ビッグデータ、ＡＩ時代に必要とされる統計的推論法の習得に必要な数学教育

樋口知之　（情報・システム研究機構　統計数理研究所）

個人的意見

以下の講演内容は、あくまでも個人的意見であり、統計数理研究所所長としての見解でないことをご承知ください。

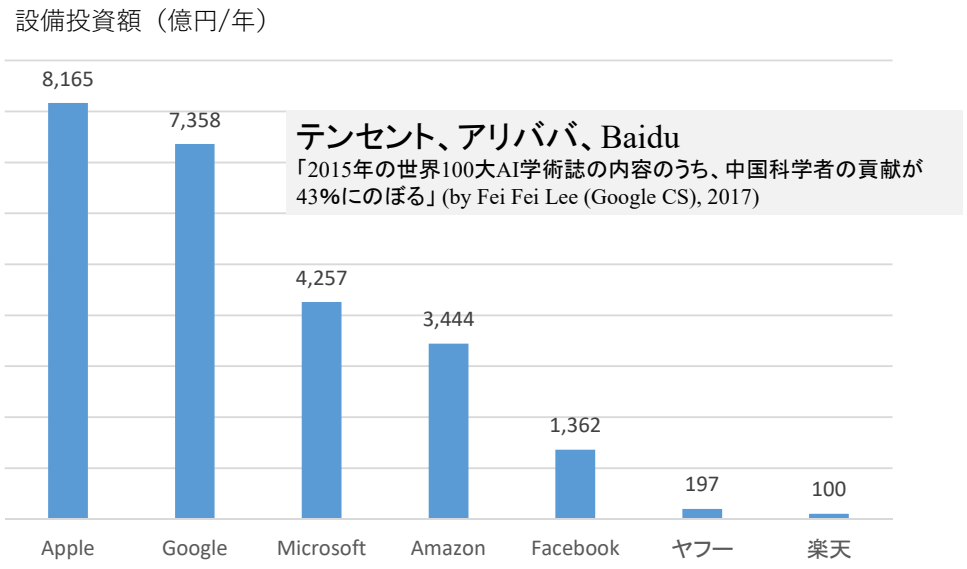
また、以下の意見に大賛成です。

- データリテラシー教育は初等教育から行う
- 初歩的な統計分析はすべての学部学生に習得させる
- 成人の学び直しとして統計学は有用

✓ 昨今、国内外で話題となっている『 p値』についての議論は本日はしません。

✓ 習得内容が学部後期あるいは大学院レベルになっています。

投資額の格差（USのAI×データ企業との比較）



テンセント、アリババ、Baidu
「2015年の世界100大AI学術誌の内容のうち、中国科学者の貢献が43%にのぼる」(by Fei Fei Lee (Google CS), 2017)

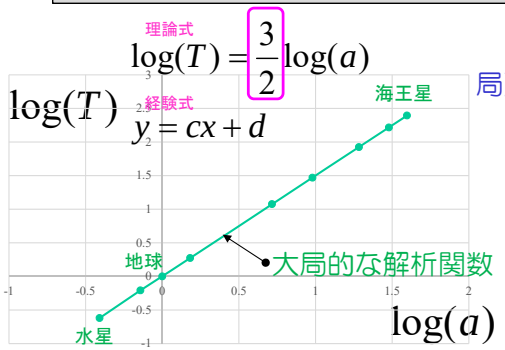
各社のFY2013 IR資料：有形固定資産の取得、100円/$で換算

（統数研・神谷特任教授のスライド(2017)から改編）
安宅和人「"シン・ニホン":AI×データ時代における日本の再生と人材育成」も参照
(http://www.meti.go.jp/committee/sankoushin/shin_sangyoukouzou/pdf/013_06_00.pdf)

