情報認識
「ガウス混合モデル」

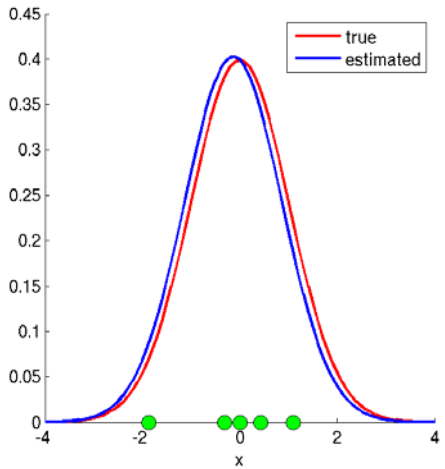
- **担当教員： 杉山 将（計算工学専攻）**
- **居室：** W8E-505
- **電子メール:** sugi@cs.titech.ac.jp

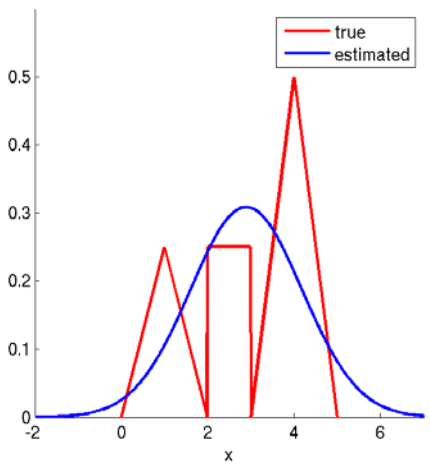
モデルの表現力

■ガウスモデル＋最尤推定：

- モデルが (大体) 正しい場合，訓練標本数が比較的少なくても推定精度が良い
- モデルが単純なため，表現できる確率密度関数の形が限られる

$$q(x;\mu,\Sigma)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$





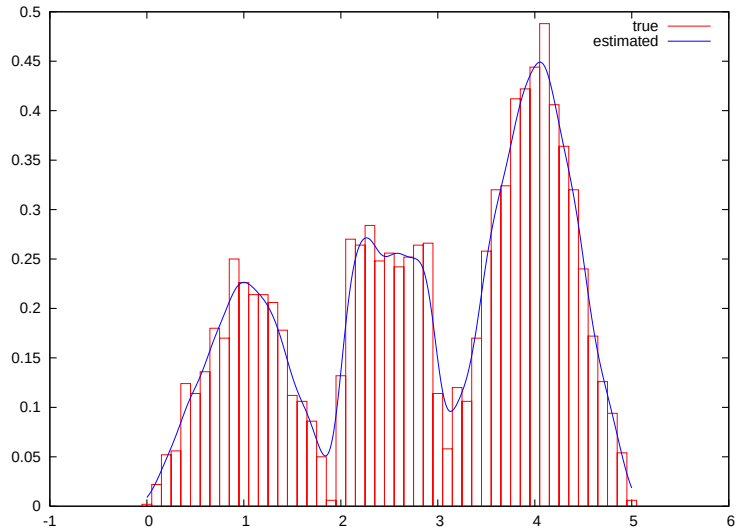
モデルの表現力 (続き)

■ ガウス核密度推定：

- 任意の確率密度関数を近似できる
- 精度良く近似するためには多数の訓練標本が必要

$$\hat{p}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

$$K(x) = \frac{1}{(2\pi)^{d/2}} \exp(-\tfrac{1}{2} x^T x)$$



ガウス混合モデル

■ **ガウス混合モデル**(Gaussian mixture model)：

$$q(x;\theta)=\sum_{j=1}^{m} w_j \phi(x;\mu_j,\sigma_j)$$

$$w_j \geq 0, \sum_{j=1}^{m} w_j = 1$$

$$\theta = \{w_j, \mu_j, \sigma_j\}_{j=1}^{m}$$

$$\mu_j \in \Re^d, \sigma_j > 0$$

$$\phi(x;\mu,\sigma)=\frac{1}{(2\pi\sigma^2)^{d/2}}\exp\left(-\frac{(x-\mu)^T(x-\mu)}{2\sigma^2}\right)$$

