Introduction to Probabilistic Machine Learning

Ralf Herbrich, Rainer Schlosser

Tutorial 5 – Recap Theory Unit 5

Overview

1. **Questions and Updates**
2. Recap: Main Concepts of Unit 5
3. Example: KL Divergence
4. Example: Iterating Approximated Factors
5. Example: Normalization Constant

Course Overview

Week	Topic Lecture	Tutorial	Exercises
07.04. & 08.04.	1 Probability Theory	Intro Julia	
14.04. & 15.04.	2 Inference Methods and Decision-Making	no tutorial	Exercise 1
21.04. & 22.04.	**no lecture**	Theory Unit 1 & 2	(14.04. – 08.05.)
28.04. & 29.04.	3 Graphical Models: Independence	Theory Unit 3	
05.05. & 06.05.	4 Graphical Models: Exact Inference	Theory Unit 4	**Exercise 2**
12.05. & 13.05.	**5 Graphical Models: Approximate Inference**	**Theory Unit 5**	(05.05. – **19.05.**)
19.05. & 20.05.	6 Bayesian Ranking	Theory Unit 6	Exercise 3
26.05. & 27.05.	7 Linear Basis Function Models	Theory Unit 7	(19.05. – 02.06.)
02.06. & 03.06.	8 Bayesian Regression	Theory Unit 8	Exercise 4
09.06. & 10.06.	**no lecture**	9 Bayesian Classification	(02.06. – 23.06.)
16.06. & 17.06.	10 Non-Bayesian Classification Learning	Theory Unit 9 & 10	
23.06. & 24.06.	11 Gaussian Processes	Theory Unit 11	Exercise 5
30.06. & 01.07.	12 Information Theory	Theory Unit 12	(23.06. – 07.07.)
07.07. & 08.07.	13 Real-World Applications		

Overview

1. Questions and Updates
2. **Recap: Main Concepts of Unit 5**
3. Example: KL Divergence
4. Example: Iterating Approximated Factors
5. Example: Normalization Constant

Sum-Product Algorithm Revisited

- The key operation for factor $f(x_1, x_2, \dots, x_n)$ and variable X_1 is

$$m_{f\to X_1}(x_1) = \sum_{\{x_2\}} \cdots \sum_{\{x_n\}} f(x_1, x_2, \dots, x_n) \prod_{j=2}^{n} m_{X_j\to f}(x_j)$$

If all $m_{X_j\to f}(x_j)$ are Gaussian, the result **might not be** Gaussian!

- Based on outgoing messages, we can compute both non-normalized marginals $p_X(\cdot)$ and $m_{X\to f}(\cdot)$

$$p_X(x) = \prod_{f\in\mathrm{ne}(X)} m_{f\to X}(x) \qquad m_{X\to f}(x) = \frac{p_X(x)}{m_{f\to X}(x)}$$

If all $m_{X_j\to f}(x_j)$ are Gaussian, the result **must be** Gaussian!

- **Idea**:
  1. We approximate all outgoing messages $m_{f\to X}(\cdot)$ by a Gaussian $\widehat{m}_{f\to X}(\cdot) = \mathcal{N}(\cdot;\mu,\sigma^2)$
  2. We measure the approximation quality in the normalized marginal, **not** the outgoing message

$$\hat{p}(\cdot) = \arg\min_{\mu,\sigma^2} \mathrm{KL}\left[\frac{m_{f\to X}(\cdot)\cdot\widehat{m}_{X\to f}(\cdot)}{\int_{-\infty}^{+\infty} m_{f\to X}(\tilde{x})\cdot\widehat{m}_{X\to f}(\tilde{x})\,\mathrm{d}\tilde{x}}, \frac{\mathcal{N}(\cdot;\mu,\sigma^2)\cdot\widehat{m}_{X\to f}(\cdot)}{\int_{-\infty}^{+\infty}\mathcal{N}(\tilde{x};\mu,\sigma^2)\cdot\widehat{m}_{X\to f}(\tilde{x})\,\mathrm{d}\tilde{x}}\right]$$

True normalized marginal with approximate incoming message

Approximate marginal with approximate incoming message

Approximate Message Passing: Example

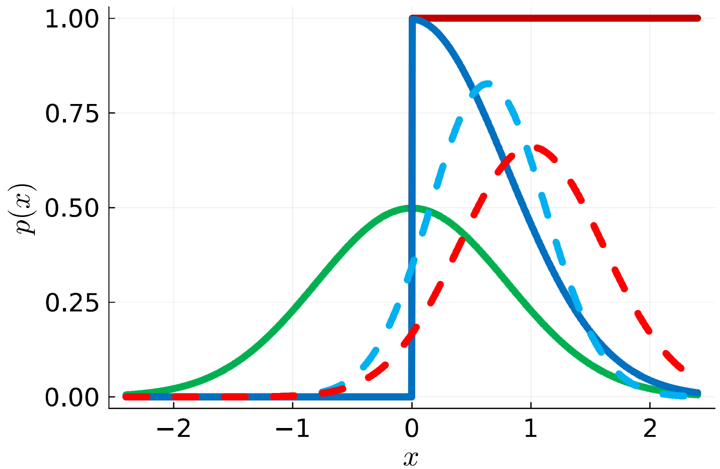
$$\boxed{f(x) = \mathbb{I}(x > 0)}$$

$$\hat{m}_{X\to f}(x) \propto \frac{\hat{p}_X(x)}{\hat{m}_{f\to X}(x)}$$

$$p_X(x) \propto f(x) \cdot \hat{m}_{X\to f}(x)$$

$$\hat{p}_X(x) = \mathcal{N}\left(x; E_{X\sim p_X}[X], \mathrm{var}_{X\sim p_X}[X]\right)$$

