Marginal Inference as Optimization

Vibhav Gogate

Exact Inference as Optimization

Assume we have a factorized distribution of the form

$$P_\Phi(\mathcal{X}) = \frac{1}{Z} \prod_{\phi \in \Phi} \phi(\boldsymbol{U}_\phi),$$

Your task is to find a tree decomposition such that

the set of beliefs in $\mathcal{T}$ defines a distribution Q by the formula

$$Q(\mathcal{X}) = \frac{\prod_{i \in \mathcal{V}_\mathcal{T}} \beta_i}{\prod_{(i-j) \in \mathcal{E}_\mathcal{T}} \mu_{i,j}}.$$

such that

$$\begin{aligned} \beta_i[\boldsymbol{c}_i] &= Q(\boldsymbol{c}_i) \\ \mu_{i,j}[\boldsymbol{s}_{i,j}] &= Q(\boldsymbol{s}_{i,j}). \end{aligned}$$

Exact Inference as Optimization

CTree-Optimize-KL:

Find $\boldsymbol{Q} = \{\beta_i : i \in \mathcal{V}_{\mathcal{T}}\} \cup \{\mu_{i,j} : (i\text{–}j) \in \mathcal{E}_{\mathcal{T}}\}$

maximizing $-\mathbb{D}(Q \| P_\Phi)$

subject to

$$\mu_{i,j}[\boldsymbol{s}_{i,j}] = \sum_{\boldsymbol{C}_i - \boldsymbol{S}_{i,j}} \beta_i(\boldsymbol{c}_i) \quad \forall (i\text{–}j) \in \mathcal{E}_{\mathcal{T}}, \forall \boldsymbol{s}_{i,j} \in Val(\boldsymbol{S}_{i,j})$$

$$\sum_{\boldsymbol{c}_i} \beta_i(\boldsymbol{c}_i) = 1 \quad \forall i \in \mathcal{V}_{\mathcal{T}}.$$

Energy Functional

$$\mathbb{D}(Q \| P_\Phi) = \ln Z - F[\tilde{P}_\Phi, Q]$$

where $F[\tilde{P}_\Phi, Q]$ *is the* energy functional

$$F[\tilde{P}_\Phi, Q] = \mathbb{E}_Q\left[\ln \tilde{P}(\mathcal{X})\right] + \mathbb{H}_Q(\mathcal{X}) = \sum_{\phi \in \Phi} \mathbb{E}_Q[\ln \phi] + \mathbb{H}_Q(\mathcal{X}).$$

Proof

$$\mathbb{D}(Q \| P_\Phi) = \mathbb{E}_Q[\ln Q(\mathcal{X})] - \mathbb{E}_Q[\ln P_\Phi(\mathcal{X})].$$

Using the product form of P_Φ, we have that

$$\ln P_\Phi(\mathcal{X}) = \sum_{\phi \in \Phi} \ln \phi(\boldsymbol{U}_\phi) - \ln Z.$$

recall that $\mathbb{H}_Q(\mathcal{X}) = -\mathbb{E}_Q[\ln Q(\mathcal{X})]$

$$\begin{aligned} \mathbb{D}(Q \| P_\Phi) &= -\mathbb{H}_Q(\mathcal{X}) - \mathbb{E}_Q\left[\sum_{\phi \in \Phi} \ln \phi(\boldsymbol{U}_\phi)\right] + \mathbb{E}_Q[\ln Z] \\ &= -F[\tilde{P}_\Phi, Q] + \ln Z. \end{aligned}$$

Exact Inference as Optimization (using Energy Functional)

CTree-Optimize:

Find $\boldsymbol{Q} = \{\beta_i : i \in \mathcal{V}_{\mathcal{T}}\} \cup \{\mu_{i,j} : (i–j) \in \mathcal{E}_{\mathcal{T}}\}$

maximizing $\tilde{F}[\tilde{P}_\Phi, \boldsymbol{Q}]$

subject to

$$\mu_{i,j}[\boldsymbol{s}_{i,j}] = \sum_{\boldsymbol{C}_i - \boldsymbol{S}_{i,j}} \beta_i(\boldsymbol{c}_i) \qquad \forall (i–j) \in \mathcal{E}_{\mathcal{T}}, \forall \boldsymbol{s}_{i,j} \in Val(\boldsymbol{S}_{i,j})$$

$$\sum_{\boldsymbol{c}_i} \beta_i(\boldsymbol{c}_i) = 1 \qquad \forall i \in \mathcal{V}_{\mathcal{T}}$$

$$\beta_i(\boldsymbol{c}_i) \geq 0 \qquad \forall i \in \mathcal{V}_{\mathcal{T}}, \boldsymbol{c}_i \in Val(\boldsymbol{C}_i).$$