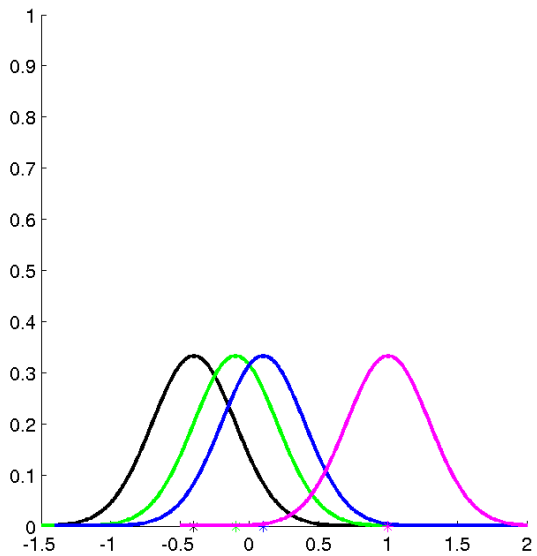
■ **注意**： $q(x;\theta)$ **は確率密度関数なので**

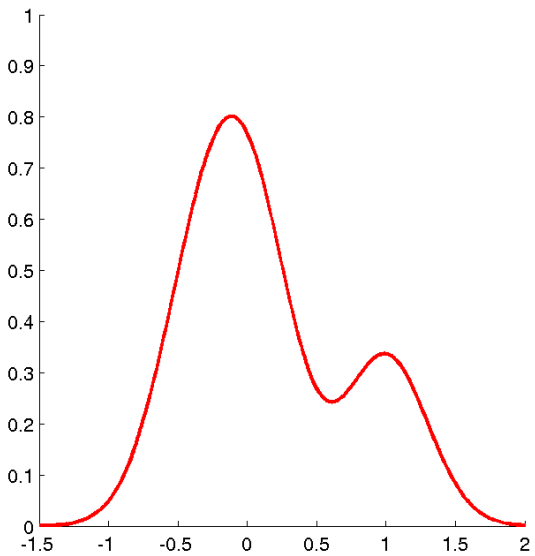
$$\int_D q(x;\theta)dx = 1 \quad \forall x \in D, q(x;\theta) \geq 0$$

を満たす．

ガウス混合モデル

- 有限個のガウスモデルの線形結合：





- 通常のガウスモデルよりも複雑な確率密度関数を表現できる．
- ガウス核密度推定より単純なので，訓練標本が比較的少ない場合でも推定精度が良い．

ガウス混合モデルの最尤推定

- **最尤推定**：(対数) 尤度 (訓練標本 $\{x_i\}_{i=1}^n$ が生成される確率) を最大にするように θ を定める

$$\log L(\theta) = \sum_{i=1}^{n} \log \sum_{j=1}^{m} w_j \phi(x_i; \mu_j, \sigma_j)$$

- 但し，**拘束条件** $w_j \geq 0, \sum_{j=1}^{m} w_j = 1$ を考慮しなければならない！

- $w_j = \dfrac{\exp(\gamma_j)}{\sum_{j'=1}^{m} \exp(\gamma_{j'})}$ とおき，$\gamma_j \in \Re$ を決定することにすれば，拘束条件は自動的に満たされる．

尤度方程式

■ 最尤推定解の**必要条件**：

$$\frac{\partial \log L}{\partial \gamma_j} = 0, \quad \frac{\partial \log L}{\partial \mu_j} = 0, \quad \frac{\partial \log L}{\partial \sigma_j} = 0$$