$$\hat{m}_{f\to X}(x) \propto \frac{\hat{p}_X(x)}{\hat{m}_{X\to f}(x)}$$



Generalization: Doubly-Truncated Gaussians

- **Doubly-Truncated Gaussian**. *Given* $l, u \in \mathbb{R}$, $\mu \in \mathbb{R}$ *and* $\sigma^2 \in \mathbb{R}^+$, *a random variable* X *has a* doubly-truncated Gaussian distribution *if*

$$p_X(x) \propto \mathbb{I}(l < x < u) \cdot \mathcal{N}(x; \mu, \sigma^2)$$

- **Moments of Doubly-Truncated Gaussian**. *Given a random variable* X *that has a doubly-truncated Gaussian distribution and* $t_a \coloneqq a/\sigma$, *we know*

$$E[X^0] = \Phi(t_{u-\mu}) - \Phi(t_{l-\mu})$$

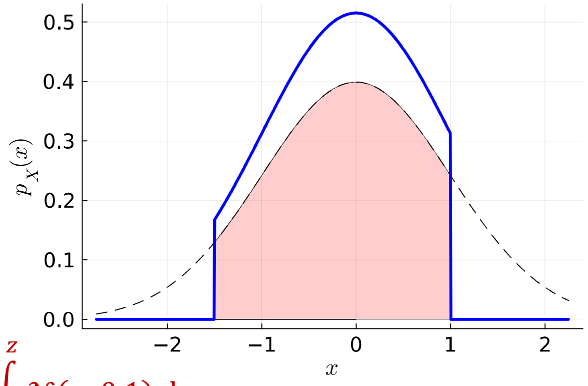
$$\Phi(z) \coloneqq \int_{-\infty}^{z} \mathcal{N}(x; 0,1)\, \mathrm{d}x$$

$$E[X^1] = \mu + \sigma \cdot \frac{\mathcal{N}(t_{l-\mu}) - \mathcal{N}(t_{u-\mu})}{\Phi(t_{u-\mu}) - \Phi(t_{l-\mu})}$$

Additive correction that goes to zero as $u \to \infty$ and $l \to -\infty$

$$E[X^2] = \mu^2 + \sigma^2 \cdot \left[1 - \frac{t_{u+\mu} \cdot \mathcal{N}(t_{u-\mu}) - t_{l+\mu} \cdot \mathcal{N}(t_{l-\mu})}{\Phi(t_{u-\mu}) - \Phi(t_{l-\mu})}\right]$$

Multiplicative correction that goes to one as $u \to \infty$ and $l \to -\infty$



Introduction to Probabilistic Machine Learning

Unit 5 – Graphical Models: Approximate Inference

Overview

1. Questions and Updates
2. Recap: Main Concepts of Unit 5
3. **Example: KL Divergence**
4. Example: Iterating Approximated Factors
5. Example: Normalization Constant

$\alpha = 1$: KL Divergence

- **Theorem (Limit $\alpha \to 1$).** *Given two probability densities $p(\cdot)$ and $q(\cdot)$ the limit of the α-divergence $D_\alpha[p,q]$ for $\alpha \to 1$ is the Kullback-Leibler divergence*

$$\lim_{\alpha \to 1} D_\alpha[p,q] = \mathrm{KL}[p,q] := \int_{-\infty}^{+\infty} \log\left(\frac{p(x)}{q(x)}\right) \cdot p(x)\,\mathrm{d}x$$

  - **Proof**: Taking limits, we have

$$\lim_{\alpha \to 1} D_\alpha[p,q] = \lim_{\alpha \to 1} \frac{1}{\alpha(1-\alpha)} \cdot \left(1 - \int_{-\infty}^{+\infty} \left[\frac{p(x)}{q(x)}\right]^\alpha \cdot q(x)\,\mathrm{d}x\right)$$

L'Hôpital's rule!

