ベイジアンモデリングと確率的画像処理

Bayesian Modeling and Probabilistic Image Processing

(付録: Appendix)

東北大学 大学院情報科学研究科
田中 和之

http://www.smapip.is.tohoku.ac.jp/~kazu/
ガウス分布 (正規分布)

平均 $\mu$, 分散 $\sigma^2$ のガウス分布 (Gaussian Distribution) の確率密度関数

$$
\rho(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \quad (-\infty < x < +\infty)
$$

平均 $E[X] = \mu$, 分散 $V[X] = \sigma^2$

平均と分散はガウス積分の公式 (Gaussian Integral Formula) から導かれる

$$
\int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\xi^2\right) d\xi = \sqrt{2\pi}
$$
多次元ガウス分布

行列 $C$ を正定値の実対称行列として、2次元ガウス分布 (Two-Dimensional Gaussian Distribution) の確率密度関数

$$
\rho(x, y) = \frac{1}{\sqrt{(2\pi)^2 \det C}} \exp \left( -\frac{1}{2} (x - \mu_x, y - \mu_y) C^{-1} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} \right)
$$

$(-\infty < x < +\infty, -\infty < y < +\infty)$において行列 $C$ が共分散行列になる。

$$
\begin{pmatrix}
V[X] & \text{Cov}[X, Y] \\
\text{Cov}[Y, X] & V[Y]
\end{pmatrix} = C
$$

$d$次元ガウス積分の公式から導かれる

$$
\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp \left( -\frac{1}{2} \bar{x}^T C^{-1} \bar{x} \right) d\bar{x} = \sqrt{(2\pi)^d \det C}
$$

一般の次元への拡張も同様。
Probabilistic image processing is formulated by assuming the degradation process and the prior process of original image and by using the Bayes formula.

\[
Pr\{\bar{Y} | X, \sigma\} = \exp\left(-\frac{1}{2\sigma^2}(X_i - Y_i)^2\right)
\]

Probabilistic Image Processing by Bayesian Network

Bayes Formulas

Prior of Original Image

\[
Pr\{X | \alpha\} = \exp\left(-\frac{1}{2\alpha}(X_i - X_j)^2\right)
\]

:Original Image

\[
\bar{X} = (X_1, X_2, ..., X_N)
\]

:Degraded Image

\[
\bar{Y} = (Y_1, Y_2, ..., Y_N)
\]

Posterior of Bayesian Image Analysis

\[
Pr\{X | \bar{Y}, \alpha, \sigma\} \sim Pr\{\bar{Y} | X, \sigma\} Pr\{X | \alpha\}
\]

Original Image

Additive White Gaussian Noise

Degraded Image

Histogram of 256^2 Random Numbers Generated with Gaussian Distribution with Average 0 and Variance \(\sigma^2=40^2\).
Statistical Estimation of Hyperparameters

Hyperparameters $\alpha$, $\sigma$ are determined so as to maximize the marginal likelihood $\Pr\{Y=y|\alpha,\sigma\}$ with respect to $\alpha$, $\sigma$.

$$(\hat{\alpha}, \hat{\sigma}) = \arg \max (\alpha,\sigma) \Pr\{Y=y|\alpha,\sigma\}$$

\[
\Pr\{Y=y|\alpha,\sigma\} = \int \Pr\{Y=y|X=x,\sigma\} \Pr\{X=z|\alpha,\gamma\} dz
\]
Gaussian Graphical Model
(Gauss Markov Random Fields)

\[ P(\tilde{x} | \tilde{y}, \alpha, \sigma) \]

\[ \propto \exp \left( -\frac{1}{2\sigma^2} \sum_{i \in V} (x_i - y_i)^2 - \frac{1}{2} \alpha \sum_{\{i, j\} \in E} (x_i - x_j)^2 \right) \]

\[ = \exp \left( -\frac{1}{2\sigma^2} \| \tilde{x} - \tilde{y} \|^2 - \frac{1}{2} \alpha \tilde{x}^T C \tilde{x} \right) \]

\[ \langle i | C | j \rangle = \begin{cases} 4, & i = j \in V \\ -1, & \{i, j\} \in E \\ 0, & \text{otherwise} \end{cases} \]

Multidimensional Gauss Integral Formulas

\[ \hat{x} = \int \tilde{x} P(\tilde{x} | \tilde{y}) d\tilde{x} = \left( I + \alpha \sigma^2 C \right)^{-1} \tilde{y} \]

\[ P(\tilde{y} | \alpha, \sigma) = \sqrt{\frac{\det(\alpha C)}{(2\pi)^{|V|}}} \frac{\exp\left(-\frac{1}{2} \alpha \tilde{y}^T C \tilde{y} \right)}{\sqrt{\det(I + \alpha \sigma^2 C)}} \]

Maximum Likelihood Estimation \(\Rightarrow\) EM Algorithm

\[ (\hat{\alpha}, \hat{\sigma}) = \arg \max_{(\alpha, \sigma)} P(\tilde{g} | \alpha, \sigma) \]
Average of Gaussian Graphical Model