Given a cluster tree $\mathcal{T}$ with a set of beliefs $\boldsymbol{Q}$ clusters in $\mathcal{T}$, we define the factored energy functional*:*

$$\tilde{F}[\tilde{P}_\Phi, \boldsymbol{Q}] = \sum_{i \in \mathcal{V}_{\mathcal{T}}} \boldsymbol{E}_{\boldsymbol{C}_i \sim \beta_i}[\ln \psi_i] + \sum_{i \in \mathcal{V}_{\mathcal{T}}} \boldsymbol{H}_{\beta_i}(\boldsymbol{C}_i) - \sum_{(i–j) \in \mathcal{E}_{\mathcal{T}}} \boldsymbol{H}_{\mu_{i,j}}(\boldsymbol{S}_{i,j}),$$

If $\boldsymbol{Q}$ is a set of calibrated beliefs for $\mathcal{T}$,

$$\tilde{F}[\tilde{P}_\Phi, \boldsymbol{Q}] = F[\tilde{P}_\Phi, Q].$$

Fixed Points

A set of beliefs $\boldsymbol{Q}$ *is a stationary point of* CTree-Optimize *if and only if there exists a set of factors* $\{\delta_{i\to j}[\boldsymbol{S}_{i,j}] : (i–j) \in \mathcal{E}_{\mathcal{T}}\}$ *such that*

$$\delta_{i\to j} \propto \sum_{\boldsymbol{C}_i - \boldsymbol{S}_{i,j}} \psi_i \left(\prod_{k \in \mathrm{Nb}_i - \{j\}} \delta_{k\to i} \right) \tag{11.10}$$

and moreover, we have that

$$\begin{aligned} \beta_i &\propto \psi_i \left(\prod_{j \in \mathrm{Nb}_i} \delta_{j\to i} \right) \\ \mu_{i,j} &= \delta_{j\to i} \cdot \delta_{i\to j}. \end{aligned}$$

IJGP as Optimization: Variational Analysis

Our optimization problem contains two approximations:

- We are using an approximation, rather than an exact, energy functional; and
- We are optimizing it over the space of pseudo-marginals, which is a relaxation (a superspace) of the space of all coherent probability distributions that factorize over the cluster graph.

CGraph-Optimize:

Find $\boldsymbol{Q}$

maximizing $\tilde{F}[\tilde{P}_\Phi, \boldsymbol{Q}]$

subject to

$$\boldsymbol{Q} \in Local[\mathcal{U}]$$

$$Local[\mathcal{U}] = \tag{11.16}$$

$$\left\{ \begin{array}{c} \{\beta_i : i \in \mathcal{V}_\mathcal{U}\} \cup \\ \{\mu_{i,j} : (i\text{–}j) \in \mathcal{E}_\mathcal{U}\} \end{array} \middle| \begin{array}{lcll} \mu_{i,j}[\boldsymbol{s}_{i,j}] & = & \sum_{\boldsymbol{C}_i - \boldsymbol{S}_{i,j}} \beta_i(\boldsymbol{c}_i) & \forall (i\text{–}j) \in \mathcal{E}_\mathcal{U}, \forall \boldsymbol{s}_{i,j} \in Val(\boldsymbol{S}_{i,j}) \\ 1 & = & \sum_{\boldsymbol{c}_i} \beta_i(\boldsymbol{c}_i) & \forall i \in \mathcal{V}_\mathcal{U} \\ \beta_i(\boldsymbol{c}_i) & \geq & 0 & \forall i \in \mathcal{V}_\mathcal{U}, \boldsymbol{c}_i \in Val(\boldsymbol{C}_i). \end{array} \right\}$$

Other Approximations

- Propagation using Approximate Messages
- Structured Variational Approximations