概括：帰納法とデータ今昔

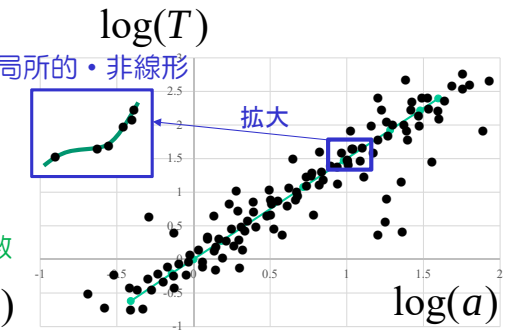
ケプラーの第三法則　師匠であったティコ・ブラーエの観測記録から推定し定式化

惑星の公転周期 T の2乗は、楕円軌道の半長軸 a の3乗に比例する



ティコ・ブラーエの観測データ

狙って観測、帰納推論から経験則、そして一般則（万有引力の法則）を導出



ビッグデータ

無目的・副産物的にデータが蓄積され、経験則のみでOK（目的－予測・判別－が達成）

ビッグデータ時代以前: 良質な空間を見つけること
1）説明変数の選択　2）線形性　3）少ないパラメータ

統計的推論法のパラダイムシフトを促す外的要因

ビッグデータ、ビッグデータ、…ビッグデータ

- （ちょっと前） クラウド、スマホ
- （今） IoT、AIアシスト（Amazon Echo, Google Home）
- （ちょっと先） Lifi ※（次世代Wifi。今の100倍）、5G（次世代通信。10Gbps）

データ駆動サービス社会

- “役に立つ”問題、解きたい問題が爆発的に増加
- モノからサービス（体験）へ、全体から個（個別、個性、固有）へ
- 発見から予測へ、因果から相関へ
- 統計的モデリングから最適化関数設計へ、統計的推測から最適化へ

目的特化型計算機

- NVIDIA GPGPU
- （量子）アニーラー、AIチップ

※Light Fidelity FI

データの観点で、何が変わったのか？

$$z_i := (y_i, \boldsymbol{x}_i) \quad i = 1, \ldots, N$$

ラベル　説明変数ベクトル　特徴ベクトル、記述子　サンプル数

$$p = \dim(\boldsymbol{x}_i)$$

データ（サンプル）次元

2000年代始め　新*NP*問題　$N << p$

今でも、理論統計学の分野では世界的に研究が盛ん

2010年代　*N*がさらに激増

- 計算量が $O(N)$ では役に立たない
- オンライン学習（バッチ学習・処理ではダメ）
- モデル等比較にはクロスバリデーションが標準

統計学は対応できているのか？

- 計算量に配慮しているか？：計算量が $O(N)$ では役に立たない
- *p*次元空間内の分布の表現をどうするのか？
- オンライン学習（推定）法の研究は？
- 統計的モデリングはエキスパートがやるのか？：深層学習とスパースモデリングの勃興
- 可視化技術は人が見るため？：深層学習（とくにCNN：Convolutional NN）の画像処理における圧倒的性能
- 計算プラットフォームの整備は？
- 目的特化型計算機への実装は？

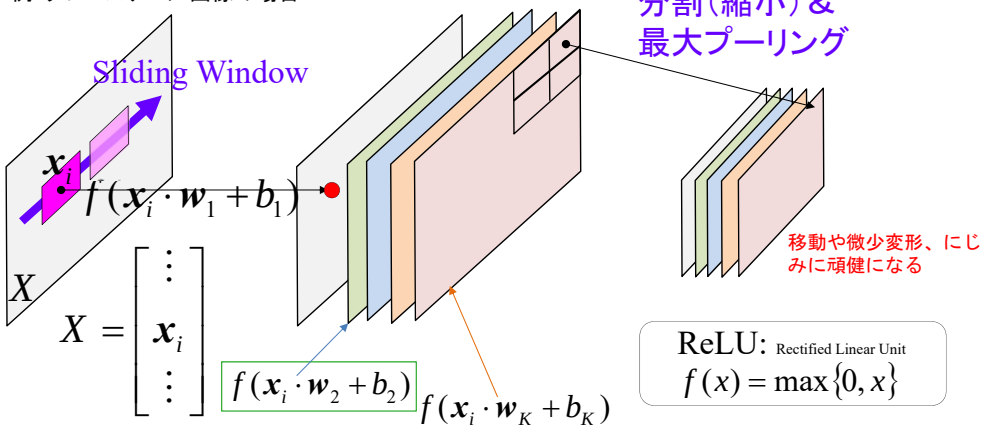
Convolutional Neural Network（CNN）の基本構造

有名なネットワーク
AlexNet [2012]
GoogLe Net [2014]

