

混合分布による学習のモデル化

赤穂 昭太郎

電子技術総合研究所

2000年3月17日

統計数理研究所研究会

「情報理論，統計科学，人工知能の接点を探る」

概要

1. はじめに：
混合分布モデルと人工知能
2. 人工知能への例題：
画像と音声からの概念獲得モデル
3. モデル選択に関する解析：
正規混合分布の尤度バイアスの非単調性について

1 混合分布モデルと人工知能

- 人工知能の問題背景
 - 観測情報が部分的
 - 環境や処理対象が複雑
 - ルールベース + メモリベース (データマイニング)
- 混合分布モデル
 - パラメトリック ↔ ノンパラメトリック
 - 統計的モデル → 複雑性の回避
 - 階層ベイズの簡単な場合 → 情報の部分性を補う
 - EM アルゴリズム*, MCMC, 平均場近似 など豊富なアルゴリズム
 - エキスパートによる分業

- アプリケーション (RWC = Real World Computing)
 - 自律ロボットナビゲーション (事情通口ロボット → 本村さん)
 - マルチモーダル対話システム

2 画像と音声からの概念獲得

長谷川 修, 吉村 隆, 麻生 英樹, 速水 悟 (電総研) との共同研究 (信学論 J80-A, 9, 1997)

2.1 マルチモーダル対話システムと学習

複数のモダリティ 知的な対話システムではユーザの好みに応じて, 複数のモダリティを扱う必要がある. モダリティの例: ジェスチャー, 音声発話, マウス入力 etc.

学習 ユーザや環境の情報をシステムに学習してもらいたい, ユーザからの情報は部分的である.

Symbol と Pattern Pattern 的な情報があふれていて, Symbolic な情報は隠れていることが多い.



複数のモダリティを使った学習を通じて, 陽には得られない知識を獲得できないか?

2.2 課題

モダリティ 音声と画像

一対多の写像の学習 もの (画像) には複数の属性がある (白い, コップ)

属性の分類の学習 どの属性かは教えない (「白い」=色, 「コップ」=名前などの情報)

課題 新しいもの (画像) に対するすべての属性を音声で答える.

学習データの例:		学習後:	
画像 X	音声 Y	画像 X	音声 Y
[白いコップ]	“しろい”	[白いペン] →	“しろい” + “ぺん”
[青いペン]	“ぺん”	[青い本] →	“あおい” + “ほん”
...
[赤い本]	“あかい”	[赤いコップ] →	“あかい” + “こっぷ”

データ 画像 (x) と 音声 (y) のペア

音声 y は x の $K(=3)$ 種類の属性 (色 y_1 , 形 y_2 , 大きさ y_3) のうちのひとつ

課題 新たに与えられた x に対するすべての属性を出力する.

処理過程の概要

1. 特徴抽出 (各モダリティ内の処理 / 情報源の種類に依存した汎用処理)
2. 正準相関分析 (情報の統合)
3. 混合モデルの学習 (一対多の対応の学習 [EM 法])
4. 新規画像の処理 (学習結果による分類と NN による想起)

2.3 正準相関分析による次元圧縮

次元の問題 特徴抽出されたもの同士の対応を学習するにはそれぞれの次元が大きすぎる

- 移す空間の次元が大きすぎる → 対応づけが困難
- 移す空間の次元が小さすぎる → 必要な情報も落ちてしまう

正準相関分析 それぞれを最も相関が高くなるような共通の空間に写像する

線形写像の場合 (一般化) 固有値問題として解ける (SVD)

非線形? ニュラルネットを使った非線形正準相関分析 (Asoh et al 1994) という話もあるが
とりあえず線形でやる

2.4 混合分布によるモデル化

混合分布 どの属性を教えているかが隠れ変数

$$f(\mathbf{y} \mid \mathbf{x}; \theta) = \sum_{k=1}^K \xi_k f_k(\mathbf{y} \mid \mathbf{x}; \theta)$$

学習 EM アルゴリズム (隠れ変数がある場合の最尤推定のための繰り返しアルゴリズム) .

この場合, 重み付きの写像の学習になる

k 番目の写像に関する \mathbf{x}, \mathbf{y} の重み (t -th ステップ)

$$q(k \mid \mathbf{y}, \mathbf{x}, \theta^{(t)}) = \frac{f_k(\mathbf{y} \mid \mathbf{x}, \theta^{(t)})}{f(\mathbf{y} \mid \mathbf{x}, \theta^{(t)})}$$

f_k の例 実験では f_k は (一番単純なものとして) 線形回帰モデルを使った .

