データ駆動科学と
ディープラーニング

東京大学　大学院新領域創成科学研究科
複雑理工学専攻
岡田真人

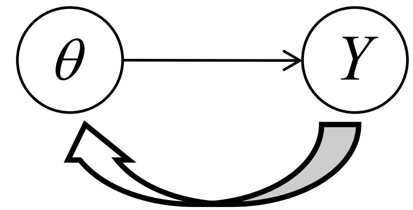
本スライドのまとめ

- 今後、ディープラーニングはデータ駆動科学においても、重要な情報数理基盤になると考えられる。
- ただし、ディープラーニングが、スパースモデリングやベイズ推論を置き換える立場にはならない。
- ディープラーニングを用いる必要がない場合には、従来のアプローチを引き続き用いるべきである。
- ベイズ推論におけるディープラーニングの役割は、観測データの生成モデルや事前確率をデータから推定する場合い使える可能性がある。その両方とも、ディープラーニングのアーキテクチャの最適化には、スパースモデリングが必須となると考えられる。
- 線形回帰や一般化線形モデルのような比較的単純なモデルで予測不可能な場合、スパースモデリングにディープラーニングを用いる必要がある。その際、インディケーターを用いた全状態探索が必須となる。
- 多層畳み込みネットワークモデルの、特徴量自動抽出機構を、どのようにデータ駆動科学に組み込むかが、今後のひとつの解題である。

ベイズ推論とディープラーニング

$$p(Y,q) = p(Y \mid q)\,p(q) = \underline{p(q \mid Y)\,p(Y)}$$

生成（因果律）



⇩

＜ベイズの定理＞

$$p(q \mid Y) = \frac{p(Y \mid q)\,p(q)}{p(Y)} \;\mu\; \exp(-\,nE(q))\,p(q)$$

$p(q \mid Y)$ ：事後確率。データが与えられたもとでの，パラメータの確率.

$p(\theta)$ ：事前確率。あらかじめ設定しておく必要がある。
これまで蓄積されてきた科学的知見

ディープラーニングで、ベイズ推論における
生成モデルや事前確率をデータから自動抽出する

スパースモデリングとは

スパースモデリング基本的な考え方は

(1) 高次元データの説明変数が次元数よりも少ない(スパース(疎) である) と仮定し,

(2) 説明変数の個数がなるべく小さくなることと, データへの適合とを同時に要請することにより,

(3) 人手に頼らない自動的な説明変数の選択を可能にする枠組みである.

スパースモデリングの具体例
線形回帰における定式化

目的変数 (機能) $y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_N x_N$ 説明変数

$= \boldsymbol{\beta}\mathbf{x}$

サンプル数p, 説明変数の数 N

回帰係数 $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, 0, \ldots, 0, \beta_N)$

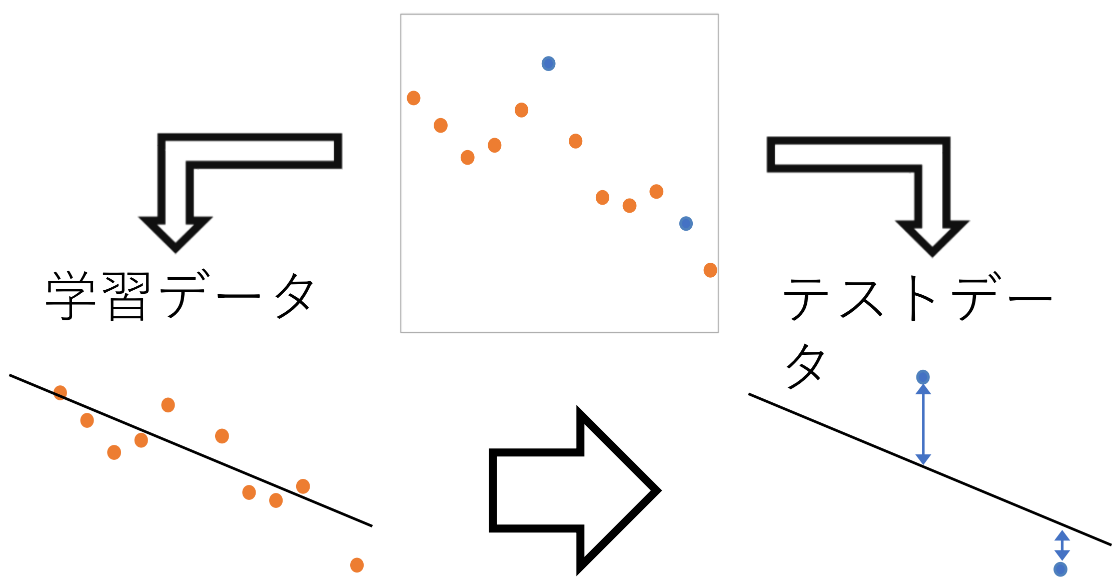
インディケータ $\mathbf{c} = (1, 1, \ldots, 0, \ldots, 0, 1)$

$\mathbf{c} = \{0,1\}^N$ $2^N - 1$ 通り

全状態探索(Exhaustive Search, ES) 法

Garside 1965, Ichikawa et al., 2014, Nagata et al., 2015, Igarashi et al., 2016, Igarashi et al., 2018