より，最尤推定解は以下を満たす (**証明は宿題**)

$$\hat{w}_j = \frac{1}{n}\sum_{i=1}^{n} \eta_{i,j} \qquad \hat{\sigma}_j = \sqrt{\frac{1}{d}\frac{\sum_{i=1}^{n} \eta_{i,j}(x_i - \hat{\mu}_j)^T (x_i - \hat{\mu}_j)}{\sum_{i'=1}^{n} \eta_{i',j}}}$$

$$\hat{\mu}_j = \frac{\sum_{i=1}^{n} \eta_{i,j} x_i}{\sum_{i'=1}^{n} \eta_{i',j}} \qquad \eta_{i,j} = \frac{\hat{w}_j \phi(x_i; \hat{\mu}_j, \hat{\sigma}_j)}{\sum_{j'=1}^{m} \hat{w}_{j'} \phi(x_i; \hat{\mu}_{j'}, \hat{\sigma}_{j'})}$$

■ しかし，この連立方程式は簡単に解けない．

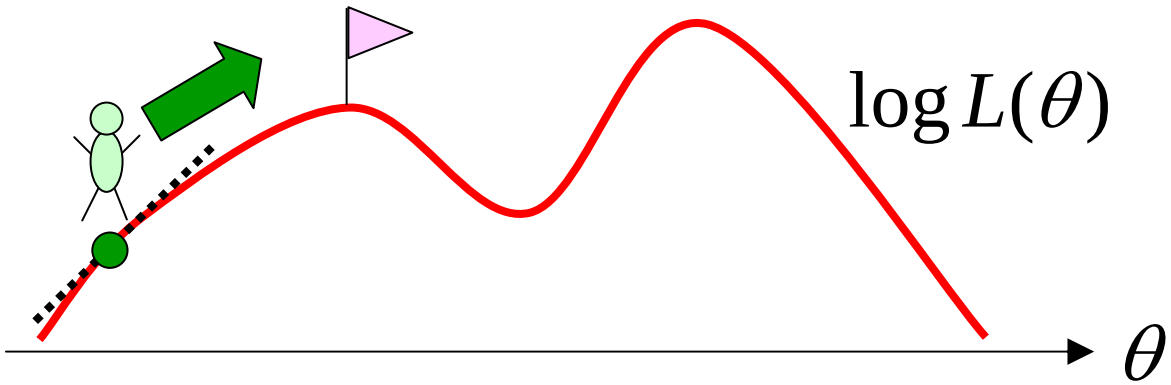
勾配法

- **勾配法**(gradient method)：
  1. 適当に初期値 $\hat{\theta}^{(0)}$ を定める．
  2. 勾配を上るようにパラメータを更新する：

$$\hat{\theta}^{(t+1)} \leftarrow \hat{\theta}^{(t)} - \varepsilon \left. \frac{\partial \log L}{\partial \theta} \right|_{\theta=\hat{\theta}^{(t)}}$$

  3. 収束するまで勾配上昇を繰り返す．

- **局所最適解**(local optimal solution)が求まる．



EMアルゴリズム

- EMアルゴリズム(expectation-maximization algorithm):
  1. 適当に初期値 $\hat{\theta}^{(0)}$ を定める.
  2. Eステップ: $\hat{\theta}^{(t)}$ の値を用いて, $\log L(\theta)$ の扱いやすい下界 $b^{(t)}(\theta)$ を求める