2.5 パラメータの初期値

EM は局所最適解に収束 好ましくない局所最適解に収束しないように適切な初期値を定めることが必要

評価基準 同じ (あるいは類似の) \mathbf{x} に対して異なる y が存在する場合にはその y は異なる属性である

射影によるクラスタリング

$$\sum_{i < j} \frac{(\mathbf{c}'\mathbf{y}_i - \mathbf{c}'\mathbf{y}_j)^2}{\|\mathbf{x}_i - \mathbf{x}_j\|} \rightarrow \text{Maximize}$$

射影された軸上で属性の数だけ分割する (射影は固有値問題として解ける).

2.6 収集データ

画像データ ● 0～9 までの 10 種類のプラスチック製の模型 (形)

- 赤, 青, 黄, 緑の 4 色
- 大, 中, 小の 3 サイズ

位置と向きに若干の変化を加えて各クラス 29 枚.

音声データ 対応する単語 17 語 / 2 名の男性話者 / 各 28 回づつ発話

あか, あお, きいろ, みどり, ぜろ, いち, に, さん, よん, ご, ろく, なな,
はち, きゅう, おおきい, ちゅうくらい, ちいさい

訓練データ 学習時にランダムに組み合わせて学習セットを作成

2.7 画像特徴

前処理

- 原画像 : 320 × 240 フルカラー画像
- RGB → YIQ 変換を施し, 輝度を捨て, I 画像と Q 画像のみ使う
- 1/1, 1/4, 1/16 に間引いた 3 種の画像を生成.

特徴量

局所高次自己相関特徴 (各 25 次元, 計 150 次元 (Kurita, Otsu et al 1992))

$$\sum_{\mathbf{x} \in \text{全画素}} F(\mathbf{x})F(\mathbf{x} + \mathbf{a}_1)F(\mathbf{x} + \mathbf{a}_2)$$

ただし, $\mathbf{a}_1, \mathbf{a}_2$ は各成分が $-1 \leq a_i \leq 1$ を満たす局所的な displacement

画像の平行移動に関して不変

2.8 音声特徴

メルケプストラム分析 前後の無音区間を切り, 音声区間のみ分析 (1 ~ 12 次)

分析条件: 16KHz サンプリング, 分析フレーム長 25ms, フレーム周期 10ms, 分析窓: ハミング窓, プリエンファシス係数 0.97.

次元の正規化 分析フレームを間引いて, 時間方向に 10 フレームとなるように正規化を行なう (計 120 次元).

2.9 学習実験

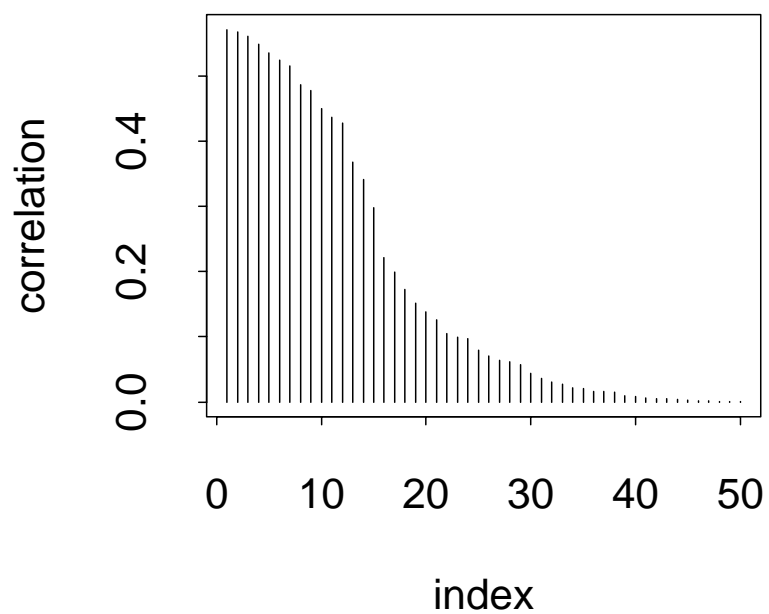
学習データ作成 DataSet(1) : 120 クラスの各クラスごと 10 サンプルの画像をランダムに選んで, 対応する全属性の音声ランダムに選んでペアを作成.

ランダムに間引いたもの : DataSet(0.8), DataSet(0.6), DataSet(0.4) とした.

正準相関分析 最適な次数を決定するのは困難だが, 10 次元とした ($\rho_i > 0.5$).