Note that $\frac{\mathrm{d}}{\mathrm{d}z} b^z = \log(b) \cdot b^z$

$$= \lim_{\alpha \to 1} \frac{1}{1-2\alpha} \cdot \left(-\int_{-\infty}^{+\infty} \log\left(\frac{p(x)}{q(x)}\right) \cdot \left[\frac{p(x)}{q(x)}\right]^\alpha \cdot q(x)\,\mathrm{d}x\right) = \int_{-\infty}^{+\infty} \log\left(\frac{p(x)}{q(x)}\right) \cdot p(x)\,\mathrm{d}x$$

- **Theorem (Moment Matching)**. *Given any distribution $p(\cdot)$ the minimizer $\mu^*, {\sigma^2}^*$ of the KL divergence $\mathrm{KL}[p(\cdot), \mathcal{N}(\cdot;\mu,\sigma^2)]$ to a Gaussian distribution has*

$$\mu^* = E_{X \sim p(\cdot)}[X] \quad \text{and} \quad {\sigma^2}^* = E_{X \sim p(\cdot)}[X^2] - (\mu^*)^2$$

Hasso Plattner Institut

Digital Engineering · Universität Potsdam

Solomon Kullback (1909 – 1994)

Richard Leibler (1914 – 2003)

Introduction to Probabilistic Machine Learning

Unit 5 – Graphical Models: Approximate Inference

KL Divergence & Alternatives

Question: What is the difference between two (discrete) distributions?

Ideas??

KL Divergence & Alternatives

Question: What is the difference between two (discrete) distributions?

Ideas??

Candidates: MSE RMSE MAE MAPE KL

$$\frac{1}{N}\sum_x (p(x)-q(x))^2$$

$$\sqrt{\frac{1}{N}\sum_x (p(x)-q(x))^2}$$

$$\frac{1}{N}\sum_x \left|p(x)-q(x)\right|$$

$$\frac{1}{N}\sum_x \left|\frac{p(x)-q(x)}{p(x)}\right|$$

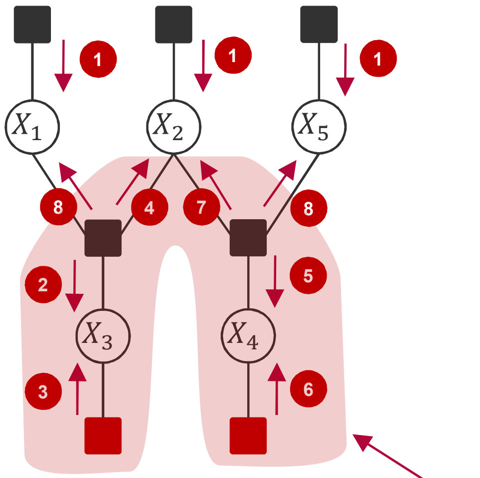
$$\sum_x p(x)\cdot\log\left(\frac{p(x)}{q(x)}\right)$$

Pros & Cons?

Overview

Expectation Propagation

- **Idea**: If we have factors in the factor graph that require approximate messages, we keep iterating on the whole path between them until convergence minimizing $\mathrm{KL}\big(p(\cdot)|\mathcal{N}(\cdot;\mu,\sigma^2)\big)$ locally for the affected marginals of the approximate factor.
- **Theorem (Minka, 2003)**: *The approximate message passing algorithm using the Kullback-Leibler divergence will always converge if the approximating distribution is in the exponential family!*



iterate until convergence

Tom Minka

KL Divergence & Alternatives

Numerical Example (discrete):

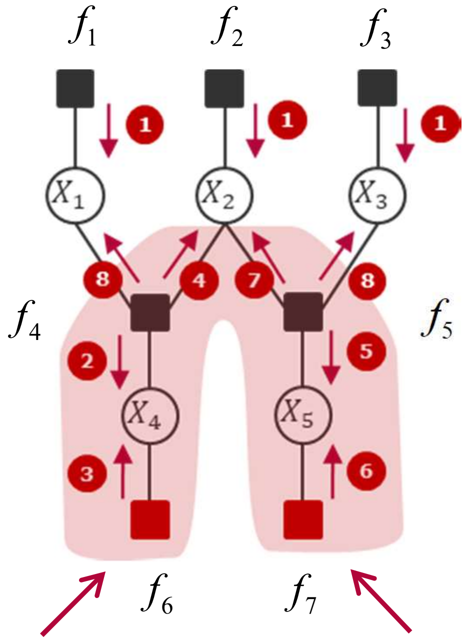
$$f_i(x_i) := (2\pi\sigma_i^2)^{-1/2} \cdot \exp\left(-\frac{1}{2\sigma_i^2}(x_i - \mu_i)^2 \right), \;\; x_i = 1, \ldots, N, \;\; i = 1, .., 3$$

$$f_4(x_1, x_2, x_4) := 1_{\{x_4 = x_1 - x_2\}}, \;\; x_1, x_2 = 1, \ldots, N, \;\; x_4 = -N, \ldots, N$$

$$f_5(x_2, x_3, x_5) := 1_{\{x_5 = x_2 - x_3\}}, \;\; x_2, x_3 = 1, \ldots, N, \;\; x_5 = -N, \ldots, N$$

$$f_6(x_4) := 1_{\{x_4 > 0\}}, \;\; x_4 = -N, \ldots, N$$

$$f_7(x_5) := 1_{\{x_5 > 0\}}, \;\; x_5 = -N, \ldots, N$$



KL Divergence & Alternatives

Numerical Example (discrete):

$$f_i(x_i) := (2\pi\sigma_i^2)^{-1/2} \cdot \exp\left(-\frac{1}{2\sigma_i^2}(x_i-\mu_i)^2\right), \quad x_i = 1,\ldots,N, \quad i = 1,..,3$$

$$f_4(x_1,x_2,x_4) := 1_{\{x_4 = x_1 - x_2\}}, \quad x_1, x_2 = 1,\ldots,N, \quad x_4 = -N,\ldots,N$$

$$f_5(x_2,x_3,x_5) := 1_{\{x_5 = x_2 - x_3\}}, \quad x_2, x_3 = 1,\ldots,N, \quad x_5 = -N,\ldots,N$$

