11/6

Lecture 18: Belief propagation, Bethe free energy

Gibbs free energy on trees only depends on 1- and 2-wise marginals:

- Average energy:

$$\mathcal{E}(x) = \sum_{(i,j)\in G} \lg\left(1/\psi_{ij}(x_i, x_j)\right), \text{ and}$$

$$\mathbb{E}_{\mu}[\mathcal{E}] = \sum_{(i,j)\in G} \underbrace{\mathbb{E}_{\mu}\left(\lg 1/\psi_{ij}(x_i,x_j)\right)}_{\substack{\text{only depends} \\ \text{on marginal dist on} \\ (x_i, x_j)}}$$

- Entropy: implied by the following:

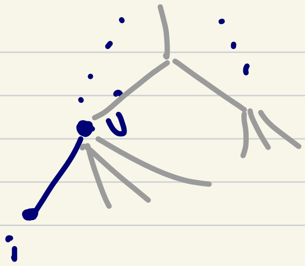
Lemma: If underlying graph is a tree, then

$$\mu(x) = \prod_{(i,j)\in G} \underbrace{\mu_{ij}(x_i, x_j)}_{\substack{\triangleq \text{ marginal} \\ \text{dist. on} \\ (x_i, x_j)}} \prod_{i\in[n]} \underbrace{\mu_i(x_i)}_{\substack{\triangleq \text{ marginal} \\ \text{dist. on} \\ x_i}}^{1-|\partial i|}$$

Pf: Induction on size of tree.

- Base case $(n=1)$ is vacuously true
- For inductive step, take any edge (i,j)

connected to leaf, where i is leaf.



$$\Pr_\mu[X = S] = \Pr(X_{n\setminus i} = S_{n\setminus i}) \cdot \Pr[X_i = S_i \mid X_{n\setminus i} = S_{n\setminus i}]$$

(Markov property)

$$= \Pr[X_{n\setminus i} = S_{n\setminus i}] \cdot \Pr(X_i = S_i \mid X_j = S_j)$$

$$= \underbrace{\Pr[X_{n\setminus i} = S_{n\setminus i}]}_{\mu_{T'}(S_{n\setminus i})} \cdot \frac{\mu_{ij}(S_i, S_j)}{\mu_j(S_j)},$$

for tree T' obtained by deleting i

contributes to $(\mu_j)^{1-|\partial j|}$

completing the inductive step.

<u>Corollary:</u>

$$H(\mu) = \sum_{(i,j)\in G} H(\mu_{ij}) - \sum_i (|\partial i| - 1) H(\mu_i)$$

So entropy term in Gibbs free energy also only depends on 1- and 2-wise marginals.

Leads us to define <u>Bethe free energy</u>.

Domain of definition: marginals $\{(\nu_i, \nu_{ij})\}$ that satisfy "local consistency", i.e.

$$\sum_{x_j \in \{\pm 1\}} \nu_{ij}(x_i, \ ) = \nu_i(x_i) \qquad \forall \ i,j, x_i$$

In CS, such marginals are said to satisfy "deg. 2 Sherali-Adams constraints" (LP version of pseudodistributions in SOS)

<u>Bethe free energy</u>:

$$G_B[\nu] \triangleq -\sum_{(i,j)\in G} H(\nu_{ij}) + \sum_{i\in[n]} (|\partial_i|-1) H(\nu_i) + \mathbb{E}_{\nu}[\mathcal{E}]$$

For trees, we have seen that

Bethe free energy = Gibbs free energy

<u>Claim 1.</u> Fixed point of BP satisfies local consistency.

<u>Pf</u>: Recall BP gives marginals

$$\nu_i(\sigma) \propto \prod_{j \in \partial i} \bar{m}_\sigma^{j \to i}$$

$$\nu_{ij}(\sigma_i, \sigma_j) \propto \psi_{ij}(\sigma_i, \sigma_j)\, m_{\sigma_i}^{i \to j}\, m_{\sigma_j}^{j \to i}.$$

Note: for any $x_i \in \{\pm 1\}$,

$$\sum_{x_j \in \{\pm 1\}} \nu_{ij}(x_i, x_j) = \sum_{x_j \in \{\pm 1\}} \psi_{ij}(x_i, x_j)\, m_{x_i}^{i \to j}\, m_{x_j}^{j \to i}$$

$= \bar{m}_{x_i}^{j \to i}$ by fixed point assumption

$$\propto m_{x_i}^{i \to j}\, \bar{m}_{x_i}^{j \to i}$$

$$\propto \nu_i(x_i)$$

This holds for all x_i, so "$\propto$" is actually "$=$". $\square$

So BP finds some point in the optimization landscape of G_B:

Convenient formula for Bethe free energy:

Let

$$\boxed{\begin{aligned} Z_i &\triangleq \sum_{\sigma\in\{\pm1\}} \prod_{j\in\partial i} \bar{m}_\sigma^{j\to i} \\ Z_{ij} &\triangleq \sum_{\sigma_i,\sigma_j\in\{\pm1\}} \psi_{ij}(\sigma_i,\sigma_j)\, m_{\sigma_i}^{i\to j}\cdot m_{\sigma_j}^{j\to i} \\ Z_{i,j} &\triangleq \sum_{\sigma\in\{\pm1\}} m_\sigma^{i\to j}\, \bar{m}_\sigma^{j\to i} \end{aligned}}$$

