$f(0) = a,\ f(-x) =$

4.2 確率的生成モデル

$p(x \mid C_k)$: クラスの条件つき確率密度

$p(C_k)$: クラスの事前確率

↓ ベイズ定理

$p(C_k \mid x)$: 事後確率

事後確率

$$p(C_1 \mid x) = \frac{p(x \mid C_1)p(C_1)}{\sum_{k=1}^{2} p(x \mid C_k)p(C_k)}$$

$$= \frac{p(x \mid C_1)p(C_1)}{p(x \mid C_1)p(C_1) + p(x \mid C_2)p(C_2)}$$

$$= \frac{1}{1 + \frac{p(x \mid C_2)p(C_2)}{p(x \mid C_1)p(C_1)}}$$

$$= \frac{1}{1 + \left(\frac{p(x \mid C_1)p(C_1)}{p(x \mid C_2)p(C_2)}\right)^{-1}}$$

$$= \frac{1}{1 + \cancel{\log}\ \cancel{\exp} \ \exp \log \left(\left(\frac{p(x \mid C_1)p(C_1)}{p(x \mid C_2)p(C_2)}\right)^{-1}\right)}$$

$$= \frac{1}{1 + e^{-a}}, \quad a = \cancel{\log}$$

$= \sigma(a)$: sigmoid 関数

$\log \exp x = x$

$\exp(\log x) = x$

$\log x = \log y$

$\rightarrow x = y$

$\log(e^{-x}) = -x$

$\log\left(\frac{1}{e^{x}}\right) =$

NO.

DATE

K クラスの場合

$$p(C_k|x) = \frac{p(x|C_k)p(C_k)}{\sum_{j=1}^{K} p(x|C_j)p(C_j)}$$

$$= \frac{\exp(\log(p(x|C_k)p(C_k)))}{\sum_{j=1}^{K} \exp(\log(p(x|C_j)p(C_j)))}$$

$$= \frac{\exp(a_k)}{\sum_{j=1}^{K} \exp(a_j)} \quad \text{where } a_k \equiv \log p(x|C_k)p(C_k)$$

(4.62)

正規化指数関数 or softmax 関数

$$a_k \gg a_j \text{ for } \forall j (\neq k) \Longrightarrow \exp(a_k) \gg \exp(a_j)$$

$$\sum_{j=1}^{K} \exp(a_j) = \sum_{j \neq k} \exp(a_j) + \exp(a_k) \simeq \exp(a_k)$$

より、

$$p(C_k|x) \simeq \frac{\exp(a_k)}{\exp(a_k)} = 1$$

極限の振るまいがシグモイド関数っぽい

$$p(C_j|x) \simeq \frac{\exp(a_j)}{\exp(a_k)} \simeq 0$$

$$\Sigma\Sigma^{-1} = I \qquad \Sigma^T(\Sigma^{-1})^T = I$$

$$(\Sigma\Sigma^{-1})^T = I^T = I \qquad \Sigma^{-1} = (\Sigma^{-1})^T$$

4.2.1

仮定: 1. $p(x|C_k)$ … Gauss分布

2. すべてのクラスが同じ共分散行列を共有する

$$\Sigma_k = \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{21} & \cdots \\ \sigma_{12} & \ddots & \\ \vdots & & \sigma_{KK} \end{pmatrix}$$

$$p(x|C_k) = \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)\right).$$

2クラスの場合

$$\langle x|\Sigma^{-1}|\mu_2\rangle = (\Sigma^{-1}|x\rangle)^\dagger|\mu_2\rangle$$

$$p(C_1|x) = \sigma\left(\log\frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)}\right)$$

$$= \cancel{\sigma(\log \exp}$$

$$= \sigma\left(\log\left(\frac{\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)p(C_1)}{\exp\left(-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\right)p(C_2)}\right)\right)$$