1. $\mathbf{c}$ を決めると記述子の係数 $\boldsymbol{\beta}$が決まる
2. その $\mathbf{c}$ に対する回帰モデル($\boldsymbol{\beta}$)の交差検証誤差(CVE)を求める
3. 各記述子の組み合わせ($\mathbf{c}$)を評価



交差検証によって，限られたデータから汎化誤差を推定する

線形回帰における全状態探索(ES-LiR) 法

インディケータ $\boldsymbol{c}$		$\boldsymbol{\beta}$ を用いて$\mathbf{y}$ の推定		CVE
$\boldsymbol{c}=(1,0,0)$	$\Longrightarrow$	$\hat{y}=b_0+b_1x_1$	$\Longrightarrow$	$error = \boldsymbol{0.20}$
$\boldsymbol{c}=(0,1,0)$	$\Longrightarrow$	$\hat{y}=b_0+b_2x_2$	$\Longrightarrow$	$error = \boldsymbol{0.21}$
$\boldsymbol{c}=(0,0,1)$	$\Longrightarrow$	$\hat{y}=b_0+b_3x_3$	$\Longrightarrow$	$error = \boldsymbol{0.23}$
$\boldsymbol{c}=(1,0,1)$	$\Longrightarrow$	$\hat{y}=b_0+b_1x_1+b_3x_3$	$\Longrightarrow$	$error = \boldsymbol{0.16}$
$\boldsymbol{c}=(1,1,0)$	$\Longrightarrow$	$\hat{y}=b_0+b_1x_1+b_2x_2$	$\Longrightarrow$	$\boxed{error = \boldsymbol{0.14}}$
$\boldsymbol{c}=(0,1,1)$	$\Longrightarrow$	$\hat{y}=b_0+b_2x_2+b_3x_3$	$\Longrightarrow$	$error = \boldsymbol{0.18}$
$\boldsymbol{c}=(1,1,1)$	$\Longrightarrow$	$\hat{y}=b_0+b_1x_1+b_2x_2+b_3x_3$	$\Longrightarrow$	$error = \boldsymbol{0.20}$
	線形回帰		CVEの計算	

2^3-1 = 8 通り，すべてのCVEをチェックする

記述子の個数Nが十分に大きい場合，計算量O(2^N)が膨大になる

スパースモデリングはディープラーニングを含めほとんどの学習機械に適用可能

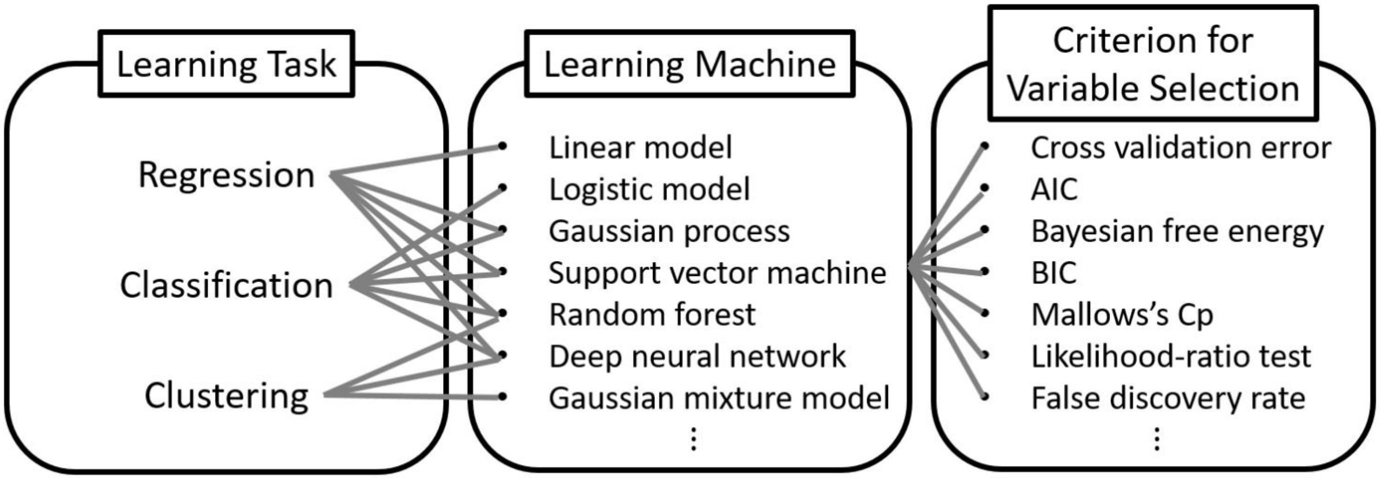


Figure 1. Constitution of exhaustive search. Exhaustive search is conducted on every triplet of learning task, learning machine and criterion for variable selection.

Yasuhiko Igarashi, Hiroko Ichikawa, Yoshinori Nakanishi-Ohno, Hikaru Takenaka, Daiki Kawabata, Satoshi Eifuku, Ryoi Tamura, Kenji Nagata, and Masato Okada
"ES-DoS: Exhaustive search and density-of-states estimation as a general framework for sparse variable selection"
International Meeting on "High-Dimensional Data-Driven Science" (HD3-2017)
Journal of Physics: Conference Series, 1036, 012001, (2018)