EM アルゴリズムによる学習

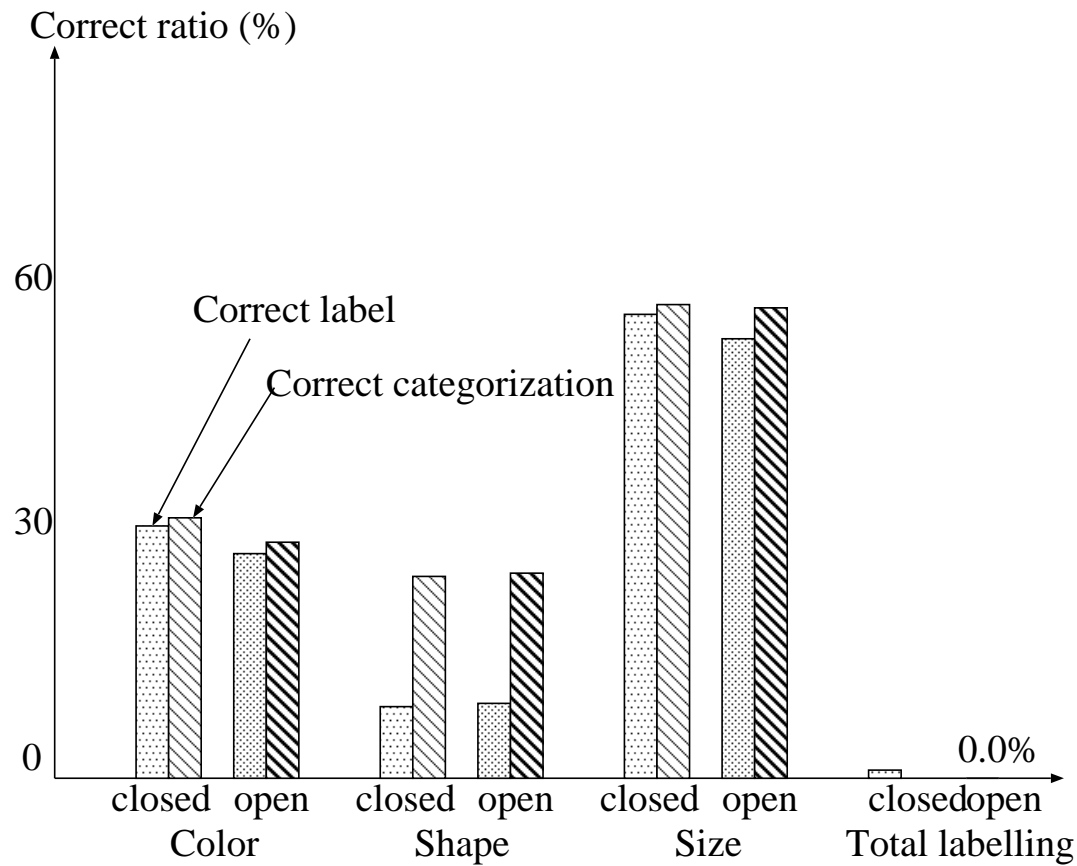
結果の評価 ランダムに発生させたテスト画像 (サンプルに含まれていない) を示し, それぞれの属性を答えられるかを評価. 学習結果により分類を行い, Nearest neighbor 法により過去の音声データとの対応づけを行なう.



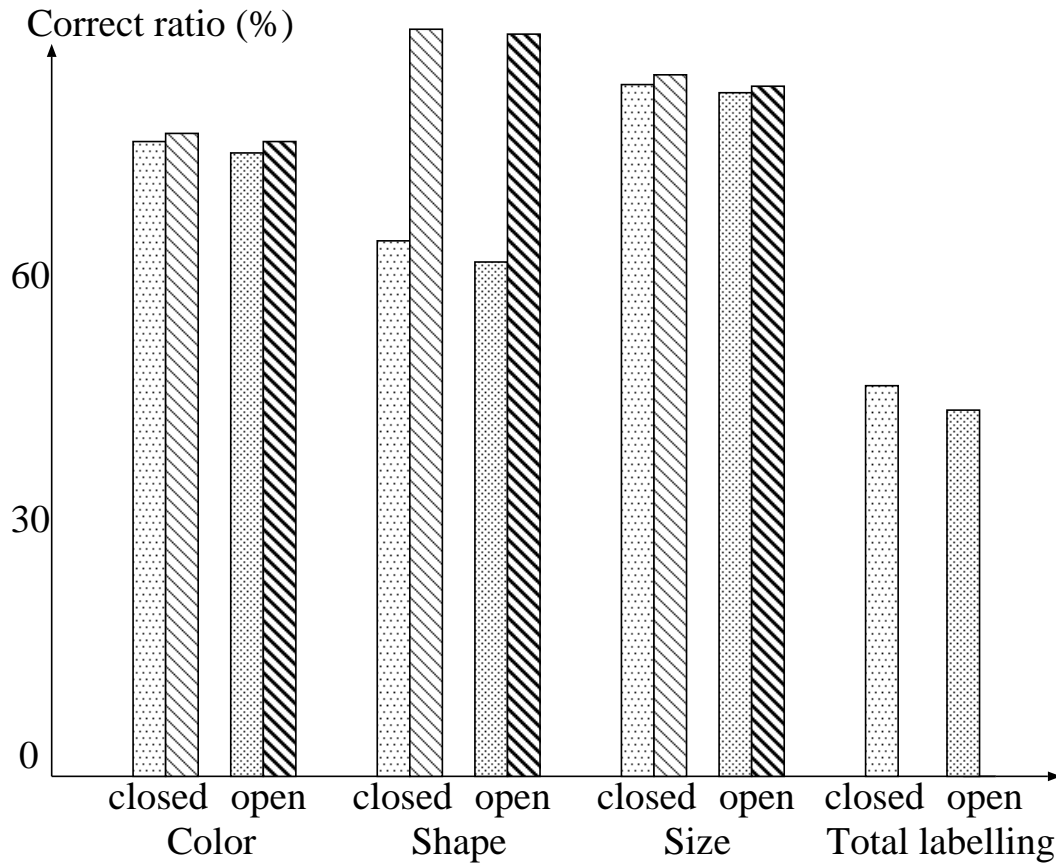
画像と音声の正準空間での相関 (属性の数 3 , DataSet(1)). 横軸 : 次元の index(50 次元まで) , 縦軸 : 各次元の相関の大きさ