仮定2より キャンセル

$$= \sigma\left(\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) - \frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \log\left(\frac{p(C_1)}{p(C_2)}\right)\right)$$

$$= \sigma\left(\cancel{\frac{1}{2}x^T\Sigma^{-1}x} + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}x^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_2^T\Sigma^{-1}x\right.$$
$$\left.\cancel{-\frac{1}{2}x^T\Sigma^{-1}x} - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \frac{1}{2}x^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_1^T\Sigma^{-1}x + \log\left(\frac{p(C_1)}{p(C_2)}\right)\right)$$

$$\left(\frac{1}{2}x^T\Sigma^{-1}\mu_2\right)^T = \mu_2^T(\Sigma^{-1})^Tx = \mu_2^T\Sigma^{-1}x$$

$$= -\frac{1}{2}\mu_2^T\Sigma^{-1}x - \frac{1}{2}\mu_2^T\Sigma^{-1}x + \frac{1}{2}\mu_1^T\Sigma^{-1}x + \frac{1}{2}\mu_1^T\Sigma^{-1}x$$

$$= (\mu_1-\mu_2)^T\Sigma^{-1}x = (\Sigma^{-1}(\mu_1-\mu_2))^Tx$$

$$\sigma(w^Tx + w_0).$$

NO.

DATE

Kクラスの場合

$$\cancel{a_k = \log\left(p(x|C_k)p(C_k)\right)}$$

$$\cancel{= \log\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)\right)\right)} \quad p(C_k)$$

$$\cancel{= \log\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\right)} - \frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k) + \log(p(C_k))$$

$$= \log\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\right) - \frac{1}{2}x^T\Sigma^{-1}x - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k$$

$$+ \underbrace{\frac{1}{2}x^T\Sigma^{-1}\mu_k + \frac{1}{2}\mu_k^T\Sigma^{-1}x^T}_{\shortparallel} + \log(p(C_k))$$

$$\mu_k^T\Sigma^{-1}x = (\Sigma^{-1}\mu_k)^T x$$

$$= \log\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\right) - \frac{1}{2}x^T\Sigma^{-1}x + w_k^T x + w_0$$

$$p(C_k|x) = \frac{\cancel{\exp\left(\log\left(\frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{|\Sigma|}}\right) - \frac{1}{2}x^T\Sigma^{-1}x\right)}\exp(w_k^T x + w_{k0})}{\sum_j \cancel{\exp\left(\log\left(\frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{|\Sigma|}}\right) - \frac{1}{2}x^T\Sigma^{-1}x\right)}\exp(w_j^T x + w_{j0})}$$

$$= \frac{\exp(w_k^T x + w_{k0})}{\sum_j \exp(w_j^T x + w_{j0})}$$

$$= \frac{\exp(a_k(x))}{\sum_j \exp(a_j(x))}$$

$$p(C_k|x) = \frac{p(x|C_k)p(C_k)}{}$$

4.2.2.

$$\left(\begin{array}{c} p(C_1)p(x_n|C_1) \propto p(C_1|x_n) \\ p(C_1|x_n) = \dfrac{p(C_1)p(x_n|C_1)}{\sum_j p(C_j)p(x_n|C_j)} \end{array} \right)$$

事前確率 $p(C_1) = \alpha,\ p(C_2) = 1-\alpha$.

$$p(x_n, C_1) = p(C_1)p(x_n|C_1) = \alpha p(x_n|C_1) = \alpha \mathcal{N}(x_n|\mu_1, \Sigma)$$

x_n かつ C_1
「N個のデータから x_n を選び、かつそれがクラス C_1 に分類される」確率
「$t_n = 1$ となるので」つまりいる?

$$p(x_n, C_2) = p(C_2)p(x_n|C_2) = (1-\alpha)\mathcal{N}(x_n|\mu_2, \Sigma).$$

仮定 1. $p(x_n|C_k)$ が Gaussian
2. Σ は共通

尤度関数

$p(\mathbf{t}, X|\pi, \mu_1, \mu_2, \Sigma)$　$p(C_1|x_n)$　$p(C_2|x_n)$

$$= \prod_{n=1}^{N} (\alpha \mathcal{N}(x_n|\mu_1, \Sigma))^{t_n} ((1-\alpha)\mathcal{N}(x_n|\mu_2, \Sigma))^{1-t_n}.$$

正解のラベルと同じラベルに分類される確率が残る.

