Fachgebiet Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 15

Exercise 1: RBM with Ternary Hidden Units (20 + 10 P)

We consider a variant of the restricted Boltzmann machine with ternary hidden units $\boldsymbol{h} \in \{-1, 0, 1\}^H$. Input features remain binary, i.e. $\boldsymbol{x} \in \{0,1\}^d$, like for the classical RBM. The probability model is given by:

$$p(\boldsymbol{x}, \boldsymbol{h}|\theta) = \frac{1}{\mathcal{Z}} \exp\Big(\sum_{j=1}^{H} \boldsymbol{w}_j^\top \boldsymbol{x} \cdot h_j + \sum_{j=1}^{H} h_j b_j\Big)$$

where $\theta = (\boldsymbol{w}_j, b_j)_{j=1}^H$ are the parameters of the model, and where and $\mathcal{Z}$ is the partition function that normalizes the probability distribution to 1.

(a) *Show* that this modified RBM can also be expressed as a product of experts

$$p(\boldsymbol{x}|\theta) = \frac{1}{\mathcal{Z}} \prod_{j=1}^{H} g_j(\boldsymbol{x}, \theta_j),$$

with

$$g_j(\boldsymbol{x}, \theta_j) = 1 + 2\cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j),$$

where cosh is the hyperbolic cosine function.

(b) *Show* that the gradient of the log-likelihood assigned to some data point $\boldsymbol{x}_n$ by the modified RBM has the form

$$\forall_{j=1}^H: \ \frac{\partial \log p(\boldsymbol{x}_n|\theta)}{\partial \boldsymbol{w}_j} = \boldsymbol{x}_n \cdot \sigma(\boldsymbol{w}_j^\top \boldsymbol{x}_n + b_j) - \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\theta)}\big[\boldsymbol{x} \cdot \sigma(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\big]$$

$$\forall_{j=1}^H: \ \frac{\partial \log p(\boldsymbol{x}_n|\theta)}{\partial b_j} = \sigma(\boldsymbol{w}_j^\top \boldsymbol{x}_n + b_j) - \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\theta)}\big[\sigma(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\big]$$

where $\sigma(t) = \dfrac{\sinh(t)}{0.5 + \cosh(t)}$.

Exercise 2: Product of Gaussian Mixture Models (20 + 10 P)

Consider the product of experts:

$$p(\boldsymbol{x}|\theta) = \frac{1}{\mathcal{Z}} \prod_{j=1}^{H} g_j(\boldsymbol{x}, \theta_j)$$

where each expert is a Gaussian mixture model in d-dimensions, and where each element of the mixture is Gaussian with identity covariance:

$$\forall_{j=1}^H: \ g_j(\boldsymbol{x}, \theta_j) = \sum_{k=1}^{C} \alpha_{jk} \frac{1}{(2\pi)^{d/2}} \exp\Big(-\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{\mu}_{jk}\|^2\Big).$$

(a) *Show* that $p(\boldsymbol{x}|\theta)$ can be rewritten as a mixture of C^H elements, where each mixture element (indexed by the vector $\boldsymbol{k} \in \{1, \dots, C\}^H$) has center

$$\boldsymbol{m}_{\boldsymbol{k}} = \frac{1}{H} \sum_{j=1}^{H} \boldsymbol{\mu}_{jk_j}.$$

(b) *Give* the centers $\boldsymbol{m}_{\boldsymbol{k}}$ of the mixture model equivalent to a product of two mixture models, where each mixture model in the product has 2 elements, where the first mixture has the two-dimensional centers $\boldsymbol{\mu}_{11} = \binom{2}{0}$ and $\boldsymbol{\mu}_{12} = \binom{4}{0}$, and where the second mixture has the two-dimensional centers $\boldsymbol{\mu}_{21} = \binom{0}{2}$ and $\boldsymbol{\mu}_{22} = \binom{0}{4}$.

Exercise 3: Programming (40 P)

Download the programming files on ISIS and follow the instructions.