Average of the posterior probability can be calculated by using the multi-dimensional Gauss integral Formula

\[ \hat{x} = \langle \hat{X} | \alpha, \sigma \rangle = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \Pr\{ \hat{X} = \hat{z} \mid \hat{Y} = \tilde{y}, \alpha, \sigma \} dz_1 dz_2 \cdots dz_{|V|} \]

\[ = \frac{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp \left( -\frac{1}{2\sigma^2} \left( \tilde{z} - \frac{I}{I + \alpha \sigma^2 C } \tilde{y} \right)^T \left( I + \alpha \sigma^2 C \right) \left( \tilde{z} - \frac{I}{I + \alpha \sigma^2 C } \tilde{y} \right) \right) dz_1 dz_2 \cdots dz_{|V|} }{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp \left( -\frac{1}{2\sigma^2} \left( \tilde{z} - \frac{I}{I + \alpha \sigma^2 C } \tilde{y} \right)^T \left( I + \alpha \sigma^2 C \right) \left( \tilde{z} - \frac{I}{I + \alpha \sigma^2 C } \tilde{y} \right) \right) dz_1 dz_2 \cdots dz_{|V|} } \]

\[ = \frac{I}{I + \alpha \sigma^2 C } \tilde{y} \]

Gaussian Integral formula

\[ \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\tilde{z} - \tilde{\mu}) \exp \left( -\frac{1}{2} (\tilde{z} - \tilde{\mu})^T (I - \alpha \sigma^2 C(\gamma))(\tilde{z} - \tilde{\mu}) \right) dz_1 dz_2 \cdots dz_{|V|} = \tilde{0} \]
Marginal Likelihood of Gaussian Graphical Model

\[
\Pr\{\bar{Y} = \bar{y} \mid \alpha, \sigma\} = \int \Pr\{\bar{Y} = \bar{y} \mid \bar{X} = \bar{z}, \sigma\} \Pr\{\bar{X} = \bar{z} \mid \alpha\} d\bar{z} = \frac{Z_{\text{POS}}(\bar{y}, \alpha, \sigma)}{(2\pi\sigma^2)^{|V|/2} Z_{\text{PR}}(\sigma)}
\]

\[
Z_{\text{PR}}(\alpha) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2} \alpha \bar{z}^T C \bar{z}\right) dz_1 dz_2 \cdots dz_{|V|} = \sqrt{\frac{(2\pi)^{|V|}}{\alpha |V| \det C}}
\]

\[
Z_{\text{POS}}(\bar{y}, \alpha, \sigma)
= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2\sigma^2} \left(\bar{z} - \frac{I}{I + \alpha \sigma^2 C} \bar{y}\right)^T \left(I + \alpha \sigma^2 C \left(\bar{z} - \frac{I}{I + \alpha \sigma^2 C} \bar{y}\right)\right)^{-1} \frac{1}{2} \alpha \bar{y}^T \frac{C}{I + \alpha \sigma^2 C} \bar{y}\right) dz_1 dz_2 \cdots dz_{|V|}
= \exp\left(-\frac{1}{2} \alpha \bar{y}^T \frac{C}{I + \alpha \sigma^2 C} \bar{y}\right) \times \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2\sigma^2} \left(\bar{z} - \frac{I}{I + \alpha \sigma^2 C} \bar{y}\right)^T \left(I + \alpha \sigma^2 C \left(\bar{z} - \frac{I}{I + \alpha \sigma^2 C} \bar{y}\right)\right)^{-1} \frac{1}{2} \alpha \bar{y}^T \frac{C}{I + \alpha \sigma^2 C} \bar{y}\right) dz_1 dz_2 \cdots dz_{|V|}
= \sqrt{\frac{(2\pi)^{|V|}}{\det(I + \alpha \sigma^2 C)}} \exp\left(-\frac{1}{2} \alpha \bar{y}^T \frac{C}{I + \alpha \sigma^2 C} \bar{y}\right)
\]

\[
\langle i \mid C \mid j \rangle = \begin{cases} 4, & i = j \in V \\ -1, & \{i, j\} \in E \\ 0, & \text{otherwise} \end{cases}
\]

Gaussian Integral formula

\[
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2} (\bar{z} - \bar{\mu})^T A (\bar{z} - \bar{\mu})\right) dz_1 dz_2 \cdots dz_{|V|} = \sqrt{\frac{(2\pi)^{|V|}}{\det A}}
\]
Marginal Likelihood in Gaussian Graphical Model

\[
\Pr\{\tilde{Y} = \tilde{y} \mid \alpha, \sigma\} = \frac{\det(\alpha C)}{\sqrt{|V| \det(I + \alpha \sigma^2 C)}} \exp\left(-\frac{1}{2} \frac{\alpha \tilde{y}}{I + \alpha \sigma^2 C} \tilde{y}^T\right)
\]