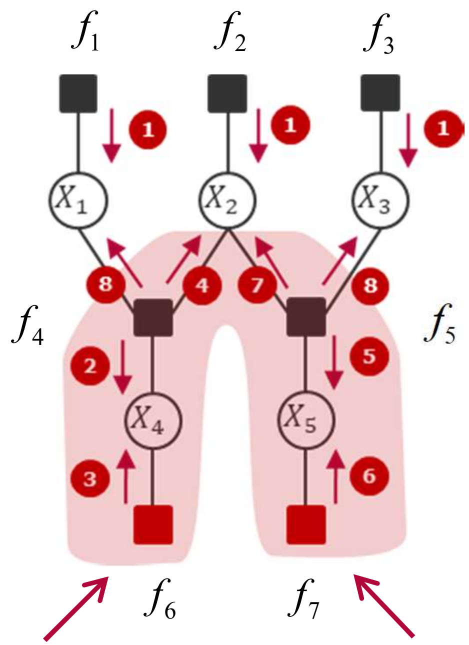
$$f_6(x_4) := 1_{\{x_4 > 0\}}, \quad x_4 = -N,\ldots,N$$

$$f_7(x_5) := 1_{\{x_5 > 0\}}, \quad x_5 = -N,\ldots,N$$

$$N = 12$$

$$\mu_1 = 8, \mu_2 = 6, \mu_3 = 4$$

$$\sigma_i^2 = 2, \quad i = 1,..,3$$



Tutorial 5

PML SS 2025

KL Divergence & Alternatives

Numerical Example (discrete):

$$f_i(x_i) := (2\pi\sigma_i^2)^{-1/2} \cdot \exp\left(-\frac{1}{2\sigma_i^2}(x_i-\mu_i)^2\right), \; x_i = 1,\dots,N, \; i = 1,..,3$$

$$f_4(x_1,x_2,x_4) := 1_{\{x_4 = x_1 - x_2\}}, \; x_1, x_2 = 1,\dots,N, \; x_4 = -N,\dots,N$$

$$f_5(x_2,x_3,x_5) := 1_{\{x_5 = x_2 - x_3\}}, \; x_2, x_3 = 1,\dots,N, \; x_5 = -N,\dots,N$$

$$f_6(x_4) := 1_{\{x_4 > 0\}}, \; x_4 = -N,\dots,N$$

$$f_7(x_5) := 1_{\{x_5 > 0\}}, \; x_5 = -N,\dots,N$$

$$N = 12$$

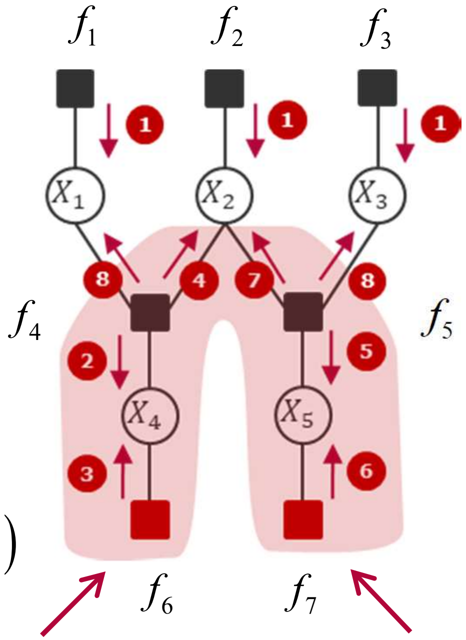
$$\mu_1 = 8, \mu_2 = 6, \mu_3 = 4$$

$$\sigma_i^2 = 2, \; i = 1,..,3$$

$$p(x_4) \Rightarrow \mu_4 := E(x_4), \; \sigma_4^2 := \sigma^2(x_4)$$

$$\Rightarrow \hat{p}(x_4) := (2\pi \cdot \sigma_4^2)^{-1/2} \cdot \exp\left(-1/2\sigma_4^{-2}(x_4-\mu_4)^2\right)$$

$$\Rightarrow m_{f_6 \to X_4}(x_4) := 1_{\{m_{f_6 \to X_4}(x_4) > 0\}} \cdot \hat{p}(x_4) / m_{X_4 \to f_6}(x_4)$$



KL Divergence & Alternatives

Numerical Example (discrete):

$$f_i(x_i) := (2\pi\sigma_i^2)^{-1/2} \cdot \exp\left(-\frac{1}{2\sigma_i^2}(x_i - \mu_i)^2\right), \ x_i = 1,\ldots,N, \ i = 1,..,3$$

$$f_4(x_1, x_2, x_4) := 1_{\{x_4 = x_1 - x_2\}}, \ x_1, x_2 = 1,\ldots,N, \ x_4 = -N,\ldots,N$$

$$f_5(x_2, x_3, x_5) := 1_{\{x_5 = x_2 - x_3\}}, \ x_2, x_3 = 1,\ldots,N, \ x_5 = -N,\ldots,N$$