Exercise 1: RBM with Ternary Hidden Units (20 + 10 P)

We consider a variant of the restricted Boltzmann machine with ternary hidden units $\boldsymbol{h} \in \{-1, 0, 1\}^H$. Input features remain binary, i.e. $\boldsymbol{x} \in \{0, 1\}^d$, like for the classical RBM. The probability model is given by:

$$p(\boldsymbol{x}, \boldsymbol{h}|\theta) = \frac{1}{\mathcal{Z}} \exp\Big(\sum_{j=1}^{H} \boldsymbol{w}_j^\top \boldsymbol{x} \cdot h_j + \sum_{j=1}^{H} h_j b_j\Big)$$

where $\theta = (\boldsymbol{w}_j, b_j)_{j=1}^H$ are the parameters of the model, and where and $\mathcal{Z}$ is the partition function that normalizes the probability distribution to 1.

(a) *Show* that this modified RBM can also be expressed as a product of experts

$$p(\boldsymbol{x}|\theta) = \frac{1}{\mathcal{Z}} \prod_{j=1}^{H} g_j(\boldsymbol{x}, \theta_j),$$

with

$$g_j(\boldsymbol{x}, \theta_j) = 1 + 2\cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j),$$

where cosh is the hyperbolic cosine function.

$$\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{\theta}) &= \sum_{\boldsymbol{h}\in\{-1,0,1\}^H} p(\boldsymbol{x}, \boldsymbol{h}|\boldsymbol{\theta}) \\
&= \sum_{\boldsymbol{h}\in\{-1,0,1\}^H} \frac{1}{\mathcal{Z}} \exp\Big(\sum_{j=1}^{H} \boldsymbol{w}_j^\top \boldsymbol{x}\cdot h_j + \sum_{j=1}^{H} h_j b_j\Big) \\
&= \sum_{\boldsymbol{h}\in\{-1,0,1\}^H} \frac{1}{\mathcal{Z}} \exp\Big(\sum_{j=1}^{H} (\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\cdot h_j\Big) \\
&= \frac{1}{\mathcal{Z}} \sum_{\boldsymbol{h}\in\{-1,0,1\}^H} \prod_{j=1}^{H} \exp\big((\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\cdot h_j\big) \\
&= \frac{1}{\mathcal{Z}} \sum_{h_1\in\{-1,0,1\}} \cdots \sum_{h_H\in\{-1,0,1\}} \prod_{j=1}^{H} \exp\big((\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\cdot h_j\big) \\
&= \frac{1}{\mathcal{Z}} \prod_{j=1}^{H} \sum_{h_j\in\{-1,0,1\}} \exp\big((\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\cdot h_j\big) \\
&= \frac{1}{\mathcal{Z}} \prod_{j=1}^{H} \big(1 + \exp(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j) + \exp(-(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j))\big) \\
&= \frac{1}{\mathcal{Z}} \prod_{j=1}^{H} \underbrace{\big(1 + 2\cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\big)}_{g_j(\boldsymbol{x},\boldsymbol{\theta}_j)}
\end{aligned}$$

$$\begin{aligned}
P(x|\theta) = \sum_{h\in\{-1,0,1\}^H} P(x,h|\theta) &= \sum_{h\in\{-1,0,1\}^H} \frac{1}{Z}\cdot \exp\Big(\sum_{j=1}^H w_j^T x \cdot h_j + \sum_{j=1}^H h_j b_j\Big) \\
&= \sum_{h\in\{-1,0,1\}^H} \frac{1}{Z} \exp\Big(\sum_{j=1}^H \big[(w_j^T x + b_j) h_j\big]\Big) \\
&= \sum_{h\in\{-1,0,1\}^H} \frac{1}{Z} \prod_{j=1}^H \exp\big((w_j^T x + b_j) h_j\big) \\
&= \frac{1}{Z} \prod_{j=1}^H \sum_{h=\{-1,0,1\}^H} \exp\big((w_j^T x + b_j) h_j\big) \\
&= \frac{1}{Z} \prod_{j=1}^H \big(1 + \exp(-(w_j^T x + b_j)) + \exp(w_j^T x + b_j)\big) \\
&= \frac{1}{Z} \prod_{j=1}^H \big(1 + 2\cosh(w_j^T x + b_j)\big) \\
&= \frac{1}{Z} \prod_{j=1}^H g_j(x, \theta_j)
\end{aligned}$$