Then

$$\boxed{G_\beta[\nu] = -\sum_{i\in[n]} \ln Z_i - \sum_{(i,j)\in G} \ln Z_{ij} + \sum_{i\in[n]}\sum_{j\in\partial i} \ln Z_{i,j}}$$

Pf:

$$-\sum_{(i,j)} H(\nu_{ij}) + \mathbb{E}_\nu \mathcal{E} = -\sum_{(i,j)} \mathbb{E}_{\nu_{ij}} \ln \frac{\psi_{ij}(x_i,x_j)}{\nu_{ij}(x_i,x_j)}$$

$$= -\sum_{(i,j)} \mathbb{E}_{\nu_{ij}} \ln \frac{Z_{ij}}{m_{x_i}^{i\to j}\, m_{x_j}^{j\to i}}$$

$$= -\sum_{(i,j)} \ln Z_{ij} + \sum_{(i,j)} \mathbb{E}_{\nu_{ij}} \left[\underline{\ln m_{x_i}^{i\to j}} + \underline{\ln m_{x_j}^{j\to i}} \right]$$

$$-H(\nu_i) = -\sum_i \mathbb{E}_{\nu_i} \ln \frac{1}{\nu_i(x_i)}$$

$$= -\sum_i \mathbb{E}_{\nu_i} \ln \frac{Z_i}{\prod_{j\in\partial i} \bar{m}^{j\to i}_{x_i}}$$

$$= -\sum_i \ln Z_i + \sum_i \mathbb{E}_{\nu_i}\left[\sum_{j\in\partial i} \underline{\ln \bar{m}^{j\to i}_{x_i}}\right]$$

$$\sum_i |\partial i|\cdot H(\nu_i) = \sum_i \sum_{j\in\partial i} \mathbb{E}_{\nu_i} \ln \frac{1}{\nu_i(x_i)}$$

$$= \sum_i \sum_{j\in\partial i} \mathbb{E}_{\nu_i} \ln \frac{Z_{i,j}}{m^{i\to j}_{x_i}\, \bar{m}^{j\to i}_{x_i}}$$

$$= \sum_i \sum_{j\in\partial i} \ln Z_{i,j} - \sum_i \sum_{j\in\partial i} \mathbb{E}_{\nu_i}\left[\underline{\ln m^{i\to j}_{x_j}} + \underline{\ln \bar{m}^{j\to i}_{x_i}}\right].$$

$\square$

Thm: For any set of $\{\psi_{ij}\}_{(ij)\in G}$, even if graph is not a tree, there is a 1-1 correspondence b/t:

fixed points of BP $\longleftrightarrow$ Stationary points for G_β!

Pf:

$$\frac{\partial G_\beta(\nu)}{\partial m_{\sigma_i}^{(i)\to j}} = \frac{1}{Z_{i;j}} \frac{\partial}{\partial m_{\sigma_i}^{(i)\to j}} Z_{i;j} - \frac{1}{Z_{ij}} \frac{\partial}{\partial m_{\sigma_i}^{(i)\to j}} Z_{ij}$$

$$= \frac{\bar{m}_{\sigma_i}^{j\to(i)}}{\sum_\sigma m_\sigma^{(i)\to j} \bar{m}_\sigma^{j\to(i)}} - \frac{\sum_{\sigma_j} \psi_{ij}(\sigma_i,\sigma_j) m_{\sigma_j}^{(j)\to i}}{\sum_{\sigma,\sigma_j} \psi_{ij}(\sigma,\sigma_j) m_\sigma^{(i)\to j} \bar{m}_\sigma^{j\to(i)}}$$

This vanishing for all i,j,σ_i is equivalent to:

$$\bar{m}_{\sigma_i}^{j\to(i)} \propto \sum_{\sigma_j} \psi_{ij}(\sigma_i,\sigma_j) m_{\sigma_j}^{(j)\to i}$$

$$\frac{\partial G_\beta(\nu)}{\partial \bar{m}_{\sigma_i}^{j\to(i)}} = \frac{1}{Z_{i;j}} \frac{\partial}{\partial \bar{m}_{\sigma_i}^{j\to(i)}} Z_{i;j} - \frac{1}{Z_i} \frac{\partial}{\partial \bar{m}_{\sigma_i}^{j\to(i)}} Z_i$$

$$= \frac{m_{\sigma_i}^{i \to j}}{\sum_{\sigma} m_{\sigma}^{i \to j} \bar{m}_{\sigma}^{j \to i}} - \frac{\prod_{k \in \partial i \setminus j} \bar{m}_{\sigma_i}^{k \to i}}{\sum_{\sigma} \prod_{k \in \partial i} \bar{m}_{\sigma_i}^{k \to i}}$$

Again, this vanishing for all i, j, σ_i is equiv. to

$$m_{\sigma_i}^{i \to j} \propto \prod_{k \in \partial i \setminus j} \bar{m}_{\sigma_i}^{k \to i} ,$$

So stationarity $\equiv$ BP fixed point. $\square$