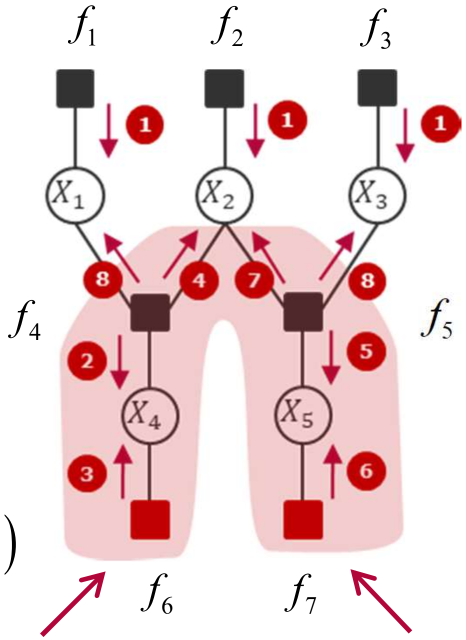
$$f_6(x_4) := 1_{\{x_4 > 0\}}, \ x_4 = -N,\ldots,N$$

$$f_7(x_5) := 1_{\{x_5 > 0\}}, \ x_5 = -N,\ldots,N$$

$$N = 12$$

$$\mu_1 = 8, \mu_2 = 6, \mu_3 = 4$$

$$\sigma_i^2 = 2, \ i = 1,..,3$$



$$p(x_4) \Rightarrow \mu_4 := E(x_4), \ \sigma_4^2 := \sigma^2(x_4)$$

$$\Rightarrow \hat{p}(x_4) := (2\pi \cdot \sigma_4^2)^{-1/2} \cdot \exp\left(-1/2\sigma_4^{-2}(x_4 - \mu_4)^2\right)$$

$$\Rightarrow m_{f_6 \to X_4}(x_4) := 1_{\{m_{f_6 \to X_4}(x_4) > 0\}} \cdot \hat{p}(x_4) / m_{X_4 \to f_6}(x_4)$$

Compare Convergence with and without Moment Matching:

Without:

With:

KL Divergence & Alternatives

Numerical Example (discrete):

$$f_i(x_i) := (2\pi\sigma_i^2)^{-1/2} \cdot \exp\left(-\frac{1}{2\sigma_i^2}(x_i - \mu_i)^2\right), \;\; x_i = 1,\ldots,N, \;\; i = 1,..,3$$

$$N = 12$$

$$\mu_1 = 8, \mu_2 = 6, \mu_3 = 4$$

$$\sigma_i^2 = 2, \;\; i = 1,..,3$$

$$f_4(x_1, x_2, x_4) := 1_{\{x_4 = x_1 - x_2\}}, \;\; x_1, x_2 = 1,\ldots,N, \;\; x_4 = -N,\ldots,N$$

$$f_5(x_2, x_3, x_5) := 1_{\{x_5 = x_2 - x_3\}}, \;\; x_2, x_3 = 1,\ldots,N, \;\; x_5 = -N,\ldots,N$$

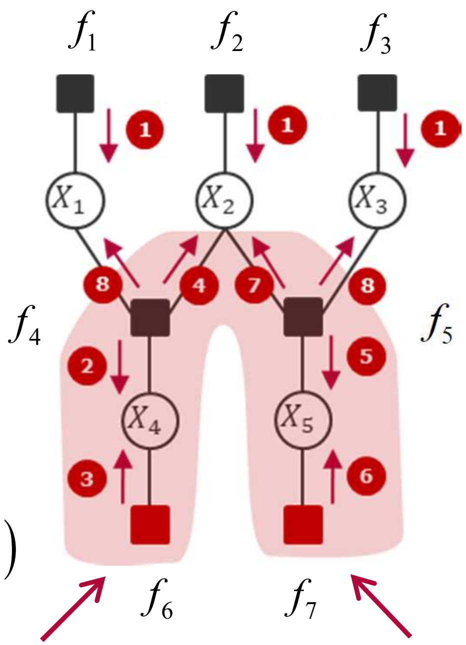
$$f_6(x_4) := 1_{\{x_4 > 0\}}, \;\; x_4 = -N,\ldots,N$$

$$f_7(x_5) := 1_{\{x_5 > 0\}}, \;\; x_5 = -N,\ldots,N$$

$$p(x_4) \Rightarrow \mu_4 := E(x_4), \;\; \sigma_4^2 := \sigma^2(x_4)$$

$$\Rightarrow \hat{p}(x_4) := (2\pi \cdot \sigma_4^2)^{-1/2} \cdot \exp\left(-1/2\sigma_4^{-2}(x_4 - \mu_4)^2\right)$$

$$\Rightarrow m_{f_6 \to X_4}(x_4) := 1_{\{m_{f_6 \to X_4}(x_4) > 0\}} \cdot \hat{p}(x_4) / m_{X_4 \to f_6}(x_4)$$