$$Z = \sum_{x=\{0,1\}^d} \prod_{j=1}^H g_j(x, \theta_j)$$

(b) *Show* that the gradient of the log-likelihood assigned to some data point $\boldsymbol{x}_n$ by the modified RBM has the form

$$\forall_{j=1}^H: \quad \frac{\partial \log p(\boldsymbol{x}_n|\theta)}{\partial \boldsymbol{w}_j} = \boldsymbol{x}_n \cdot \sigma(\boldsymbol{w}_j^\top \boldsymbol{x}_n + b_j) - \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\theta)}\big[\boldsymbol{x}\cdot\sigma(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\big]$$

$$\forall_{j=1}^H: \quad \frac{\partial \log p(\boldsymbol{x}_n|\theta)}{\partial b_j} = \sigma(\boldsymbol{w}_j^\top \boldsymbol{x}_n + b_j) - \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\theta)}\big[\sigma(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\big]$$

where $\sigma(t) = \dfrac{\sinh(t)}{0.5 + \cosh(t)}$.

$$P(x_n|\theta) = \frac{1}{Z} \prod_{j=1}^H \big(1 + 2\cosh(w_j^T x_n + b_j)\big)$$

$$\begin{aligned}
\log P(x_n|\theta) &= \log\Big(\frac{1}{Z} \prod_{j=1}^H \big(1 + 2\cosh(w_j^T x_n + b_j)\big)\Big) \\
&= \sum_{j=1}^H \log\big(1 + 2\cosh(w_j^T x_n + b_j)\big) - \log Z = \sum_{j=1}^H \log(g_j(x,\theta_j)) - \log Z
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \log Z}{\partial w_j} = \frac{1}{Z} \cdot \frac{\partial Z}{\partial w_j} &= \frac{1}{Z} \cdot \frac{\partial \sum_{x\in\{0,1\}^d} \prod_{i=1}^H g_i(x,\theta_i)}{\partial w_j} \\
&= \frac{1}{Z} \cdot \sum_{x\in\{0,1\}^d} \frac{\partial \prod_{i=1}^H g_i(x,\theta_i)}{\partial w_j} \\
&= \frac{1}{Z} \cdot \sum_{x\in\{0,1\}^d} \underbrace{\Big(\prod_{i\neq j} g_i(x,\theta_i)\Big)}_{\text{const}} \cdot \frac{\partial g_j(x,\theta_j)}{\partial w_j} \\
&= \frac{1}{Z} \cdot \sum_{x\in\{0,1\}^d} \Big(\prod_{i\neq j} g_i(x,\theta_i)\Big) \cdot 2\sinh(w_j^T x + b_j) \cdot x
\end{aligned}$$

$$= \sum_{x_n \in \{0,1\}^d} \frac{1}{Z} \cdot \prod_{i=1}^{H} g_i(x, \theta_i) \cdot \frac{2 \sinh(w_j^T x + b_j) \cdot x}{g_j(x, \theta_j)}$$

$$= \frac{2}{2} \sum_{x_n \in \{0,1\}^d} P(X|\theta) \cdot \frac{2 \sinh(w_j^T x + b_j) \cdot x}{1 + 2\cosh(w_j^T x + b_j)}$$

$$= \mathbb{E}_{x \sim P(x|\theta)} \cdot \frac{\sinh(w_j^T x + b_j)}{0.5 + \cosh(w_j^T x + b_j)} \, x$$

$$= \mathbb{E}_{x \sim P(x|\theta)} \cdot \left[\delta(w_j^T x + b_j) \cdot x \right]$$

$$\frac{\partial \log P(x_n|\theta)}{\partial w_j} = \frac{2 \sinh(w_j^T x_n + b_j) \cdot x_n}{1 + 2\cosh(w_j^T x_n + b_j)} - \frac{\partial \log Z}{\partial w_j}$$

$$= x_n \delta(w_j^T x_n + b_j) - \mathbb{E}_{x \sim P(x|\theta)} \cdot \left[\delta(w_j^T x + b_j) \cdot x \right]$$

$$\frac{\partial \log Z}{\partial b_j} = \frac{1}{Z} \cdot \frac{\partial Z}{\partial b_j} = \frac{1}{Z} \cdot \frac{\partial \sum_{x \in \{0,1\}^d} \prod_{i=1}^{H} g_i(x, \theta_j)}{\partial b_j}$$

