

## 課題5

- 目的：認識における Bottom-up と Top-down 処理の絶妙な関係（鶏と卵）を味わう
  - 「認識」と「生成」との間のギャップ（例：読める  $\iff$  書ける）
  - Bottom-up 的手法：data driven（モデルなし，ニューラルネットなど）
  - Top-down 的手法：model-based（テンプレートマッチングなど）
  - 通常的手法はどちらかに属する．それぞれ弱点がある！「確率的な生成モデル」は，両方の長所を同時にもつ．
  - 生成モデル：対象を的確に生成できるモデルを作っておけば，正確な認識ができる．
  - モデルのパラメータは例題データから学習することができる（オプション）．
- 課題 5.1（必須）
  - ノイズを含んだデータ 2 例 r2494294obs, r7773obs を講義のホームページからダウンロードし，自分の作ったプログラムを使い，もとの値の推定を試みる．  
<http://www.cs.miyazaki-u.ac.jp/~date/lectures/neural/index.html>  
自分の作ったプログラムが正しくもとの値の推定をできているかは，チェック用データ r20070703obs を使い，その結果が r20070703map となっているか確認することでチェックできる（map はデータの値+3 になっていることに注意）．  
講義で習った手法とは異なる手法を用い（例， $y_i > 0.5 \rightarrow x_i = 1$  など），問題の解決を試み性能の良し悪しを比較する
- 課題 5.2（自由課題．特別演習で本課題をおこなった人は 1. が必須）
  1. 得られた  $\vec{x}_{\text{map}}$  の事後確率  $\text{Prob}(\vec{x}_{\text{map}}|\vec{y})$  を計算する．自分で問題を 100 例生成し，「事後確率」と「もとのデータをどのくらい正しく復元できたか」（復元率）との関係を考察する．
  2. パラメータ ( $p_{ij}$ ) の値を未知として，データから推定してみる（EM アルゴリズム）．
  3. 事後確率分布から正確にサンプル  $\vec{x}$  を得る．
  4. Gibbs Sampler を使用し事後確率分布  $\text{Prob}(\vec{x}|\vec{y})$  からサンプル  $\vec{x}$  を得る．
  5. 状態数を 10 ~ 50 値にして，音声データ，画像をモデル化する．音声データのサンプルは講義のページに a200p50s6.dat など置いてある．
  6. その他，自由にいろいろなデータを学習させることを試してみる …
- 提出物：プログラムとレポート．
  - 何を試して，どういう結果になったか．
  - このモデルでイメージできた点，不満足な点．
  - レポート提出だけでなく，口頭による説明（特に動的計画法について）を求めます．
- レポートは電子メール（宛先：date@cs.miyazaki-u.ac.jp）または紙で提出．
- 使用するプログラムは何でもよく，自分が書きやすい言語を使えばよい．プログラムの書き方などがわからなければ締め切り間際でなく早めに相談にくること．
- 提出期日：7月31日（金）