畳み込み操作とロバスト処理の多段結合が肝

- 要はフィルタバンク $(\boldsymbol{w}_k, k = 1, \ldots K)$
- 入力に近い層で有効

例：グレースケール画像の場合



一部参考　http://cs231n.stanford.edu/slides/winter1516_lecture7.pdf

急がば回れ：　まずは画像へ変換

2017年

DeepVariant：ゲノムのmutation caller（変異検出器）

Creating a universal SNP and small indel variant caller with deep neural networks

https://www.biorxiv.org/content/early/2016/12/21/092890

リファレンス配列にアライメントされたリード配列の積み重なり（pileup）を画像データとして入力し学習させている

結果、非常に高精度の変異（多様性）検出ができるようになっており、生殖細胞系列における変異（germline variant）の検出のコンペティションにおいて賞を受けている.
pileupの画像は、通常の研究現場では、何らかの方法で見つけられた変異の候補を人間が専用のviewerで、本物かどうか確認・検討する際に目にするもの
その画像そのものを変異検出に用いた点がすごいところ

プラント内で既に蓄積されている様々な数値（温度、圧力、流量等）を解析し、現在のデータと比較することによって、プラントの異常の検知や未来の変動の予測を行うシステムを開発　　エッジコンピューティング，IoTでもこのような流れか？

この20年間の内挿手法の劇的な性能向上

線形・非ガウス→スパースモデリング

変数（要素）間の関係　最適化関数のクラス

Tibshirani, R. (1996)

Cand'es, E.J. and Tao, T. (2005)

非線形・ガウス→深層学習

Hinton, G.E. (2006)

線形・非ガウスモデルと非線形・ガウスモデルがマシンで得られる時代になった！

モデリングの技

スパース（疎性を利用した）最適化 1：

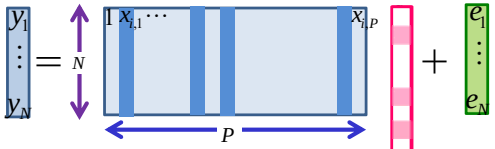
LASSO: Least Absolute Shrinkage and Selection Operator

例：多変量回帰

$$y_i = a_0 + a_1 x_{i,1} + a_2 x_{i,2} + \cdots\cdots + a_P x_{i,P} + e_i \quad (i = 1,\ldots,N)$$

$$\mathbf{y} = X\mathbf{a} + \mathbf{e}$$

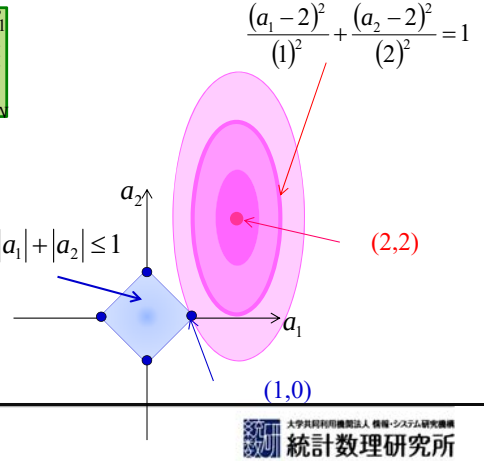
モデルでは表現できない部分（誤差というのは不適切）



新NP問題 $(N << P)$

$$\mathbf{a}* = \min_{\mathbf{a}} \left\{ |\mathbf{y} - X\mathbf{a}|^2 + \lambda |\mathbf{a}| \right\}$$

$$\min_{\mathbf{a}} . |\mathbf{y} - X\mathbf{a}|^2 \text{ subj. to } |\mathbf{a}| \le \alpha$$