尤度と事前確率を
パラメトリックな関数形で表現
パラメータの最尤量を求める.

対数尤度

$$\log p(\mathbf{t}, \mathbf{X} \mid \alpha, \mu_1, \mu_2, \Sigma)$$

$$= \log\left(\prod_{n=1}^{N} \left(\alpha \mathcal{N}(\mathbf{x}_n \mid \mu_1, \Sigma)\right)^{t_n} \left((1-\alpha)\mathcal{N}(\mathbf{x}_n \mid \mu_2, \Sigma)\right)^{1-t_n}\right)$$

$$= \sum_{n=1}^{N}\left(t_n \log\left(\alpha \mathcal{N}(\mathbf{x}_n \mid \mu_1, \Sigma)\right) + (1-t_n)\log\left((1-\alpha)\mathcal{N}(\mathbf{x}_n \mid \mu_2, \Sigma)\right)\right)$$

$$= \sum_{n=1}^{N}\Big(\underbrace{t_n \log\alpha + (1-t_n)\log(1-\alpha)}_{\alpha} + \underbrace{t_n \log \mathcal{N}(\mathbf{x}_n \mid \mu_1, \Sigma)}_{\mu_1,\ \Sigma} + \underbrace{(1-t_n)\log \mathcal{N}(\mathbf{x}_n \mid \mu_2, \Sigma)}_{\mu_2,\ \Sigma}\Big)$$

↓ α の極値

$$\frac{\partial}{\partial\alpha}\log p(\mathbf{t}, \mathbf{X} \mid \alpha, \mu_1, \mu_2, \Sigma) \qquad t_n \cancel{= \alpha t_n - \alpha + \alpha t_n}$$

$$= \sum_{n=1}^{N}\left(\frac{t_n}{\alpha} - \frac{1-t_n}{1-\alpha}\right) = \sum_{n=1}^{N}\frac{(1-\alpha)t_n - (1-t_n)\alpha}{\alpha(1-\alpha)}$$

$$= \frac{1}{\alpha(1-\alpha)}\sum_{n=1}^{N}(t_n - \alpha) = \frac{1}{\alpha(1-\alpha)}\left(\sum_{n=1}^{N} t_n - N\alpha\right)$$

$$\frac{\partial}{\partial\alpha}\log p = 0 \Longrightarrow \alpha = \frac{1}{N}\sum_{n=1}^{N} t_n = \frac{N_1}{N} = \frac{N_1}{N_1 + N_2}.$$

α の最尤値 … 全データ数 N に対する. クラス C_1 のデータ数 N_1 の割合.

NO.

DATE

μ_1の最尤量

$$\mathcal{N}(x_n|\mu_1,\Sigma) = \underbrace{\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{\sqrt{|\Sigma|}}}_{=N_0}\exp\left(-\frac{1}{2}(x_n-\mu_1)^T\Sigma^{-1}(x_n-\mu_1)\right)$$

$$\sum_{n=1}^{N} t_n \log \mathcal{N}(x_n|\mu_1,\Sigma)$$

$$= \sum_{n=1}^{N} t_n\left(-\frac{1}{2}(x_n-\mu_1)^T\Sigma^{-1}(x_n-\mu_1)\right) + \mathrm{Const}$$

$$= -\frac{1}{2}\sum_{n=1}^{N} t_n (x_n-\mu_1)^T\Sigma^{-1}(x_n-\mu_1) + \mathrm{Const}$$