$$= \frac{1}{Z} \cdot \sum_{x \in \{0,1\}^d} \frac{\partial \prod_{i=1}^{H} g_i(x, \theta_j)}{\partial b_j}$$

$$= \frac{1}{Z} \cdot \sum_{x_n \in \{0,1\}^d} \underbrace{\left(\prod_{i \neq j} g_i(x, \theta_i) \right)}_{\text{const}} \cdot \frac{\partial g_j(x, \theta_j)}{\partial w_j}$$

$$= \frac{1}{Z} \cdot \sum_{x \in \{0,1\}^d} \left(\prod_{i \neq j} g_i(x, \theta_i) \right) \cdot \sinh(w_j^T x + b_j)$$

$$= \sum_{x_n \in \{0,1\}^d} \frac{1}{Z} \cdot \prod_{i=1}^{H} g_i(x, \theta_i) \cdot \frac{\sinh(w_j^T x + b_j)}{g_j(x, \theta_j)}$$

$$= \frac{2}{2} \sum_{x_n \in \{0,1\}^d} P(X|\theta) \cdot \frac{\sinh(w_j^T x + b_j)}{1 + 2\cosh(w_j^T x + b_j)}$$

$$= \mathbb{E}_{x \sim P(x|\theta)} \cdot \frac{\sinh(w_j^T x + b_j)}{0.5 + \cosh(w_j^T x + b_j)}$$

$$= \mathbb{E}_{x \sim P(x|\theta)} \cdot \left[\delta(w_j^T x + b_j) \right]$$

$$\frac{\partial \log P(x_n|\theta)}{\partial b_j} = \frac{2 \sinh(w_j^T x_n + b_j) \cdot}{1 + 2\cosh(w_j^T x_n + b_j)} - \frac{\partial \log Z}{\partial b_j}$$

$$= \delta(w_j^T x_n + b_j) - \mathbb{E}_{x \sim P(x|\theta)} \cdot \left[\delta(w_j^T x + b_j) \right]$$

(b) We build upon the the PoE formulation from exercise 1(a), which we rewrite as follows:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}} \prod_{j=1}^{H} \left(1 + 2\cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)\right) = \frac{1}{\mathcal{Z}} \exp\Big(\underbrace{\log \prod_{j=1}^{H} (1 + 2\cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j))}_{f(\boldsymbol{x},\boldsymbol{\theta}):=} \Big)$$

$$= \frac{1}{\mathcal{Z}} \exp\left(f(\boldsymbol{x},\boldsymbol{\theta})\right) = \frac{\exp\left(f(\boldsymbol{x},\boldsymbol{\theta})\right)}{\sum_{\boldsymbol{x}\in\{0,1\}^d} \exp\left(f(\boldsymbol{x},\boldsymbol{\theta})\right)},$$

where in the last step we expanded the normalization constant $\mathcal{Z}$. Therefore, the log-likelihood has the following form:

$$\log p(\boldsymbol{x}_n|\boldsymbol{\theta}) = f(\boldsymbol{x}_n,\boldsymbol{\theta}) - \log \sum_{\boldsymbol{x}\in\{0,1\}^d} \exp\left(f(\boldsymbol{x},\boldsymbol{\theta})\right)$$

We use this form to compute the gradient as follows:

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{x}_n|\boldsymbol{\theta}) &= \nabla_{\boldsymbol{\theta}} f(\boldsymbol{x}_n,\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}} \log \sum_{\boldsymbol{x}\in\{0,1\}^d} \exp\left(f(\boldsymbol{x},\boldsymbol{\theta})\right) \\
&= \nabla_{\boldsymbol{\theta}} f(\boldsymbol{x}_n,\boldsymbol{\theta}) - \frac{1}{\sum_{\boldsymbol{x}\in\{0,1\}^d} \exp\left(f(\boldsymbol{x},\boldsymbol{\theta})\right)} \cdot \sum_{\boldsymbol{x}\in\{0,1\}^d} \exp\left(f(\boldsymbol{x},\boldsymbol{\theta})\right) \cdot \nabla_{\boldsymbol{\theta}} f(\boldsymbol{x},\boldsymbol{\theta}) \\
&= \nabla_{\boldsymbol{\theta}} f(\boldsymbol{x}_n,\boldsymbol{\theta}) - \sum_{\boldsymbol{x}\in\{0,1\}^d} \underbrace{\frac{\exp\left(f(\boldsymbol{x},\boldsymbol{\theta})\right)}{\sum_{\boldsymbol{y}\in\{0,1\}^d} \exp\left(f(\boldsymbol{y},\boldsymbol{\theta})\right)}}_{p(\boldsymbol{x}|\boldsymbol{\theta})} \cdot \nabla_{\boldsymbol{\theta}} f(\boldsymbol{x},\boldsymbol{\theta}) \\
&= \nabla_{\boldsymbol{\theta}} f(\boldsymbol{x}_n,\boldsymbol{\theta}) - \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\boldsymbol{\theta})}[\nabla_{\boldsymbol{\theta}} f(\boldsymbol{x},\boldsymbol{\theta})]
\end{aligned}$$

That is,

$$\boxed{\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{x}_n|\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} f(\boldsymbol{x}_n,\boldsymbol{\theta}) - \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\boldsymbol{\theta})}[\nabla_{\boldsymbol{\theta}} f(\boldsymbol{x},\boldsymbol{\theta})]} \tag{1}$$

Remember our definition of $f(\boldsymbol{x},\boldsymbol{\theta})$:

$$f(\boldsymbol{x},\boldsymbol{\theta}) = \log \prod_{j=1}^{H} (1 + 2\cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)) = \sum_{j=1}^{H} \log(1 + 2\cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)) \tag{2}$$

Therefore, the corresponding gradients are given as:

$$\nabla_{\boldsymbol{w}_j} f(\boldsymbol{x},\boldsymbol{\theta}) = \underbrace{\frac{2\sinh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)}{1 + 2\cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)}}_{\sigma(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)} \boldsymbol{x} = \sigma(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j) \cdot \boldsymbol{x} \tag{3}$$

$$\frac{\partial f(\boldsymbol{x},\boldsymbol{\theta})}{\partial b_j} = \underbrace{\frac{2\sinh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)}{1 + 2\cosh(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)}}_{\sigma(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)} = \sigma(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j) \tag{4}$$

Inserting (3) and (4) into (1) gives our target form.

Exercise 2: Product of Gaussian Mixture Models ($20 + 10$ P)

Consider the product of experts:

$$p(\boldsymbol{x}|\theta) = \frac{1}{\mathcal{Z}} \prod_{j=1}^{H} g_j(\boldsymbol{x}, \theta_j)$$

where each expert is a Gaussian mixture model in d-dimensions, and where each element of the mixture is Gaussian with identity covariance:

$$\forall_{j=1}^{H} : \; g_j(\boldsymbol{x}, \theta_j) = \sum_{k=1}^{C} \alpha_{jk} \frac{1}{(2\pi)^{d/2}} \exp\Big(-\frac{1}{2} \|\boldsymbol{x} - \boldsymbol{\mu}_{jk}\|^2 \Big).$$

(a) *Show* that $p(\boldsymbol{x}|\theta)$ can be rewritten as a mixture of C^H elements, where each mixture element (indexed by the vector $\boldsymbol{k} \in \{1, \ldots, C\}^H$) has center

$$\boldsymbol{m}_{\boldsymbol{k}} = \frac{1}{H} \sum_{j=1}^{H} \boldsymbol{\mu}_{jk_j}.$$

$$P(x|\theta) = \frac{1}{Z} \prod_{j=1}^{H} \sum_{k=1}^{C} \alpha_{jk} \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\|x-\mu_{jk}\|^2\right)$$

$$= \sum_{k_1=1}^{C} \cdots \sum_{k_H=1}^{C} \left(\frac{1}{Z}\prod_{i=1}^{H} \alpha_{ik_i}\right) \cdot \left(\prod_{i=1}^{H} \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\|x-\mu_{ik_i}\|^2\right)\right)$$

$$= \sum_{k_1=1}^{C} \cdots \sum_{k_H=1}^{C} \left(\frac{1}{Z}\prod_{i=1}^{H} \alpha_{ik_i}\right) \cdot \left(\frac{1}{(2\pi)^{\frac{dH}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^{H}\|x-\mu_{ik_i}\|^2\right)\right)$$