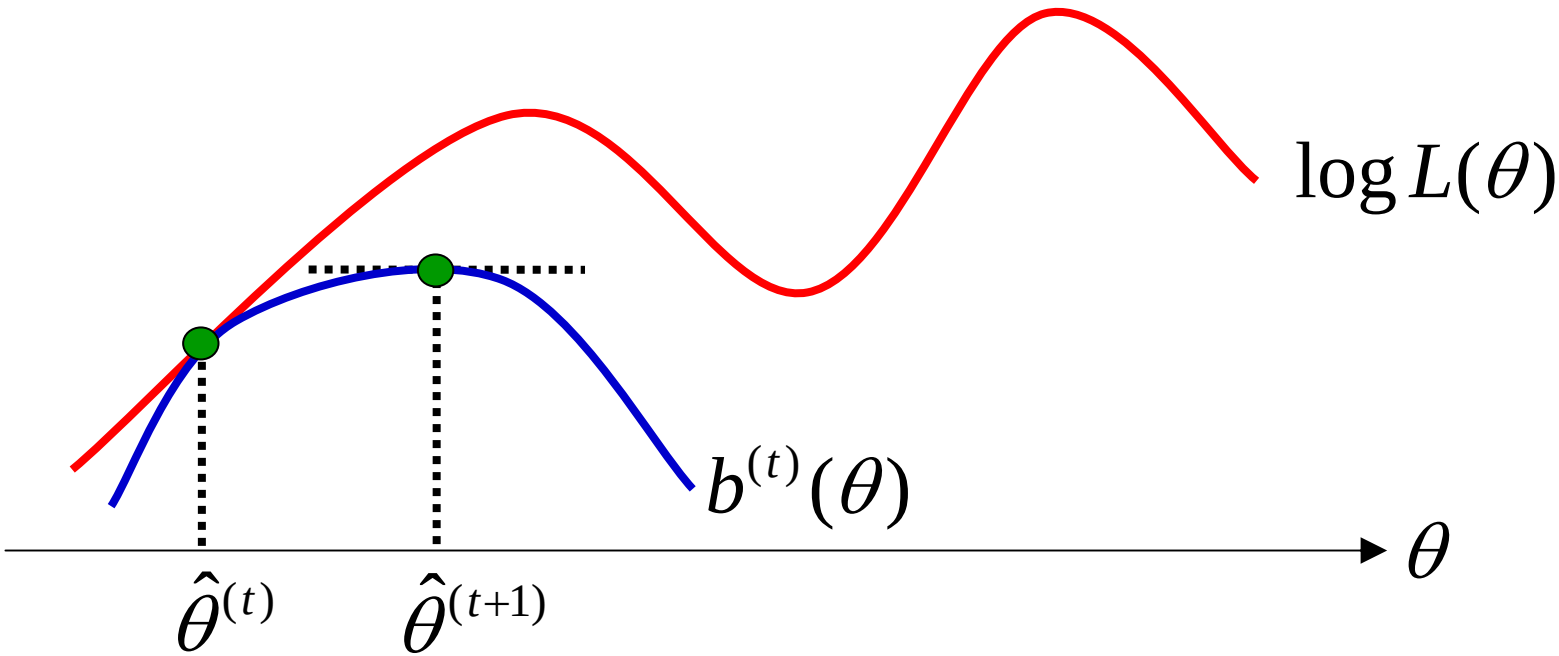
$$\forall \theta, \quad \log L(\theta) \geq b^{(t)}(\theta)$$

  3. Mステップ: $b^{(t)}(\theta)$ を最大にする $\hat{\theta}^{(t+1)}$ を求める

$$\hat{\theta}^{(t+1)} = \arg\max_{\theta} b^{(t)}(\theta)$$

  4. 収束するまで更新を繰り返す.

