coupling prevents efficient Bayesian estimation
- However, can still use parameter priors to avoid overfitting of ~~MLE~~ MAP
- Typical priors are L_1, L_2
  - Drive parameters toward zero
- L_1 provably induces sparse solutions
  - Performs feature selection / structure learning