Compare Convergence with and without Moment Matching:

Without: 1 Iteration $\hat{\mu}_4 = 2.776, \;\; 3.86635$

With:

KL Divergence & Alternatives

Numerical Example (discrete):

$$f_i(x_i) := (2\pi\sigma_i^2)^{-1/2} \cdot \exp\left(-\frac{1}{2\sigma_i^2}(x_i-\mu_i)^2\right), \;\; x_i = 1,\dots,N,\;\; i=1,..,3$$

$$f_4(x_1,x_2,x_4) := 1_{\{x_4 = x_1 - x_2\}},\;\; x_1,x_2 = 1,\dots,N,\;\; x_4 = -N,\dots,N$$

$$f_5(x_2,x_3,x_5) := 1_{\{x_5 = x_2 - x_3\}},\;\; x_2,x_3 = 1,\dots,N,\;\; x_5 = -N,\dots,N$$

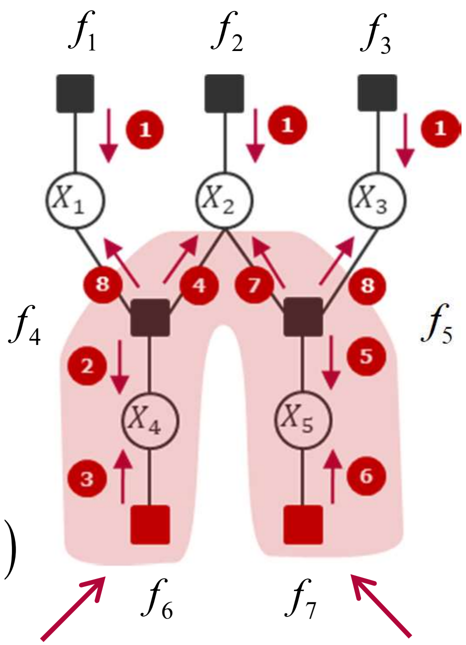
$$f_6(x_4) := 1_{\{x_4 > 0\}},\;\; x_4 = -N,\dots,N$$

$$f_7(x_5) := 1_{\{x_5 > 0\}},\;\; x_5 = -N,\dots,N$$

$$N = 12$$

$$\mu_1 = 8, \mu_2 = 6, \mu_3 = 4$$

$$\sigma_i^2 = 2,\;\; i = 1,..,3$$



$$p(x_4) \Rightarrow \mu_4 := E(x_4),\;\; \sigma_4^2 := \sigma^2(x_4)$$

$$\Rightarrow \hat{p}(x_4) := (2\pi \cdot \sigma_4^2)^{-1/2} \cdot \exp\left(-1/2\sigma_4^{-2}(x_4-\mu_4)^2\right)$$

$$\Rightarrow m_{f_6 \to X_4}(x_4) := 1_{\{m_{f_6 \to X_4}(x_4) > 0\}} \cdot \hat{p}(x_4) / m_{X_4 \to f_6}(x_4)$$

Compare Convergence with and without Moment Matching:

Without: 1 Iteration $\hat{\mu}_4 = 2.776,\;\; 3.86635$

With: 7 Iterations $\hat{\mu}_4 = 2.776,\;\; 3.270,\;\; 3.345,\;\; 3.357,\;\; 3.358,\;\; 3.35851,\;\; 3.35855,\;\; 3.35856$

Overview

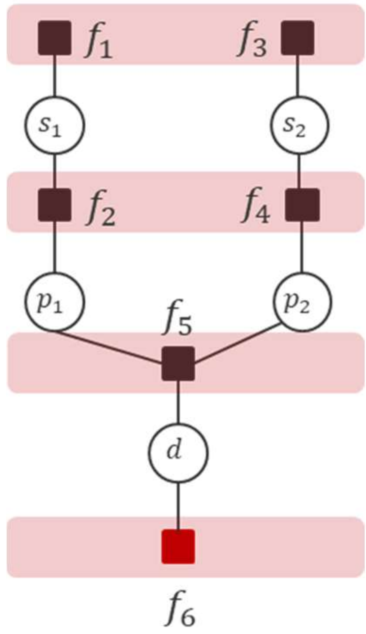
1. Questions and Updates
2. Recap: Main Concepts of Unit 5
3. Example: KL Divergence
4. Example: Iterating Approximated Factors
5. **Example: Normalization Constant**

Normalization Constant

Numerical Example (discrete):

(i) $$Z = \sum_{s_1}\sum_{s_2}\sum_{p_1}\sum_{p_2}\sum_{d} f_1(s_1)\cdot f_2(s_1,p_1)\cdot f_3(s_2)\cdot f_4(s_2,p_2)\cdot f_5(p_1,p_2,d) = 1$$

(ii) $$Z = \sum_{s_1}\sum_{s_2}\sum_{p_1}\sum_{p_2}\sum_{d} f_1(s_1)\cdot f_2(s_1,p_1)\cdot f_3(s_2)\cdot f_4(s_2,p_2)\cdot f_5(p_1,p_2,d)\cdot f_6(d) < 1$$