$$\downarrow \frac{\partial}{\partial \mu_1} = 0$$

$$-\frac{1}{2}\sum_{n=1}^{N} t_n\left(-1\,\Sigma^{-1}(x_n-\mu_1) - (x_n-\mu_1)^T\Sigma^{-1}\right)$$

$$-2\,(x_n-\mu_1)^T\Sigma^{-1}$$

$$=$$

$$= \sum_{n=1}^{N} t_n (x_n-\mu_1)^T\Sigma^{-1} = 0$$

$$\times\Sigma$$

$$\sum_{n=1}^{N} t_n x_n - \underbrace{\left(\sum_{n=1}^{N} t_n\right)}_{=N_1}\mu_1 = 0$$

$$\therefore \mu_1 = \frac{1}{N_1}\sum_{n=1}^{N} t_n x_n \cdots$$ C_1に割り当てられる

すべての入力ベクトルの和 $\longrightarrow \frac{\partial x}{\partial x} = 1$.

$$\downarrow t_n \to (1-t_n)$$

$$\mu_2 = \frac{1}{N_2}\sum_{n=1}^{N}(1-t_n)x_n \cdots$$ C_2に…

$$\left(\begin{pmatrix}B_1\\B_2\end{pmatrix}\right)\begin{pmatrix}a & b\\ c & d\end{pmatrix} = (a+c \quad b+d$$

$$\frac{\partial x_1}{\partial x_1} = 1$$

$$\frac{\partial x}{\partial x} = e.$$

$$x + \Delta - x =$$

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}e\\ f\end{pmatrix} = \begin{pmatrix}ae+bf\\ ce+df\end{pmatrix}$$

$$(a \;\; b)\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}x_1\\ x_2\end{pmatrix}$$

$$f(x+\Delta) = x+\Delta$$

$$f(x) = a^T x + b$$

$$f(x+\Delta) = a^T x + a^T\Delta + \cancel{b}$$

$$-)\; f(x) = a^T x \quad + \cancel{b}$$

$$a^T\Delta$$

$$\frac{\partial}{\partial x}(a\cdot x) = a.$$

$$f(x+\Delta) = x+\Delta$$

$$= f(x) + 1\cdot\Delta$$

KOKUYO

$\langle a|\hat{A}|b\rangle = \hat{A}|b\rangle\langle a|$

<u>Σの最大値</u>

$$\sum_{n=1}^{N}\left(t_n \log \mathcal{N}(x_n|\mu_1,\Sigma) + (1-t_n)\log \mathcal{N}(x_n|\mu_2,\Sigma)\right)$$

$$= \sum_{n=1}^{N}\left[t_n\left(-\frac{1}{2}(x_n-\mu_1)^T\Sigma^{-1}(x_n-\mu_1) + \log\left(\frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{|\Sigma|}}\right)\right)\right.$$

$$\left.+(1-t_n)\left(-\frac{1}{2}(x_n-\mu_2)^T\Sigma^{-1}(x_n-\mu_2) + \log\left(\frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{|\Sigma|}}\right)\right)\right]$$

$$= -\frac{1}{2}\underbrace{\sum_{n=1}^{N} t_n}_{=N_1}\log|\Sigma| - \frac{1}{2}\underbrace{\sum_{n=1}^{N} t_n (x_n-\mu_1)^T\Sigma^{-1}(x_n-\mu_1)}_{\mathrm{Tr}(\Sigma^{-1}S_1)} \qquad \underbrace{-\frac{D}{2}N\log(2\pi)}_{=\,+\,\mathrm{const}},$$

$$-\frac{1}{2}\underbrace{\sum_{n=1}^{N}(1-t_n)}_{=N_2}\log|\Sigma| - \frac{1}{2}\underbrace{\sum_{n=1}^{N}(1-t_n)(x_n-\mu_2)^T\Sigma^{-1}(x_n-\mu_2)}_{\mathrm{Tr}(\Sigma^{-1}S_2)} \cancel{+\,\mathrm{const}}.$$

$$= -\frac{N}{2}\log|\Sigma| - \frac{1}{2}\underbrace{\sum_{n\in C_1}(x_n-\mu_1)^T\Sigma^{-1}(x_n-\mu_1)}_{\mathrm{Tr}} \qquad S$$