2.10 認識結果

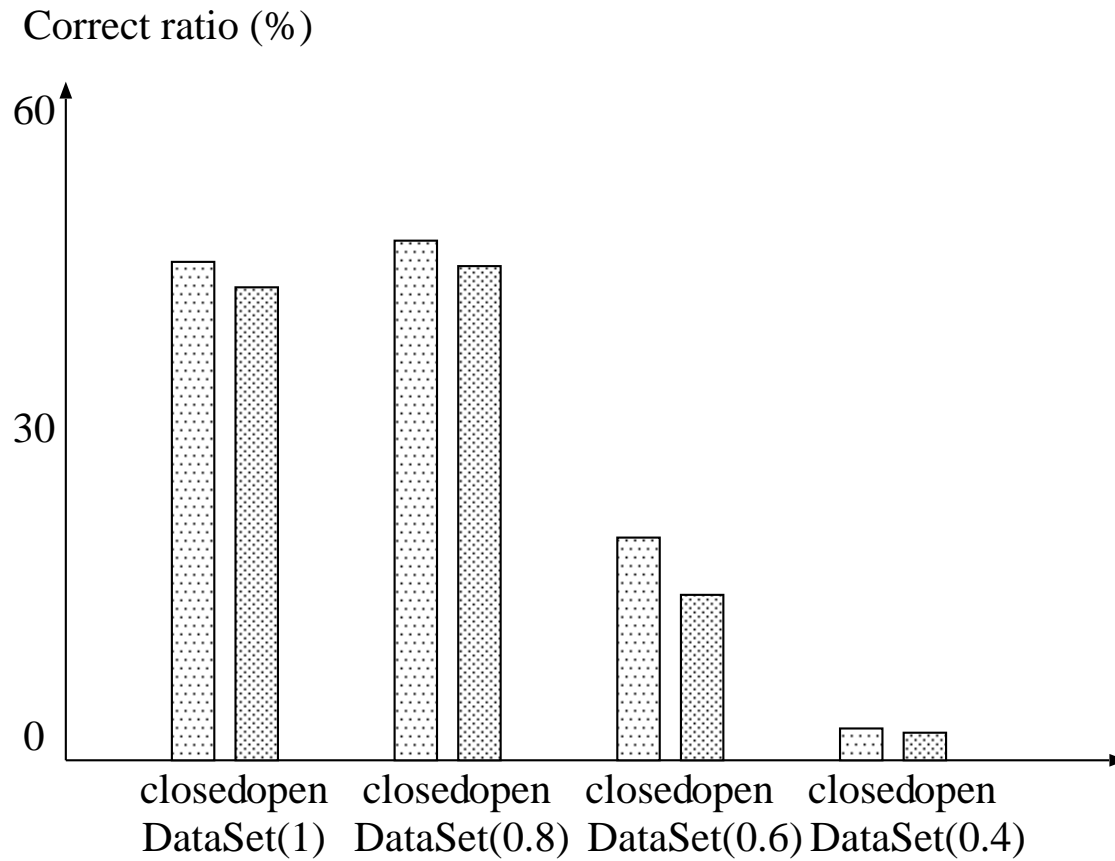
EM アルゴリズムの初期解での正解率 (3 クラス, DataSet(1) の学習)



EM アルゴリズム 30 ステップ後の正解率 (3 クラス, DataSet(1) の学習)



間引いたデータに対する学習の全属性の正解率 (3 クラス)



2.11 まとめと今後の課題

情報統合 正準相関分析

複数概念の獲得 混合分布モデル + EM アルゴリズム

問題点 1 概念内のばらつきが大きくなると、概念内を分類したほうが評価値が高くなることもある (今回の実験では数字の種類).