Normalization Constant

Numerical Example (discrete):

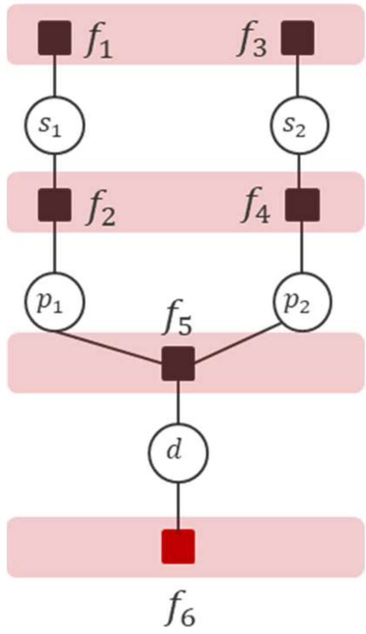
(i) $$Z = \sum_{s_1} \sum_{s_2} \sum_{p_1} \sum_{p_2} \sum_{d} f_1(s_1) \cdot f_2(s_1, p_1) \cdot f_3(s_2) \cdot f_4(s_2, p_2) \cdot f_5(p_1, p_2, d) = 1$$

(ii) $$Z = \sum_{s_1} \sum_{s_2} \sum_{p_1} \sum_{p_2} \sum_{d} f_1(s_1) \cdot f_2(s_1, p_1) \cdot f_3(s_2) \cdot f_4(s_2, p_2) \cdot f_5(p_1, p_2, d) \cdot f_6(d) < 1$$

Apply the Sum-Product-Algorithm WITHOUT any normalization

and obtain **unnormalized** marginals $\tilde{p}(s_1), \tilde{p}(s_2), \tilde{p}(p_1), \tilde{p}(p_2), \tilde{p}(d)$

(iii)



Normalization Constant

Numerical Example (discrete):

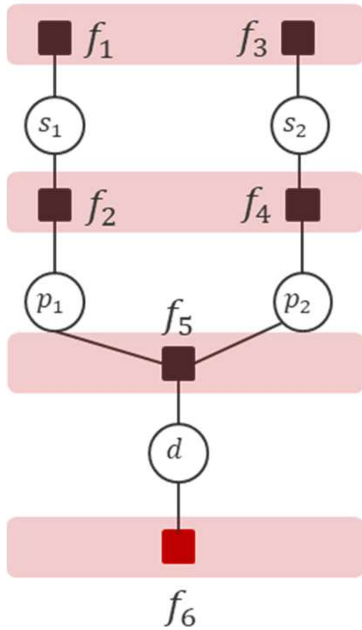
(i) $$Z = \sum_{s_1}\sum_{s_2}\sum_{p_1}\sum_{p_2}\sum_{d} f_1(s_1)\cdot f_2(s_1,p_1)\cdot f_3(s_2)\cdot f_4(s_2,p_2)\cdot f_5(p_1,p_2,d) = 1$$

(ii) $$Z = \sum_{s_1}\sum_{s_2}\sum_{p_1}\sum_{p_2}\sum_{d} f_1(s_1)\cdot f_2(s_1,p_1)\cdot f_3(s_2)\cdot f_4(s_2,p_2)\cdot f_5(p_1,p_2,d)\cdot f_6(d) < 1$$

Apply the Sum-Product-Algorithm WITHOUT any normalization

and obtain **unnormalized** marginals $\tilde{p}(s_1), \tilde{p}(s_2), \tilde{p}(p_1), \tilde{p}(p_2), \tilde{p}(d)$. Then:

(iii) $$Z = \sum_{s_1}\tilde{p}(s_1) = \sum_{s_2}\tilde{p}(s_2) = \sum_{p_1}\tilde{p}(p_1) = \sum_{p_2}\tilde{p}(p_2) = \sum_{d}\tilde{p}(d)$$