$A|a\rangle \cancel{\langle b|}$

$A|a\rangle\langle b|$

$$-\frac{1}{2}\sum_{n\in C_2}(x_n-\mu_2)^T\Sigma^{-1}(x_n-\mu_2)$$

$$\left(\Sigma^{-1}(x_n-\mu_2)\right)^{TT}$$

$$= -\frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{k=1}^{2}\sum_{n\in C_k}\Sigma^{-1}(x_n-\mu_k)(x_n-\mu_1)^T$$

$(x_n-\mu_1$

$$= -\frac{N}{2}\log|\Sigma| - \frac{N}{2}\underbrace{\Sigma^{-1}\underbrace{\sum_{k=1}^{2}\frac{N_k}{N}\underbrace{\frac{1}{N_k}\sum_{n\in C_k}(x_n-\mu_k)(x_n-\mu_k)^T}_{=S_k}}_{}}_{=S}$$

$$\cancel{\mathrm{Tr}(\Sigma^{-1}S)}$$

$$\mathrm{Tr}(\Sigma^{-1}(S_1'+S_2'))$$

$$= -\frac{N}{2}\log|\Sigma| - \frac{1}{2}\underbrace{\left(\mathrm{Tr}(\Sigma^{-1}S_1') + \mathrm{Tr}(\Sigma^{-1}S_2')\right)}_{} \qquad = \mathrm{Tr}\,\Sigma^{-1}\left(N\underbrace{\left(\frac{N_1}{N}\underbrace{\frac{1}{N_1}S_1'}_{=S_1} + \frac{N_2}{N}\underbrace{\frac{1}{N_2}S_2'}_{=S_1}\right)}_{=S}\right)$$

$$= -\frac{N}{2}\log|\Sigma| - \frac{N}{2}\mathrm{Tr}(\Sigma^{-1}S)$$

$f(X) = \log|X| \qquad f(X+\Delta) = \log|X+\Delta|$

$$|X+\Delta| = |X(1+X^{-1}\Delta)|$$
$$= |X||1+X^{-1}\Delta|$$

$$= \log|X+\Delta| - \log|X| + \log|X|$$
$$= \log\frac{|X+\Delta|}{|X|} + \log|X|$$
$$= \log|1+X^{-1}\Delta| + \log|X|$$

$$\frac{\partial}{\partial \Sigma^{-1}} \log p(t, X \mid \alpha, \mu_1, \mu_2, \Sigma)$$

$$\frac{\partial}{\partial X}\log|X| = (X^{-1})^T$$

$$= -\frac{N}{2}\frac{\partial}{\partial \Sigma^{-1}}\left(-\log|\Sigma^{-1}| + \mathrm{Tr}(\Sigma^{-1}S)\right)$$

$$\frac{\partial}{\partial X}\mathrm{Tr}(XA) = A^T$$

$$= -\frac{N}{2}(-\Sigma^T + S^T) = 0$$

$$\log|\Sigma| = \log\frac{1}{|\Sigma^{-1}|}$$
$$= -\log|\Sigma^{-1}|$$

$$\rightarrow \cancel{S}\ \Sigma = S$$

各クラスの共分散行列の重み付き平均

NO.

DATE

4.2.3.