Extremum Conditions for $\alpha$ and $\sigma$

\[
\frac{\partial}{\partial \alpha} \Pr\{Y = y \mid \alpha, \sigma\} = 0, \quad \frac{\partial}{\partial \sigma} \Pr\{Y = y \mid \alpha, \sigma\} = 0
\]

\[
(\hat{\alpha}, \hat{\sigma}) = \arg \max_{(\alpha, \sigma)} \Pr\{\tilde{Y} = \tilde{y} \mid \alpha, \sigma\}
\]

\[
\alpha(\tau) \leftarrow \left(\frac{1}{|V|} \text{Tr} \frac{\sigma(t-1)^2 C}{I + \alpha(t-1)\sigma(t-1)^2 C} + \frac{1}{|V|} \tilde{y}^T \frac{C}{I + \alpha(t-1)\sigma(t-1)^2 C} \tilde{y}\right)^{-1}
\]

\[
\sigma(t) \leftarrow \sqrt{\left(\frac{\sigma(t-1)^2 I}{|V|} \text{Tr} \frac{I + \alpha(t-1)\sigma(t-1)^2 C}{I + \alpha(t-1)\sigma(t-1)^2 C} \tilde{y}^T \frac{\alpha(t-1)^2 \sigma(t-1)^4 C^2}{(I + \alpha(t-1)\sigma(t-1)^2 C)^2} \tilde{y}\right)}
\]

Iterated Algorithm

EM Algorithm

\[
(\alpha(t+1), \sigma(t+1)) \leftarrow \arg \max_{(\alpha, \sigma)} Q(\alpha, \sigma \mid \alpha(t), \sigma(t))
\]
Bayesian Image Analysis by Gaussian Graphical Model

Iteration Procedure of EM algorithm in Gaussian Graphical Model

\[ (\hat{\alpha}, \hat{\sigma}) = \arg \max_{(\alpha, \sigma)} P(g | \alpha, \sigma) \]

\[ (\alpha(t + 1), \sigma(t + 1)) \leftarrow \arg \max_{(\alpha, \sigma)} Q(\alpha, \sigma | \alpha(t), \sigma(t), \hat{y}). \]

\[ \hat{x}(t) = (I + \alpha(t)\sigma(t)^2 C)^{-1} \hat{y} \]

\[ \sigma = 40 \]

\[ \hat{\sigma} = 37.624 \]

\[ \hat{\alpha} = 0.000713 \]
扱いやすい確率モデルのグラフ表現

扱いやすい確率モデルの数理構造

\[
\sum_{A=T,F} \sum_{B=T,F} \sum_{C=T,F} f(A,D) g(B,D) h(C,D)
\]

\[
= \left( \sum_{A=T,F} f(A,D) \right) \left( \sum_{B=T,F} g(B,D) \right) \left( \sum_{C=T,F} h(C,D) \right)
\]

別々に和を計算できる

扱いやすくない確率モデルの数理構造

\[
\sum_{A=T,F} \sum_{B=T,F} \sum_{C=T,F} f(A,B) g(B,C) h(C,A)
\]

別々に和を計算することが難しい

木構造をもつグラフ表現

閉路を含むグラフ表現
転送行列法＝確率伝搬法（1）

1次元鎖

$$\Pr\{\vec{X} = \vec{x}\} = \frac{1}{Z} \prod_{i=1}^{N-1} W_{i,i+1}(x_i, x_{i+1})$$

$$L_{k-1\rightarrow k}(x_k) \equiv \sum_{x_1} \sum_{x_2} \cdots \sum_{x_{k-1}} \left( \prod_{i=1}^{k-1} W_{i,i+1}(x_i, x_{i+1}) \right)$$

$$R_{k+1\rightarrow k}(x_k) \equiv \sum_{x_{k+1}} \sum_{x_{k+2}} \cdots \sum_{x_{N-1}} \left( \prod_{i=k}^{N-1} W_{i,i+1}(x_i, x_{i+1}) \right)$$
転送行列法＝確率伝搬法 (2)

漸化式

\[ L_{k \rightarrow k+1}(x_{k+1}) = \sum_{x_k} L_{k-1 \rightarrow k}(x_k) W_{k, k+1}(x_k, x_{k+1}) \]

\[ R_{k \rightarrow k-1}(x_{k-1}) = \sum_{x_k} W_{k-1, k}(x_{k-1}, x_k) R_{k+1 \rightarrow k}(x_k) \]

\[ \Pr\{X_m = x_m\} = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_{m-1}} \sum_{x_{m+1}} \cdots \sum_{x_{N-1}} \Pr\{\bar{X} = \bar{x}\} \]

\[ = \frac{L_{m-1 \rightarrow m}(x_m) R_{m+1 \rightarrow m}(x_m)}{\sum_{x_m} L_{m-1 \rightarrow m}(x_m) R_{m+1 \rightarrow m}(x_m)} \]
Belief Propagation for Tree Graphical Model

\[ \sum \sum \sum W_{12}(x_1, x_2)W_{13}(x_1, x_3)W_{14}(x_1, x_4)W_{15}(x_1, x_5) \]

\[ = W_{12}(x_1, x_2) \left( \sum_{x_3} W_{13}(x_1, x_3) \right) \left( \sum_{x_4} W_{14}(x_1, x_4) \right) \left( \sum_{x_5} W_{15}(x_1, x_5) \right) \]

\[ \sum_{x_3} W_{13}(x_1, x_3) \cdot \sum_{x_4} W_{14}(x_1, x_4) \cdot \sum_{x_5} W_{15}(x_1, x_5) \]

\[ M_{3\rightarrow 1}(x_1) \cdot M_{4\rightarrow 1}(x_1) \cdot M_{5\rightarrow 1}(x_1) \]
The message from node 1 to node 2 can be expressed in terms of all the messages incoming to node 1 except the own message.