$$\|x-m_k\|^2 = \|x\|^2 - 2x^T m_k + \|m_k\|^2$$

$$H\cdot\|x-m_k\|^2 = H\cdot\|x\|^2 - 2Hx^T m_k + H\cdot\|m_k\|^2$$

$$= H\cdot\|x\|^2 - 2x^T(Hm_k) + H\cdot\|m_k\|^2$$

$$= H\cdot\|x\|^2 - 2x^T\cdot\left(\sum_{j=1}^{H}\mu_{jk_j}\right) + H\cdot\|m_k\|^2$$

$$\Rightarrow H\cdot\|x\|^2 - 2x^T\cdot\left(\sum_{j=1}^{H}\mu_{jk_j}\right) = \underline{H\cdot\|x-m_k\|^2 - H\|m_k\|^2}$$

$$\sum_{i=1}^{H}\|x-\mu_{ik_i}\|^2 = \sum_{i=1}^{H}\left[\|x\|^2 - 2x^T\mu_{ik_i} + \|\mu_{ik_i}\|^2\right]$$

$$= H\cdot\|x\|^2 - 2x^T\sum_{i=1}^{H}\mu_{ik_i} + \sum_{i=1}^{H}\|\mu_{ik_i}\|^2$$

$$= H\cdot\|x-m_k\|^2 - H\cdot\|m_k\|^2 + \sum_{i=1}^{H}\|\mu_{ik_i}\|^2$$

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{H}\|x-\mu_{ik_i}\|^2\right) = \exp\left(-\frac{H}{2}\|x-m_k\|^2\right)\cdot\exp\left(\frac{H}{2}\|m_k\|^2 - \frac{1}{2}\sum_{i=1}^{H}\|\mu_{ik_i}\|^2\right)$$

$$= \sum_{k_1=1}^{C} \cdots \sum_{k_H=1}^{C} \underbrace{\left(\frac{1}{Z}\prod_{i=1}^{H}\alpha_{ik_i}\right)\cdot\exp\left(\frac{H}{2}\|m_k\|^2 - \frac{1}{2}\sum_{i=1}^{H}\|\mu_{ik_i}\|^2\right)}_{\text{new mixing coef.}} \underbrace{\frac{1}{(2\pi)^{\frac{dH}{2}}}\exp\left(-\frac{H}{2}\|x-m_k\|^2\right)}_{\text{Gaussian with mean } m_k,\ \Sigma = \frac{1}{H} I}$$

(b) *Give* the centers $\boldsymbol{m_k}$ of the mixture model equivalent to a product of two mixture models, where each mixture model in the product has 2 elements, where the first mixture has the two-dimensional centers $\boldsymbol{\mu}_{11} = \binom{2}{0}$ and $\boldsymbol{\mu}_{12} = \binom{4}{0}$, and where the second mixture has the two-dimensional centers $\boldsymbol{\mu}_{21} = \binom{0}{2}$ and $\boldsymbol{\mu}_{22} = \binom{0}{4}$.

H×C

$$\boldsymbol{m_k} = \frac{1}{H} \sum_{j=1}^{H} \boldsymbol{\mu}_{jk_j}.$$

index k is a vector !

$H=2$

$k = \binom{1}{1}$ $\quad m_k = \frac{1}{2} \cdot (\mu_{1k_1} + \mu_{2k_2}) = \frac{1}{2}(\mu_{11} + \mu_{21}) = \frac{1}{2}\left[\binom{2}{0} + \binom{0}{2}\right] = \binom{1}{1}$

$k = \binom{1}{2}$ $\quad m_k = \frac{1}{2}(\mu_{11} + \mu_{22}) = \frac{1}{2}\left[\binom{2}{0} + \binom{0}{4}\right] = \binom{1}{2}$

$k = \binom{2}{1}$ $\quad m_k = \frac{1}{2}(\mu_{12} + \mu_{21}) = \frac{1}{2}\left[\binom{4}{0} + \binom{0}{2}\right] = \binom{2}{1}$

$k = \binom{2}{2}$ $\quad m_k = \frac{1}{2}(\mu_{12} + \mu_{22}) = \binom{2}{2}$