問題点 2 概念数の自動決定

今後の課題 柔軟なマルチモーダルインタフェースにどう結びつけていくか?

3 正規混合分布の尤度バイアスの非単調性について

Bert Kappen (Nijmegen Univ.) との共同研究 (Neural Computation (in press), 2000)

3.1 混合分布と汎化の問題

汎化の問題 要素の数を増やせば増やす程データへのフィットをよくできるが、テストサンプルへのフィットは悪くなる → モデル選択

RBBM Boltzmann machine を連続化した Radial Basis Boltzmann Machine :
正規混合分布の特殊な場合

特徴 分岐現象 (相転移) による階層的なモデル構造を示す

問題 分岐の様子とそれに伴う学習曲線の振舞いはどうなっているのか?

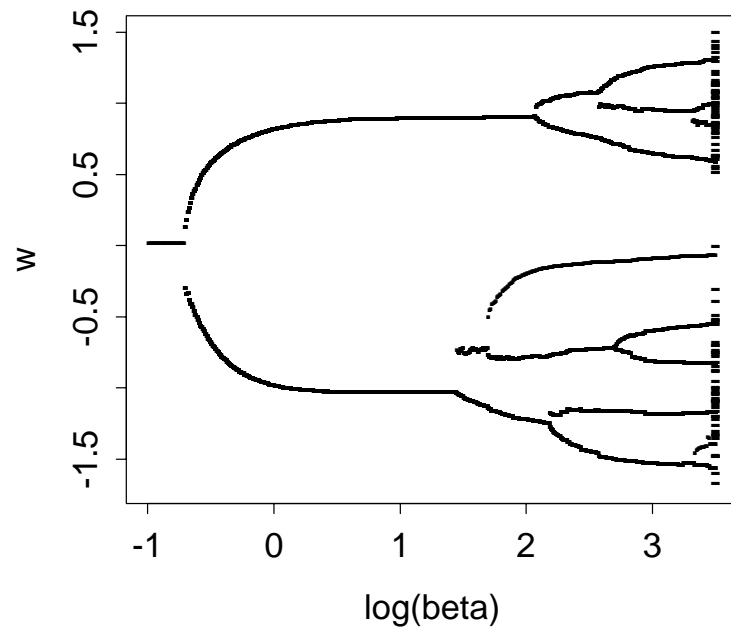
3.2 Radial Basis Boltzmann Machine (RBBM)

$$p(\mathbf{x} | W; \beta) = \frac{1}{h} \sum_{i=1}^h \sqrt{\frac{\beta}{\pi}} \exp(-\beta \|\mathbf{x} - \mathbf{w}_i\|^2). \quad (1)$$

RBBM の特徴

- 重みが等しい正規混合分布
- 分散共分散行列が等方的ですべて等しい
- モデルの複雑さは component の数 h とコントロールパラメータ β (逆温度) によって決まる .
- Boltzmann machine の連続値への一般化 (の特殊な場合)
(Rose et al 1990, Kappen 1993, 1995, Nijman and Kappen 1996)

温度を変化させたときの最尤解の振舞いの例 ($h = P = 100$) $u[0.5, 1.5] + N[-1, 0.3^2]/2$



- 小さな β では全体が一つのクラスタ
- β を大きくして行くと, 分岐 (相転移) が起きて小さなクラスタにわかれる
- 階層的クラスタリング
- h は十分大きく取ればモデルの複雑度は β のみが支配する .

3.3 RBMM の分岐点

分布 $q(\boldsymbol{x})$ の (真の) 最尤解

β が β_c よりも小さいとき すべての正規分布が一つに退化する

$$\boldsymbol{w}_i = \langle \boldsymbol{x} \rangle$$

最初分岐点 (Rose 1990) $\beta_c = \frac{1}{2\lambda_1}$

$\lambda_1 = \boldsymbol{x}$ の分散共分散行列の最大固有値

β が β_c より大きいとき

Assumption

1. x : 1 次元
 2. $q(x)$: 平均を軸にして対称
 3. h : 偶数
- 対称にすると，漸近展開したときに，低次の項が消えてくれる．
 - h が偶数というのは解を対称にするための条件で， h は大きいとしているのであまり本質的ではない

第 1 分岐点の振舞い Assumption の下で，以下の二つの場合に分類される (κ_4 : q の 4 次
キュムラント)

1. $\kappa_4 \neq 0$ のとき，分岐は 2-way.

分岐点近傍では

$$\Delta\beta \simeq \frac{s_4}{6(\sigma^2)^4}(\Delta w_i)^2, \quad (2)$$

$\Delta\beta = \beta - \beta_c > 0$, $\Delta w_i = w_i - \langle x \rangle$. σ^2 : 分散, s_4 : 4 次モーメント

2. $\kappa_4 = 0$ のとき，分岐は h -way.