Summation over all the nodes except node 2
確率伝搬法 (Belief Propagation)

閉路のないグラフ上の確率モデル

\[ P(a, b, c, d, x_1, x_2) = W_A(a, x_1)W_B(b, x_1)W_{12}(x_1, x_2)W_C(c, x_2)W(d, x_2) \]

\[ x_3 \in a \quad x_4 \in b \]
\[ x_5 \in c \quad x_6 \in d \]
確率伝搬法 (Belief Propagation)

閉路のないグラフ上の確率モデル

\[ W_A(a, x_1) \quad W_B(b, x_1) \]

\[ a \quad b \]

\[ W_{12}(x_1, x_2) \]

\[ W_C(c, x_2) \quad W_D(d, x_2) \]

\[ P(a, b, c, d, x_1, x_2) = W_A(a, x_1)W_B(b, x_1) \times W_{12}(x_1, x_2)W_C(c, x_2)W(d, x_2) \]
閉路のないグラフ上の確率伝搬法

$$\Pr\{\vec{X} = \vec{x}\} = \frac{1}{Z} \prod_{i=1}^{N-1} W_{i,i+1}(x_i, x_{i+1})$$

閉路が無いことが重要!!

同じノードは2度通らない

$$M_{k \rightarrow k+1}(x_{k+1}) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_k} \left( \prod_{i=1}^{k} W_{i,i+1}(x_i, x_{i+1}) \right)$$

$$= \sum_{x_k} M_{k-1 \rightarrow k}(x_k) M_{k-2 \rightarrow k}(x_k) M_{k-3 \rightarrow k}(x_k) W_{k,k+1}(x_k, x_{k+1})$$
確率伝搬法

閉路のあるグラフ上の確率モデル

\[ P(\tilde{x}) = P(x_1, x_2, \cdots, x_L) = \prod_{ij \in B} \Phi_{ij}(x_i, x_j) \]

\[ P_1(x_1) = \sum \sum \cdots \sum P(x_1, x_2, x_3, \cdots, x_L) \]

\[ = \frac{M_{2\rightarrow1}(x_1)M_{3\rightarrow1}(x_1)M_{4\rightarrow1}(x_1)M_{5\rightarrow1}(x_1)}{\sum M_{2\rightarrow1}(z_1)M_{3\rightarrow1}(z_1)M_{4\rightarrow1}(z_1)M_{5\rightarrow1}(z_1)} \]

\[ = M_{2\rightarrow1}(x_1) \]
閉路のあるグラフ上の確率モデル

\[ M_{1\rightarrow 2}(x_2) = \frac{\sum W_{12}(z_1, x_2) M_{3\rightarrow 1}(z_1) M_{4\rightarrow 1}(z_1) M_{5\rightarrow 1}(z_1)}{\sum \sum W_{12}(z_1, z_2) M_{3\rightarrow 1}(z_1) M_{4\rightarrow 1}(z_1) M_{5\rightarrow 1}(z_1)} \]

閉路のあるグラフ上でも局所的な
構造だけに着目してアルゴリズムを
構成することは可能。
ただし、得られる結果は厳密では
なく近似アルゴリズム

メッセージに対する固定
定点方程式

\[ \vec{M} = \Psi(\vec{M}) \]
固定点方程式と反復法

固定点方程式

\[ \vec{M}^* = \Psi(\vec{M}^*) \]

反復法

繰り返し出力を入力に入れるこことにより、固定点方程式の解が数値的に得られる。
Probabilistic Image Processing by Bayesian Network

The systems have double loop structures of an inner loop and an outer loop. Inner loop is the belief propagation in Bayesian network and outer loop is a statistical learning method to estimate hyperparameters $\alpha$ and $\sigma$. 

\[ \text{arg max}_{(a,b)} Q(a,b|\alpha(t),\sigma(t)) \quad (t = 1, 2, 3, 4, 5, \ldots) \] 

$Q(a,b|\alpha(t),\sigma(t))$ is calculated by Loopy Belief Propagation.
確率伝播法の情報論的解釈

Kullback-Leibler Divergence

\[ D[Q|P] = \sum_x Q(x) \ln \left( \frac{Q(x)}{P(x)} \right) \geq 0 \]

\[ Q(x) = P(x) \Rightarrow D[Q|P] = 0 \]

\[ D[Q|P] = \sum_x Q(x) \sum_{ij \in N} \ln W_{ij}(x_i, x_j) + \sum_x Q(x) \ln Q(x) + \ln Z \]

\[ = F[Q] + \ln Z \]