入力 : $x_i \in \{0,1\}^D$ ~~$i=1,\dots,D$~~

x_1	x_2	x_3	$\cdots$	x_D
特徴1	2	3	$\cdots$	D
0 or 1	0 or 1	$\cdots$		0 or 1
1	0	0		0
0	1	0	$\cdots$	
$\vdots$				

一般的な分布 … 2^D-1 個の独立な変数を含む

2^D 個の要素

$\downarrow$

$\begin{cases} x_i \text{ は互いに独立} \\ \mu_{ki} \text{ で定まる(ひとつ)ベルヌーイ分布に従う} \end{cases}$ と仮定

$$p(x|C_k) = \prod_{i=1}^{D} \mu_{ki}^{x_i}(1-\mu_{ki})^{1-x_i}$$

$\downarrow$ $a_k = \log(p(x|C_k)p(C_k))$ (4.63)

$$a_k(x) = \log\left(\prod_{i=1}^{D} \mu_{ki}^{x_i}(1-\mu_{ki})^{1-x_i}\, p(C_k)\right)$$

$$= \underbrace{\sum_{i=1}^{D}\left(x_i \log \mu_{ki} + (1-x_i)\log(1-\mu_{ki})\right)}_{\text{入力}x_i\text{の線形関数}} + \log p(C_k)$$

$\longrightarrow$ 決定境界が超平面

4.2.4. 指数型分布族

入力が… $\begin{cases} \cdot \text{Gaussian (連続)} & p(x|C_k) \propto N(x|\mu, \Sigma) \\ \cdot \text{離散値} & x_i \in \{0,1\}^D \end{cases}$

のとき、~~事後確率~~ $p(C_k|x)$ が… $\begin{cases} \text{シグモイド ~~関数~~ (2クラス)} \\ \text{ソフトマックス (K>2クラス)} \end{cases}$

を活性化関数とする一般化線形モデルで与えられる。

↓ 尤度 $p(x|C_k)$ をGaussianから指数型分布族に拡張する。

$$\cancel{p(x|C_k)}$$
$$p(x|\lambda_k) = h(x) g(\lambda_k) \exp\left(\lambda_k^T u(x)\right)$$

特に $u(x) = x$ となる分布について

$x \to x/s$ ~~変数変換~~ (2.236) 尤度にパラメータ s の導入

$$p(x|\lambda_k, s) = \frac{1}{s} h\left(\frac{1}{s}x\right) g(\lambda_k) \exp\left(\frac{1}{s}\lambda_k^T x\right)$$

$$p(x \mid \lambda_k, s) = \frac{1}{s} h\left(\frac{1}{s}x\right) g(\lambda_k) \exp\left(\frac{1}{s}\lambda_k^T x\right)$$

<u>2クラスの場合</u>

$$a = \log\left(\frac{p(x \mid \lambda_{1,s})\, p(C_1)}{p(x \mid \lambda_{2,s})\, p(C_2)}\right) \qquad (4.58)$$

$$= \log\left(\frac{\cancel{\frac{1}{s} h\left(\frac{1}{s}x\right)} g(\lambda_1) \exp\left(\frac{1}{s}\lambda_1^T x\right) p(C_1)}{\cancel{\frac{1}{s} h\left(\frac{1}{s}x\right)} g(\lambda_2) \exp\left(\frac{1}{s}\lambda_2^T x\right) p(C_2)}\right)$$

$$= \underline{\frac{1}{s}(\lambda_1 - \lambda_2)^T x + \log g(\lambda_1) - \log g(\lambda_2) + \log p(C_1) - \log p(C_2)} \qquad (4.85)$$

$p(C_1 \mid x) \geq p(C_2 \mid x)$ となる境界は
xの線形関数.

$K > 2$ クラスの場合

$$a_k = \log\left(p(x \mid \lambda_k, s)\, p(C_k)\right)$$

$$= \log\left(\frac{1}{s} h\left(\frac{1}{s}x\right) g(\lambda_k) \exp\left(\frac{1}{s}\lambda_k^T x\right) p(C_k)\right)$$

$$= \underbrace{\frac{1}{s}\lambda_k^T x + \log g(\lambda_k) + \log p(C_k)}_{= a_k(x)} + \log\left(\frac{1}{s} h\left(\frac{1}{s}x\right)\right)$$

xの線形関数.