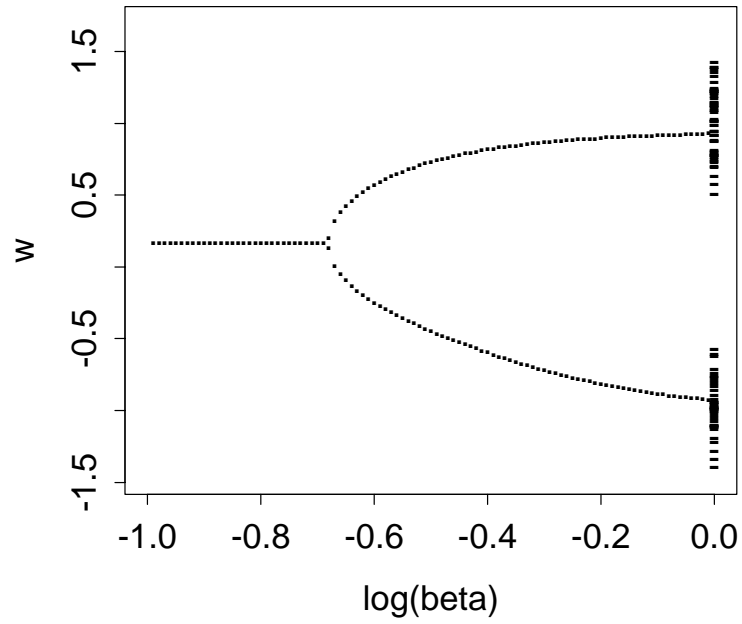
分岐点近傍では

$$\Delta\beta \simeq \frac{1}{2(\sigma^2)^2}\sigma_w^2, \quad (3)$$

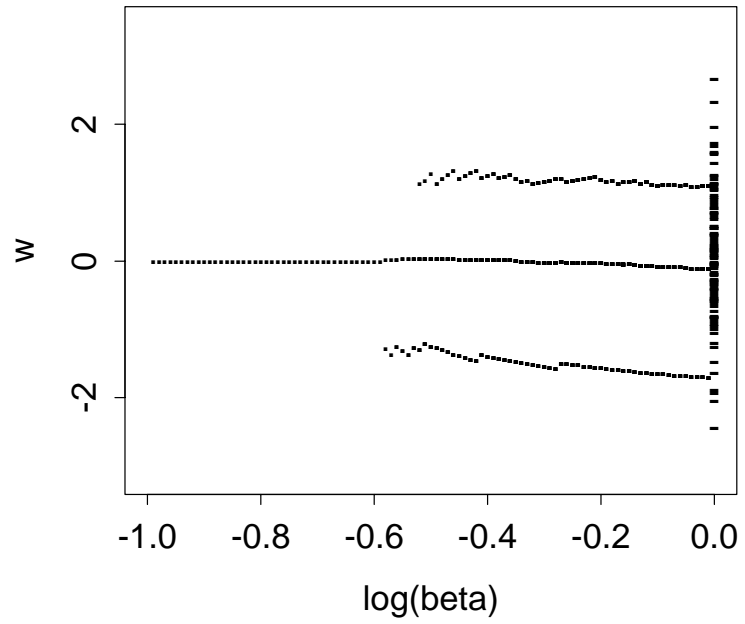
$$\sigma_w^2 = \frac{1}{h} \sum_i \Delta w_i^2, \quad \Delta w_i = w_i - \langle x \rangle.$$

- h が無限にあれば， $p(w)$ が正規分布になる .
- $q(x)$ の 4 次キュムラントの意味での正規性

$\kappa_4 \neq 0$ のときの分岐の例 2つの正規分布の混合分布



$\kappa_4 = 0$ のときの分岐の例 標準正規分布



3.4 汎化誤差の非単調性

汎化の問題 サンプル尤度はモデルを複雑にすればするほど大きくなる .

しかし , 真の分布に対する尤度は逆に小さくなる可能性がある .

モデル選択 モデルの複雑度を調節する指標によりモデル選択を行う

(対数) 尤度のバイアス サンプル尤度と真の尤度との差の期待値

AIC 尤度のバイアスに着目してバイアス補正により複雑度 (パラメータ数) を調節
真の分布が分布に入っているという条件

NIC AIC の一般化 (TIC, Moody's effective number of parameters)

NIC の振舞いを調べると , 見かけのパラメータ数と異なる振舞いがわかる

Neural 情報量基準 (NIC)

訓練サンプル $X^{(P)} = \{\mathbf{x}_1, \dots, \mathbf{x}_P\}$

サンプル対数尤度 $R_{\text{emp}}^{(P)}(W; \beta) \equiv \frac{1}{P} \sum_{i=1}^P \log p(\mathbf{x}_i | W; \beta)$