Free Energy

\[ Z = \sum_x \prod_{ij \in N} W_{ij}(x_i, x_j) \]

\[ \min_Q \left\{ F[Q] \left| \sum_x Q(x) = 1 \right. \right\} = F[P] = -\ln Z \]

\( P(x_1, x_2, \cdots, x_L) = \frac{1}{Z} \prod_{ij \in N} W_{ij}(x_i, x_j) \)
確率伝搬法の情報論的解釈

\[ D[Q|P] = \sum_x Q(x) \ln \left( \frac{Q(x)}{P(x)} \right) \geq 0 \]

\[ P(x) = \frac{1}{Z} \prod_{ij \in E} W_{ij}(x_i, x_j) \]

KL Divergence

Free Energy

\[ D[Q|P] = F[Q] + \ln(Z) \]

\[ F[Q] = \sum_x Q(x) \sum_{ij \in E} \ln W_{ij}(x_i, x_j) + \sum_x Q(x) \ln Q(x) \]

\[ Q_{ij}(x_i, x_j) = \sum_{x \backslash \{x_i, x_j\}} Q(x) \]

\[ = \sum_{ij \in E} \sum_{x_i} \sum_{x_j} \left( \sum_{x \backslash \{x_i, x_j\}} Q(x) \right) \ln W_{ij}(x_i, x_j) + \sum_{x} Q(x) \ln Q(x) \]
確率伝搬法の情報論的解釈

$$D[Q|P] = F[Q] + \ln(Z)$$

$$F[Q] = \sum_{i,j \in E} \sum_{\xi, \zeta} Q_{ij}(\xi, \zeta) \ln W_{ij}(\xi, \zeta)$$

$$+ \sum_{x} Q(x) \ln Q(x)$$

$$\approx \sum_{i,j \in E} \sum_{\xi, \zeta} Q_{ij}(\xi, \zeta) \ln W_{ij}(\xi, \zeta)$$

$$+ \sum_{i \in V} Q_{i}(\xi) \ln Q_{i}(\xi)$$

$$+ \sum_{i,j \in E} \left( \sum_{\xi, \zeta} Q_{ij}(\xi, \zeta) \ln Q_{ij}(\xi, \zeta) - Q_{i}(\xi) \ln Q_{i}(\xi) - Q_{j}(\xi) \ln Q_{j}(\xi) \right)$$

$$P(x) = \frac{1}{Z} \prod_{i,j \in E} W_{ij}(x_i, x_j)$$

$$Q_i(x_i) \equiv \sum_{x \mid x_i} Q(x)$$

$$Q_{ij}(x_i, x_j) \equiv \sum_{x \mid \{x_i, x_j\}} Q(x)$$

Bethe Free Energy
確率伝搬法の情報論的解釈

\[ D[Q|P] \approx F_{\text{Bethe}} \{Q_i, Q_{ij}\} + \ln Z \]

\[ F_{\text{Bethe}} \{Q_i, Q_{ij}\} = \sum_{ij \in E} \sum_{\xi} \sum_{\zeta} Q_{ij}(\xi, \zeta) \ln W_{ij}(\xi, \zeta) + \sum_{i \in V} \sum_{\xi} Q_i(\xi) \ln Q_i(\xi) \]

\[ + \sum_{ij \in E} \left( \sum_{\xi} \sum_{\zeta} Q_{ij}(\xi, \zeta) \ln Q_{ij}(\xi, \zeta) - \sum_{\xi} Q_i(\xi) \ln Q_i(\xi) - \sum_{\xi} Q_j(\xi) \ln Q_j(\xi) \right) \]

\[ \arg \min_{Q} D[Q|P] \Rightarrow \arg \min_{\{Q_i, Q_{ij}\}} F_{\text{Bethe}} \{Q_i, Q_{ij}\} \]

\[ \sum_{\xi} Q_i(\xi) = \sum_{\xi} \sum_{\zeta} Q_{ij}(\xi, \zeta) = 1 \]

\[ Q_i(\xi) = \sum_{\zeta} Q_{ij}(\xi, \zeta) \]
確率伝搬法の情報論的解釈

\[
\text{argmin}_{\{Q_i, Q_{ij}\}} \left\{ F_{\text{Bethe}}[\{Q_i, Q_{ij}\}] \mid Q_i(\xi) = \sum_{\zeta} Q_{ij}(\xi, \zeta), \sum_{\xi} Q_i(\xi) = \sum_{\xi, \zeta} Q_{ij}(\xi, \zeta) = 1 \right\}
\]