EMアルゴリズム

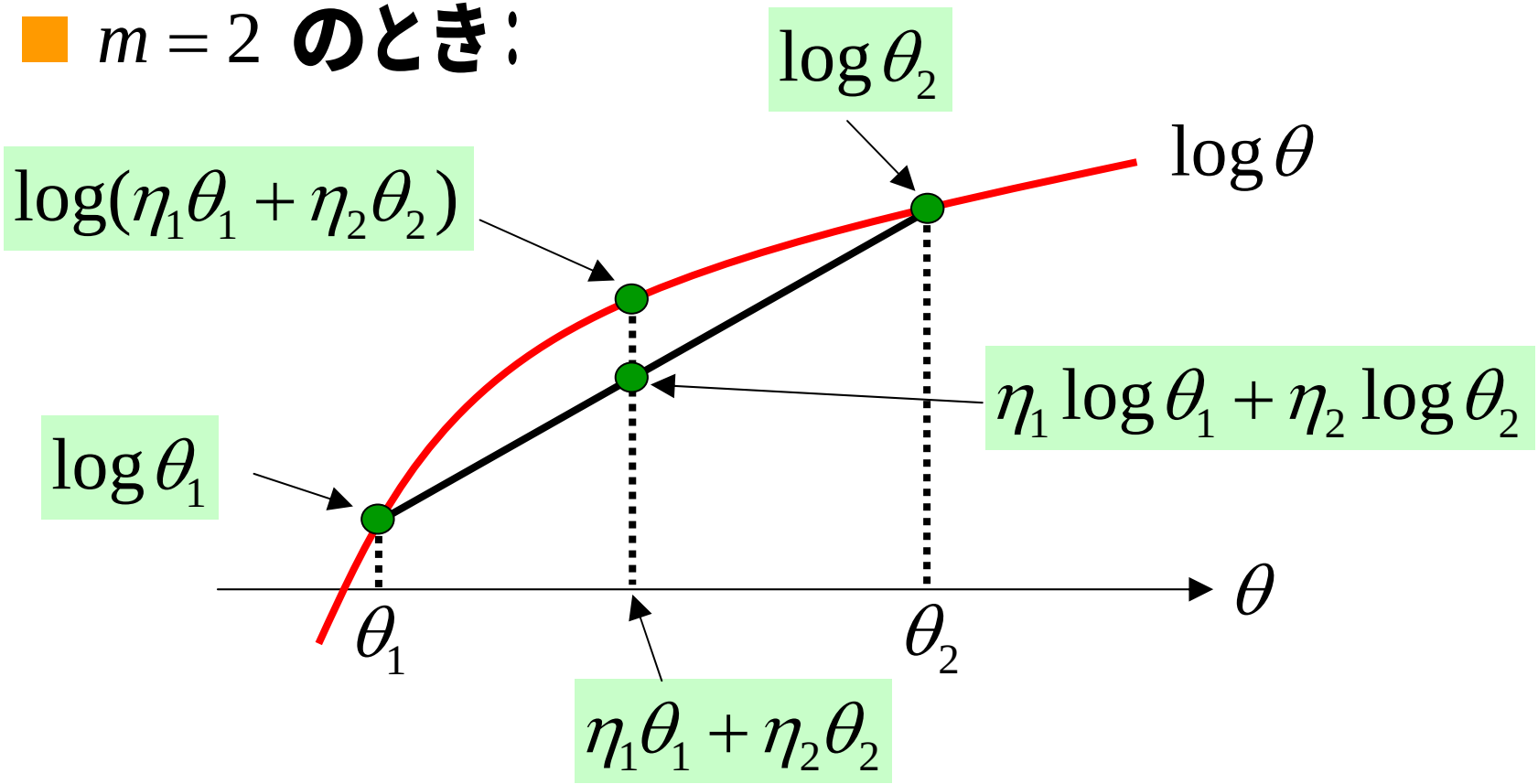


EMアルゴリズムの導出

■ ジェンセンの不等式(Jensen's inequality):

$$\log \sum_{j=1}^{m} \eta_j \theta_j \geq \sum_{j=1}^{m} \eta_j \log \theta_j \qquad \eta_j \geq 0, \sum_{j=1}^{m} \eta_j = 1$$

■ $m = 2$ のとき：



EMアルゴリズムの導出 (続き)

- ジェンセンの不等式より,

$$\log \sum_{j=1}^{m} v_j = \log \sum_{j=1}^{m} \eta_j \frac{v_j}{\eta_j} \geq \sum_{j=1}^{m} \eta_j \log \frac{v_j}{\eta_j}$$

$$\eta_j \geq 0, \sum_{j=1}^{m} \eta_j = 1$$

- 対数尤度の下界を求める:

$$\log L(\theta) = \sum_{i=1}^{n} \log \sum_{j=1}^{m} w_j \phi(x_i; \mu_j, \sigma_j) \geq b^{(t)}(\theta)$$

$$b^{(t)}(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \eta_{i,j}^{(t)} \log \frac{w_j \phi(x_i; \mu_j, \sigma_j)}{\eta_{i,j}^{(t)}}$$

$$\eta_{i,j}^{(t)} = \frac{\hat{w}_j^{(t)} \phi(x_i; \hat{\mu}_j^{(t)}, \hat{\sigma}_j^{(t)})}{\sum_{j'=1}^{m} \hat{w}_{j'}^{(t)} \phi(x_i; \hat{\mu}_{j'}^{(t)}, \hat{\sigma}_{j'}^{(t)})}$$