真の対数尤度 $R_{\text{exp}}(W; \beta) \equiv \langle \log p(\mathbf{x} | W; \beta) \rangle$

尤度のバイアス $\langle R_{\text{emp}}^{(P)}(W^{(P)}; \beta) \rangle - \langle R_{\text{exp}}(W^{(P)}; \beta) \rangle \simeq \frac{h_{\text{NIC}}}{P}$

$$h_{\text{NIC}}(\beta) \equiv \text{Tr}[H(W^*)^{-1} D(W^*)]$$

W^* : 真の最尤解,

$$H_{ij}(W) = -\langle \partial_i \partial_j \log p \rangle, D_{ij}(W) = \langle \partial_i \log p \partial_j \log p \rangle$$

1 セットの訓練サンプル $R_{\text{emp}}^{(P)}(W^{(P)}; \beta) - R_{\text{exp}}(W^{(P)}; \beta) \simeq \frac{h_{\text{NIC}}}{P} + \frac{U}{\sqrt{P}}$

U は平均 0 オーダー 1 で, 階層的モデル間では一致 (holds in RBBM)

疑問 RBBM で β を変化させた時 h_{NIC} はどうなるか?

分岐点より温度が高い場合 h_{NIC} を陽に計算できる .

$\beta < \beta_c$ のとき

$$h_{\text{NIC}}(\beta) = 2\beta \text{Tr}[V\mathbf{x}], \quad (4)$$

$V\mathbf{x}$ は $q(\mathbf{x})$ の共分散行列 .

分岐点より前 尤度のバイアスは β に比例して増加する

分岐点に近付くと (パラメータ数の) 1 に近付く

第二分岐点以降 おそらく定性的には各分岐点の前では線形に増加する

分岐後どうなるか? 直観的な予想 : パラメータが二つに増えるから NIC も増加するはず?

一般には解析が難しいので，Assumption (1次元対称な分布)の下で考える

1. $\kappa_4 \neq 0, s_4 \neq (\sigma^2)^2$ ならば

$$\lim_{\beta \downarrow \beta_c} \frac{\partial}{\partial \beta} h_{\text{NIC}}(\beta) = -\infty \quad (5)$$

$s_4 \neq (\sigma^2)^2$ は二つの δ 関数の混合分布を除くすべての分布について成り立つ。

2. $q(x) = (\delta(x-1) + \delta(x+1))/2$ のときは，

$$\partial h_{\text{NIC}}(\beta_c) / \partial \beta = -4.$$

結果

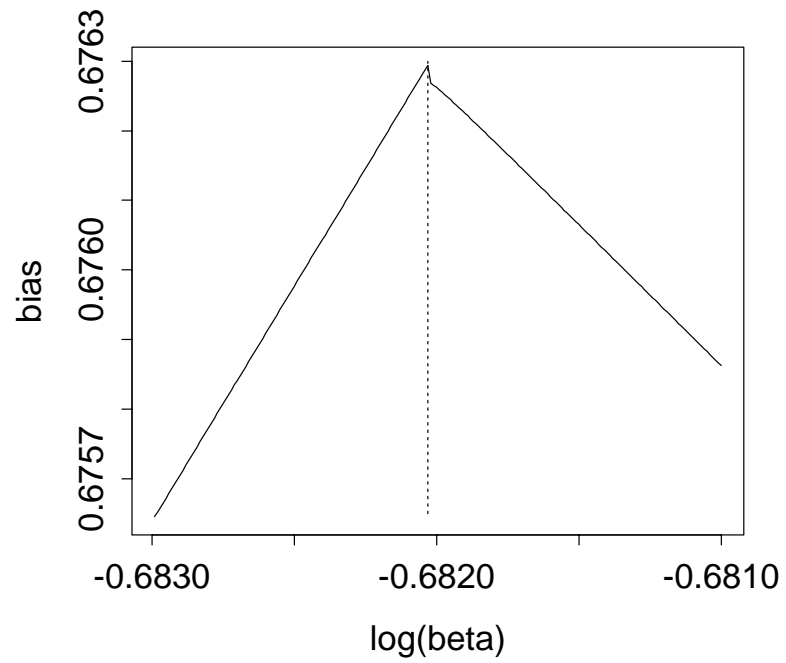
NIC の非単調性 分岐によってみかけのパラメータ数が増えても RBBM の NIC が減少することがある