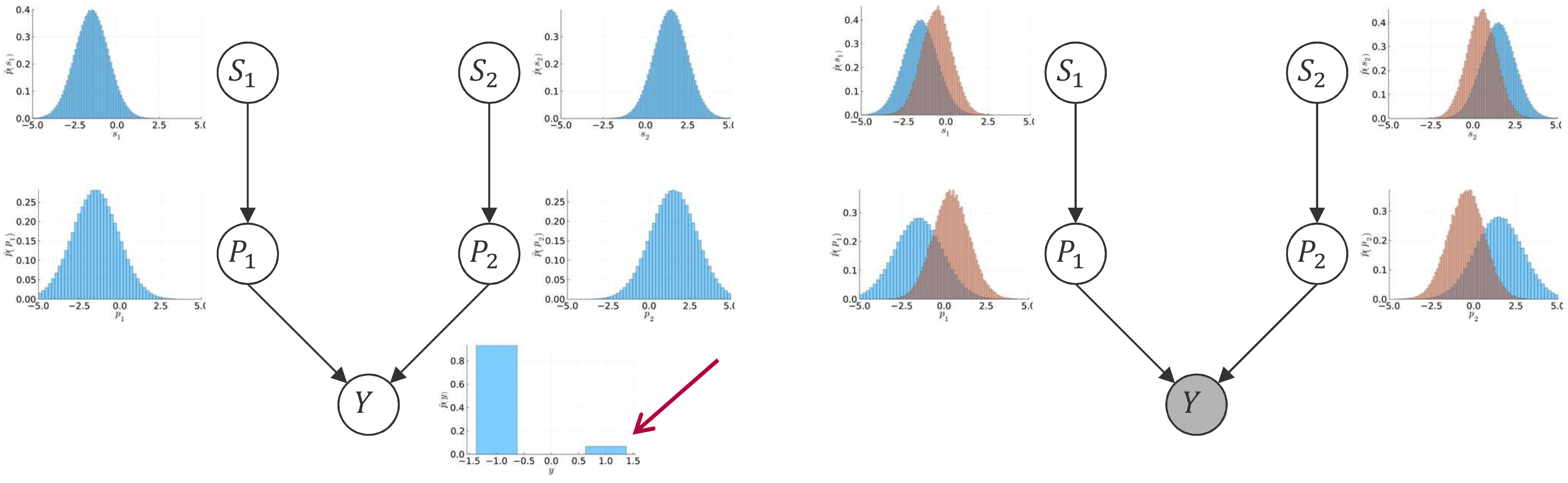


Sampling a Bayesian Network: Example (ctd)

$$\boxed{\mathcal{N}(s_1;\mu_1,\sigma_1^2)\cdot\mathcal{N}(s_2;\mu_2,\sigma_2^2)\cdot\mathcal{N}(p_1;s_1,\beta^2)\cdot\mathcal{N}(p_2;s_2,\beta^2)\cdot\mathbb{I}(y(p_1-p_2)>0)}$$

$\mu_1 \ll \mu_2$

Without match outcome **With match outcome ($y = 1$)**



Weighted Sums of Gaussian Random Variables

Linear Combination (**Addition**) of Gaussian RVs:

(i) $X \sim \mathcal{N}(\mu_X, \sigma_X^2) \qquad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

$$\Rightarrow a \cdot X + b \cdot Y \sim ??$$

Weighted Sums of Gaussian Random Variables

Linear Combination (**Addition**) of Gaussian RVs:

(i) $$X \sim \mathcal{N}(\mu_X, \sigma_X^2) \qquad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

$$\Rightarrow a \cdot X + b \cdot Y \sim \mathcal{N}(a \cdot \mu_X + b \cdot \mu_Y, a^2 \cdot \sigma_X^2 + b^2 \cdot \sigma_Y^2)$$

Weighted Sums of Gaussian Random Variables

Linear Combination (**Addition**) of Gaussian RVs:

(i) $$X \sim \mathcal{N}(\mu_X, \sigma_X^2) \qquad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

$$\Rightarrow a \cdot X + b \cdot Y \sim \mathcal{N}(a \cdot \mu_X + b \cdot \mu_Y, a^2 \cdot \sigma_X^2 + b^2 \cdot \sigma_Y^2)$$

Multiplication Theorem for **Products** of Gaussian Densities:

(ii)

Weighted Sums of Gaussian Random Variables

Linear Combination (**Addition**) of Gaussian RVs:

(i) $$X \sim \mathcal{N}(\mu_X, \sigma_X^2) \qquad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

$$\Rightarrow a \cdot X + b \cdot Y \sim \mathcal{N}(a \cdot \mu_X + b \cdot \mu_Y, a^2 \cdot \sigma_X^2 + b^2 \cdot \sigma_Y^2)$$

Multiplication Theorem for **Products** of Gaussian Densities:

(ii) $$\mathcal{N}(x; \mu_1, \sigma_1^2) \cdot \mathcal{N}(x; \mu_2, \sigma_2^2) \propto \mathcal{N}\left(x; \frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)$$

Weighted Sums of Gaussian Random Variables

Linear Combination (**Addition**) of Gaussian RVs:

(i) $$X \sim \mathcal{N}(\mu_X, \sigma_X^2) \qquad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

$$\Rightarrow a \cdot X + b \cdot Y \sim \mathcal{N}(a \cdot \mu_X + b \cdot \mu_Y, a^2 \cdot \sigma_X^2 + b^2 \cdot \sigma_Y^2)$$

Multiplication Theorem for **Products** of Gaussian Densities:

(ii) $$\mathcal{N}(x; \mu_1, \sigma_1^2) \cdot \mathcal{N}(x; \mu_2, \sigma_2^2) \propto \mathcal{N}\left(x; \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)$$

$$= \mathcal{N}\left(x; \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \cdot \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \cdot \mu_2, \underbrace{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}}_{=\left(\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^2 \cdot \sigma_1^2 + \left(\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2 \cdot \sigma_2^2 \ ??}\right)$$

Summary

- Recap I: KL Divergence / Moment Matching / Truncated Gaussians
- Recap II: Iterated Sum-Product-Algorithm (Approximate MP)
- Recap III: Normalization Constant
- Recap IV: Sums of Gaussians

See you next Week!