Lagrange Multipliers to ensure the constraints

\[
L_{\text{Bethe}}[\{Q_i, Q_{ij}\}] = F_{\text{Bethe}}[\{Q_i, Q_{ij}\}]
\]

\[
- \sum_{i \in V} \sum_{j \in \partial i} \sum_{\xi} Q_{i,j}(\xi) \left( Q_i(\xi) - \sum_{\zeta} Q_{ij}(\xi, \zeta) \right)
\]

\[
- \sum_{i \in V} \left( \sum_{\xi} Q_i(\xi) - 1 \right) - \sum_{ij \in E} v_{ij} \left( \sum_{\xi} \sum_{\zeta} Q_{ij}(\xi, \zeta) - 1 \right)
\]
確率伝搬法の情報論的解釈

\[ L_{\text{Bethe}} \left[ \{Q_i, Q_{ij} \} \right] = F_{\text{Bethe}} \left[ \{Q_i, Q_{ij} \} \right] - \sum_{i \in V} \sum_{j \in \partial i} \sum_{\xi \in \mathcal{V}} \lambda_{i,j}(\xi) \left( Q_i(\xi) - \sum_{\zeta} Q_{ij}(\xi, \zeta) \right) \]

\[ - \sum_{i \in V} \nu_i \left( \sum_{\xi} Q_i(\xi) - 1 \right) - \sum_{ij \in E} \nu_{ij} \left( \sum_{\xi} Q_{ij}(\xi, \zeta) - 1 \right) \]

\[ = \sum_{ij \in E} \left( \sum_{\xi} \sum_{\zeta} Q_{ij}(\xi, \zeta) \ln W_{ij}(\xi, \zeta) + \sum_{i \in V} \sum_{\xi} Q_i(\xi) \ln Q_i(\xi) \right) + \sum_{ij \in E} \left( \sum_{\xi} \sum_{\zeta} Q_{ij}(\xi, \zeta) \ln Q_{ij}(\xi, \zeta) - \sum_{\xi} Q_i(\xi) \ln Q_i(\xi) - \sum_{\xi} Q_j(\xi) \ln Q_j(\xi) \right) \]

\[ - \sum_{i \in V} \sum_{j \in \partial i} \sum_{\xi \in \mathcal{V}} \lambda_{i,j}(\xi) \left( Q_i(\xi) - \sum_{\zeta} Q_{ij}(\xi, \zeta) \right) - \sum_{i \in V} \nu_i \left( \sum_{\xi} Q_i(\xi) - 1 \right) - \sum_{ij \in E} \nu_{ij} \left( \sum_{\xi} \sum_{\zeta} Q_{ij}(\xi, \zeta) - 1 \right) \]

Extremum Condition

\[ \frac{\partial}{\partial Q_i(x_i)} L_{\text{Bethe}} \left[ \{Q_i, Q_{ij} \} \right] = 0 \]

\[ \frac{\partial}{\partial Q_{ij}(x_i, x_j)} L_{\text{Bethe}} \left[ \{Q_i, Q_{ij} \} \right] = 0 \]
確率伝搬法の情報論的解釈

\[
\frac{\partial}{\partial Q_i(x_i)} L_{\text{Bethe}} \left[ \{Q_i, Q_{ij}\} \right] = 0
\]

\[
\frac{\partial}{\partial Q_{ij}(x_i, x_j)} L_{\text{Bethe}} \left[ \{Q_i, Q_{ij}\} \right] = 0
\]

\[
Q_1(x_1) \propto M_{2 \rightarrow 1}(x_1) M_{3 \rightarrow 1}(x_1) \\
\times M_{4 \rightarrow 1}(x_1) M_{5 \rightarrow 1}(x_1)
\]

\[
Q_{12}(x_1, x_2) \propto M_{3 \rightarrow 1}(x_1) M_{4 \rightarrow 1}(x_1) M_{5 \rightarrow 1}(x_1) \\
\times W_{12}(x_1, x_2) \\
\times M_{6 \rightarrow 2}(x_2) M_{7 \rightarrow 2}(x_2) M_{8 \rightarrow 2}(x_2)
\]

Extremum Condition
確率伝搬法の情報論的解釈

\[ Q_1(x_1) = \sum Q_{12}(x_1, x_2) \]

\[ M_{2\to1}(x_1) \propto \sum W_{12}(x_1, x_2) M_{6\to2}(x_2) \]
\[ \times M_{7\to2}(x_2) M_{8\to2}(x_2) \]

Message Update Rule

\[ Q_1(x_1) \propto M_{2\to1}(x_1) M_{3\to1}(x_1) \]
\[ \times M_{4\to1}(x_1) M_{5\to1}(x_1) \]

\[ Q_{12}(x_1, x_2) \propto M_{3\to1}(x_1) M_{4\to1}(x_1) M_{5\to1}(x_1) \]
\[ \times W_{12}(x_1, x_2) \]
\[ \times M_{6\to2}(x_2) M_{7\to2}(x_2) M_{8\to2}(x_2) \]
確率伝搬法の情報論的解釈

\[
M_{1\rightarrow 2}(x_2) = \sum_{x_1} \sum_{x_2} W_{12}(x_1, x_2) M_{3\rightarrow 1}(x_1) M_{4\rightarrow 1}(x_1) M_{5\rightarrow 1}(x_1)
\]