EMアルゴリズムの導出 (続き)

$$b^{(t)}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{m}\eta_{i,j}^{(t)}\log w_j\phi(x_i;\mu_j,\sigma_j) - \sum_{i=1}^{n}\sum_{j=1}^{m}\eta_{i,j}^{(t)}\log\eta_{i,j}^{(t)}$$

■ **下界を最大にするパラメータを求める：**

$$\frac{\partial b^{(t)}}{\partial\gamma_j} = 0, \quad \frac{\partial b^{(t)}}{\partial\mu_j} = 0, \quad \frac{\partial b^{(t)}}{\partial\sigma_j} = 0$$

$$\hat{w}_j^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\eta_{i,j}^{(t)}$$

$$\hat{\mu}_j^{(t+1)} = \frac{\sum_{i=1}^{n}\eta_{i,j}^{(t)}x_i}{\sum_{i'=1}^{n}\eta_{i',j}^{(t)}}$$

$$\hat{\sigma}_j^{(t+1)} = \sqrt{\frac{1}{d}\frac{\sum_{i=1}^{n}\eta_{i,j}^{(t)}(x_i-\hat{\mu}_j^{(t)})^T(x_i-\hat{\mu}_j^{(t)})}{\sum_{i'=1}^{n}\eta_{i',j}^{(t)}}}$$

EMアルゴリズム

■Eステップ：下界は実際には計算する必要はない（下界は期待値なので“expectation”と呼ばれる）

$$\eta_{i,j}^{(t)} = \frac{w_j^{(t)} \phi(x_i; \mu_j^{(t)}, \sigma_j^{(t)})}{\sum_{j'=1}^{m} w_{j'}^{(t)} \phi(x_i; \mu_{j'}^{(t)}, \sigma_{j'}^{(t)})}$$

■Mステップ:パラメータを更新する

$$\hat{w}_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \eta_{i,j}^{(t)}$$

$$\hat{\mu}_j^{(t+1)} = \frac{\sum_{i=1}^{n} \eta_{i,j}^{(t)} x_i}{\sum_{i'=1}^{n} \eta_{i',j}^{(t)}}$$

$$\hat{\sigma}_j^{(t+1)} = \sqrt{\frac{1}{d} \frac{\sum_{i=1}^{n} \eta_{i,j}^{(t)} (x_i - \hat{\mu}_j^{(t)})^T (x_i - \hat{\mu}_j^{(t)})}{\sum_{i'=1}^{n} \eta_{i',j}^{(t)}}}$$

小レポート (第11回)

1. ガウス混合モデルの最尤推定量が以下の条件を満たすことを証明せよ.

- $$\hat{w}_j = \frac{1}{n}\sum_{i=1}^{n}\eta_{i,j}$$

- $$\hat{\mu}_j = \frac{\sum_{i=1}^{n}\eta_{i,j}x_i}{\sum_{i'=1}^{n}\eta_{i',j}}$$

$$\eta_{i,j} = \frac{\hat{w}_j\phi(x_i;\hat{\mu}_j,\hat{\sigma}_j)}{\sum_{j'=1}^{m}\hat{w}_{j'}\phi(x_i;\hat{\mu}_{j'},\hat{\sigma}_{j'})}$$

- $$\hat{\sigma}_j = \sqrt{\frac{1}{d}\frac{\sum_{i=1}^{n}\eta_{i,j}(x_i-\hat{\mu}_j)^T(x_i-\hat{\mu}_j)}{\sum_{i'=1}^{n}\eta_{i',j}}}$$

小レポート (続き)

2. **計算機実験**：ガウス混合モデルのＥＭアルゴリズムを実装し，適当な一次元の確率密度関数を推定せよ．
3. **計算機実験**：ガウス混合モデルを用いて手書き文字認識を行なえ．混合数 m は交差確認法を用いて決定せよ．

- 次回の講義は1月15日 (火)