Map Estimation for MRFs and CRFs

- MLE has the problem of overfitting
- MLE maximizes the likelihood function
- Use parameter priors
- In bayesian networks it was computationally elegant
- Conjugate priors in the parameter priors (integrate a dirichlet prior into the likelihood to obtain a closed form posterior)
- In MRFs and CRFs the likelihood cannot be maintained in closed form given the parameter coupling
- Maximum a posteriory (MAP) estimate of the parameters

Gaussian Parameter Prior (L2 regularization/dense)

$$P(\theta : \underset{\downarrow \text{close to } 0}{\sigma^2}) = \prod_{i=1}^{K} \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{\frac{-\theta_i^2}{2\sigma^2}\right\}$$

0 mean univariate Gaussian

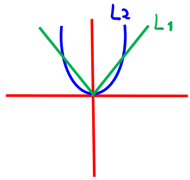
small variance → better

σ^2 hyperparameter

(same as dirichlet)

Laplacian Parameter Prior (L1 regularization) sparse

$$P(\underset{\downarrow \text{close to } 0}{\theta : \beta}) = \prod_{i=1}^{K} \frac{1}{2\beta} \exp\left\{-\frac{|\theta_i|}{\beta}\right\}$$



$$\text{Argmax}_\theta\, P(D,\theta) = \text{Argmax}_\theta\, \underbrace{P(D|\theta)}_{\text{likelihood}}\, \underbrace{P(\theta)}_{\text{Prior}}$$

$$= \text{Argmax}_\theta \left(\ell(\theta : D) + \log P(\theta)\right)$$

L2 Regularization

L1 performs feature selection/structure learning