Message Passing Rule of Belief Propagation

\[
\sum_{x_1} = 12, 115, 114, 113, 112 \rightarrow M_1 \rightarrow 2
\]

\[
1, 33, 44, 5, 13 \rightarrow M_3 \rightarrow 1
\]

\[
21, 1 \rightarrow M_4 \rightarrow 1
\]

\[
1, 2, 7, 6 \rightarrow M_5 \rightarrow 1
\]
統計物理学による確率伝搬法の一般化

一般化された確率伝搬法

クラスター変分法という統計物理学の手法がその一般化の鍵
確率伝搬法の統計物理学的位置付け

木構造を持つグラフィカルモデルではベーテ近似は転送行列法と等価である。

閉路を持つグラフィカルモデル上のベイジアンネットでの確率伝搬法はベーテ近似またはその拡張版であるクラスター変分法に等価である（Yedidia, Weiss and Freeman, NIPS2000）。

ベーテ近似
転送行列法（木構造）
もともとの確率伝搬法
クラスター変分法（菊池近似）
一般化された確率伝搬法
統計的性能の解析的評価

\[
\text{MSE}(\alpha \mid \sigma, \bar{x}) = \frac{1}{|V|} \int \left\| h(\bar{y}, \alpha, \sigma) - \bar{x} \right\|^2 \Pr\{\bar{Y} = \bar{y} \mid \bar{X} = \bar{x}, \sigma\} \, d\bar{y}
\]
統計的性能の解析的評価

\[
MSE(\alpha \mid \sigma, \bar{x}) \equiv \frac{1}{|V|} \int \left\| \hat{h}(\bar{y}, \alpha, \sigma) - \bar{x} \right\|^2 \Pr\{\bar{Y} = \bar{y} \mid \bar{X} = \bar{x}\} d\bar{y}
\]

\[
= \frac{1}{|V|} \int \left\| (I + \alpha \sigma^2 C)^{-1} \bar{y} - \bar{x} \right\|^2 \Pr\{\bar{Y} = \bar{y} \mid \bar{X} = \bar{x}\} d\bar{y}
\]

\[
\hat{h}(\bar{y}, \alpha, \sigma) = \int \bar{x} P(\bar{x} \mid \bar{y}) d\bar{x}
\]

\[
= (I + \alpha \sigma^2 C)^{-1} \bar{y}
\]

\[
\Pr\{\bar{Y} = \bar{y} \mid \bar{X} = \bar{x}, \sigma\} \propto \prod_{i \in V} \exp\left( -\frac{1}{2\sigma^2} (x_i - y_i)^2 \right)
\]

\[
= \exp\left( -\frac{1}{2\sigma^2} \left\| \bar{x} - \bar{y} \right\|^2 \right)
\]

\[
\langle i \mid C \mid j \rangle = \begin{cases} 
4, & i = j \in V \\
-1, & \{i, j\} \in E \\
0, & \text{otherwise}
\end{cases}
\]
Statistical Performance Estimation for Gaussian Markov Random Fields

\[ \text{MSE}(\alpha | \sigma, \tilde{x}) = \frac{1}{|V|} \int \left[ \left( \frac{I}{I + \alpha \sigma^2 C} \tilde{y} - \tilde{x} \right)^2 \times \Pr \{ \tilde{y} = \tilde{y} | \tilde{X} = \tilde{x}, \sigma \} \right] d\tilde{y} \]

\[ \begin{align*}
&= \frac{1}{|V|} \int \left[ \frac{I}{I + \alpha \sigma^2 C} \tilde{y} - \tilde{x} \right]^2 \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^{|V|} \exp \left( -\frac{1}{2\sigma^2} \| \tilde{y} - \tilde{x} \|^2 \right) d\tilde{y} \\
&= \frac{1}{|V|} \int \left[ \frac{I}{I + \alpha \sigma^2 C} (\tilde{y} - \tilde{x}) + \frac{\alpha \sigma^2 C}{I + \alpha \sigma^2 C} \tilde{x} \right]^2 \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^{|V|} \exp \left( -\frac{1}{2\sigma^2} \| \tilde{y} - \tilde{x} \|^2 \right) d\tilde{y} \\
&= \frac{1}{|V|} \int \left[ \frac{I}{I + \alpha \sigma^2 C} (\tilde{y} - \tilde{x}) + \frac{\alpha \sigma^2 C}{I + \alpha \sigma^2 C} \tilde{x} \right]^T \left( \frac{I}{I + \alpha \sigma^2 C} (\tilde{y} - \tilde{x}) + \frac{\alpha \sigma^2 C}{I + \alpha \sigma^2 C} \tilde{x} \right) \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^{|V|} \exp \left( -\frac{1}{2\sigma^2} \| \tilde{y} - \tilde{x} \|^2 \right) d\tilde{y} \\
&= \frac{1}{|V|} \int (\tilde{y} - \tilde{x})^T \left( \frac{I}{I + \alpha \sigma^2 C} (\tilde{y} - \tilde{x}) \right) \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^{|V|} \exp \left( -\frac{1}{2\sigma^2} \| \tilde{y} - \tilde{x} \|^2 \right) d\tilde{y} \\
&\quad - \frac{1}{|V|} \int (\tilde{y} - \tilde{x})^T \frac{\alpha \sigma^2 C}{I + \alpha \sigma^2 C} \tilde{x} \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^{|V|} \exp \left( -\frac{1}{2\sigma^2} \| \tilde{y} - \tilde{x} \|^2 \right) d\tilde{y} \\
&\quad - \frac{1}{|V|} \int \tilde{x}^T \frac{\alpha \sigma^2 C}{I + \alpha \sigma^2 C} \tilde{x} \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^{|V|} \exp \left( -\frac{1}{2\sigma^2} \| \tilde{y} - \tilde{x} \|^2 \right) d\tilde{y} \\
&\quad + \frac{1}{|V|} \int \tilde{x}^T \frac{\alpha \sigma^2 C}{I + \alpha \sigma^2 C} \tilde{x} \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^{|V|} \exp \left( -\frac{1}{2\sigma^2} \| \tilde{y} - \tilde{x} \|^2 \right) d\tilde{y} \\
&= \frac{1}{|V|} \text{Tr} \left( \frac{\sigma^2 I}{I + \alpha \sigma^2 C} \right) + \frac{1}{|V|} \tilde{x}^T \frac{\alpha \sigma^2 C^2}{(I + \alpha \sigma^2 C)^2} \tilde{x} \\
&\quad \{ i | C \} = \begin{cases}
4, & i = j \in V \\
-1, & \{ i, j \} \in E \\
0, & \text{otherwise}
\end{cases}
\]
Statistical Performance Estimation for Gaussian Markov Random Fields