汎化誤差 分岐点のまわりでは訓練誤差はほぼ定数 (傾き 0) なので, 分岐点の直前よりも分岐点の直後の方が (NIC の意味で) 汎化がよい

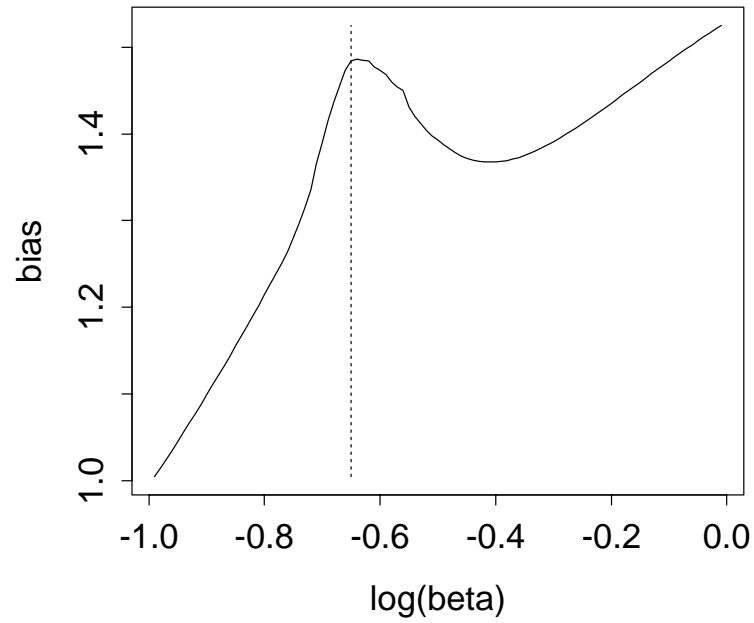
モデル選択 分岐点は汎化誤差の極大点になっているので, それより前と後にそれぞれモデル選択の局所最適解が存在している.

課題 $\kappa_4 = 0$ の場合は分岐が h -way なので解析が困難

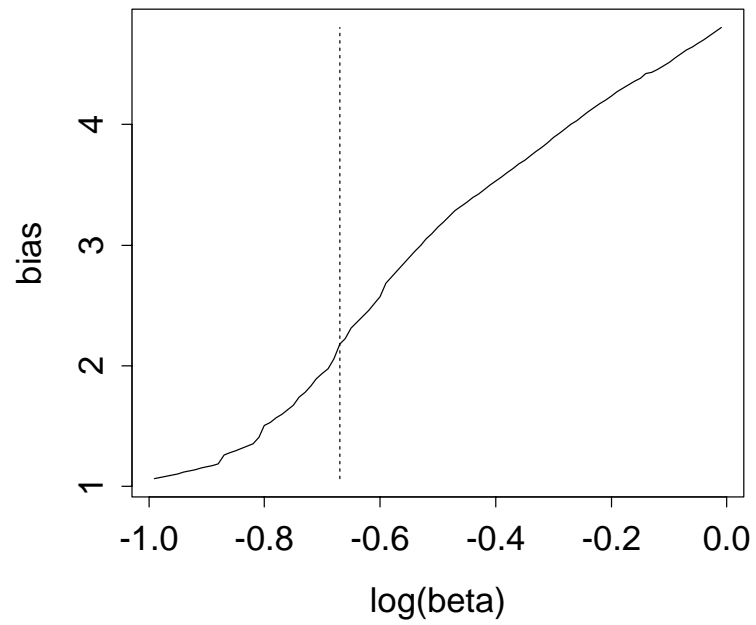
$\kappa_4 \neq 0$ の場合の NIC の振舞い (1 run)



$\kappa_4 \neq 0$ の場合の NIC の振舞い (50 run の平均)



$\kappa_4 = 0$ の場合の NIC の振舞い (1 run)



実際の汎化誤差との関係

NIC の仮定 NIC による汎化誤差の式は局所的に正規分布である必要がある

分岐点 本当に分岐点の近くでは 1-正規モデルと 2-正規モデルを行ったり来たりするので、NIC は使えない。

成立範囲 分岐点のゆらぎの大きさは $O(1/\sqrt{P})$ 程度なので、これより離れたところでは汎化誤差の式が成り立っている。

実験結果 実験的には、カーブがなまされる程度で、それほど極端なことは起きない

$-\infty$ の傾き 実験的には少ししか出ない

3.5 まとめと今後の課題

RBBM という正規混合モデルの一種について解析を行った。

分岐と正規性 分岐は 4 次のキュムラント κ_4 に依存して、正規的なら h -way それ以外なら 2-way になる。

NIC (分岐前) 分岐点より前では NIC は線形に増加する

NIC (分岐後) 分岐点を越えたところでは、 $\kappa_4 \neq 0$ のときは尤度のバイアスが β の増加とともに減少する

$\kappa_4 = 0$ に関しては実験的にはこのような非単調性は観測されなかった。

課題 高次元、第 2 分岐点以降に関しても定性的にはここでの解析と同様の結果が得られると考えられるが、ちゃんと示すのは大変そう。