\[
\text{MSE}(\alpha \mid \sigma, \bar{x}) = \frac{1}{|V|} \int \left\| \mathbf{h}(\bar{y}, \alpha, \sigma) - \bar{x} \right\|^2 \times \text{Pr}\{\bar{Y} = \bar{y} \mid \bar{X} = \bar{x}, \sigma\} \, d\bar{y}
\]

\[
= \frac{1}{|V|} \int \left\| \frac{I}{I + \alpha\sigma^2 \mathbf{C}} \bar{y} - \bar{x} \right\|^2 \times \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^{|V|} \exp \left( -\frac{1}{2\sigma^2} \left\| \bar{x} - \bar{y} \right\|^2 \right) d\bar{y}
\]

\[
= \frac{1}{|V|} \text{Tr} \left( \frac{\sigma^2 I}{(I + \alpha\sigma^2 \mathbf{C})^2} \right) + \frac{1}{|V|} \bar{x}^T \frac{\alpha^2 \sigma^4 \mathbf{C}^2}{(I + \alpha\sigma^2 \mathbf{C})^2} \bar{x}
\]

\[
\langle i \mid \mathbf{C} \mid j \rangle = \begin{cases} 
4, & i = j \in V \\
-1, & \{i, j\} \in E \\
0, & \text{otherwise}
\end{cases}
\]
Statistical Performance Estimation for Binary Markov Random Fields

$$\text{MSE}(\alpha \mid \sigma, \tilde{x}) = \frac{1}{|V|} \int \| \tilde{h} (\tilde{y}, \alpha, \sigma) - \tilde{x} \|^2 \Pr \{ \tilde{Y} = \tilde{y} \mid \tilde{X} = \tilde{x}, \sigma \} \, d\tilde{y}$$

$$s_i = \pm 1$$

Light intensities of the original image can be regarded as spin states of ferromagnetic system.

$$\Pr \{ \tilde{X} = \tilde{s} \mid \tilde{Y} = \tilde{y}, \alpha, \sigma \} \propto \exp \left( \frac{1}{\sigma^2} \sum_{i \in V} y_i s_i + \alpha \sum_{\{i, j\} \in E} s_i s_j \right)$$

Free Energy of Ising Model with Random External Fields

It can be reduced to the calculation of the average of free energy with respect to locally non-uniform external fields $y_1, y_2, \ldots, y_{|V|}$. 

$$\log \left( \sum_{s_1 = \pm 1} \sum_{s_2 = \pm 1} \cdots \sum_{s_{|V|} = \pm 1} \exp \left( \frac{1}{2 \sigma^2} \sum_{i = 1}^{|V|} y_i s_i + \alpha \sum_{\{i, j\} \in E} s_i s_j \right) \right)$$

where $s_i$ is the spin state of the $i$-th pixel in the image.
Statistical Performance Estimation for Markov Random Fields

\[ \text{MSE}(\alpha \mid \sigma, \bar{x}) = \frac{1}{|V|} \int \left\| \tilde{h}(\bar{y}, \alpha, \sigma) - \bar{x} \right\|^2 \Pr\{\bar{Y} = \bar{y} \mid \bar{X} = \bar{x}, \sigma\} \, d\bar{y} \]

Spin Glass Theory in Statistical Mechanics
- Loopy Belief Propagation

Multi-dimensional Gauss Integral Formulas

\[ \text{MSE}(\alpha \mid \sigma, \bar{x}) \]

\[ \text{MSE}(\alpha \mid \sigma, \bar{x}) \]

\( \sigma = 1 \)

\( \sigma = 40 \)