スパース最適化 ２：行列、テンソル分解

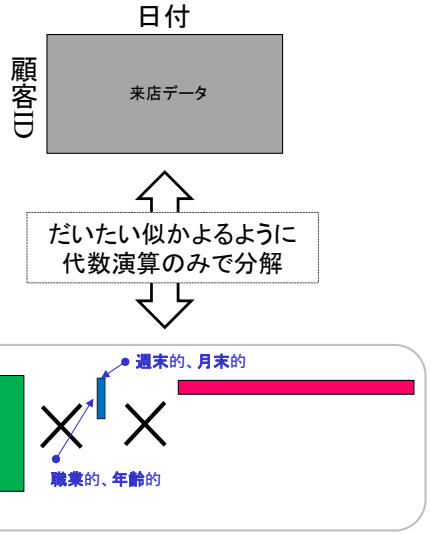
ID付きPOSデータ

i軸：顧客ID

k 軸：日時

$y_{ijk} = 0,1,2\ldots$

j 軸:商品



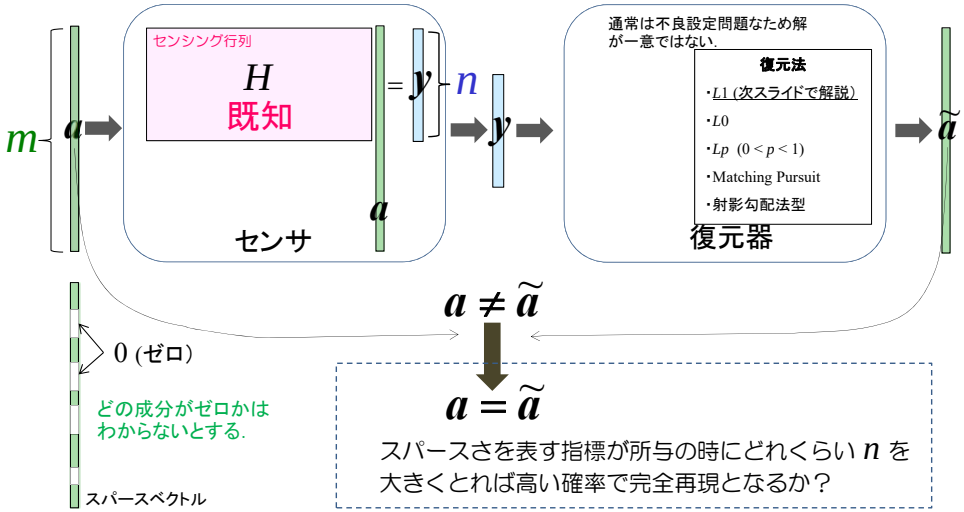
『バラバラを癖で束ねる』を実現

スパース最適化３：

圧縮センシング:－Compressed Sensing－

E.J. Candes and T. Tao (*IEEE IT*, 2006), D.L. Donoho (*IEEE IT*, 2006)※2003年頃から研究が始まる

$y = Ha$　観測データには観測ノイズは含まれていないとする。



$a \neq \tilde{a}$

$a = \tilde{a}$

スパースさを表す指標が所与の時にどれくらい n を大きくとれば高い確率で完全再現となるか？

スパース最適化４：

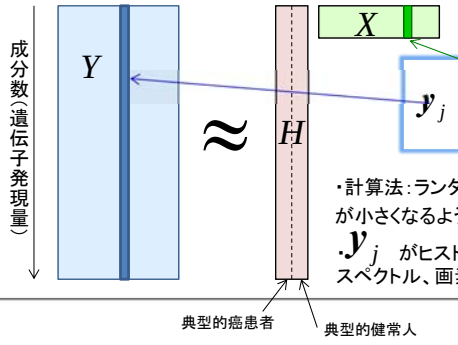
NMF: Non-Negative Matrix Factorization

D. D. Lee and H. S. Seung, (*Nature*, 1999) (NIPS, 2000)

$Y_{m\times n}$ ： **非負値データ**行列。m次元特徴ベクトルをnサンプル分横に並べた行列

$$\min . \left\| Y - H_{m\times r} \cdot X_{r\times n} \right\|^2_{F \text{ or } ID}$$

サンプル数（患者番号）　非負。基底特徴ベクトルを表す　　非負。結合重みを表す



Fの場合、分解結果を確率モデルで表すと

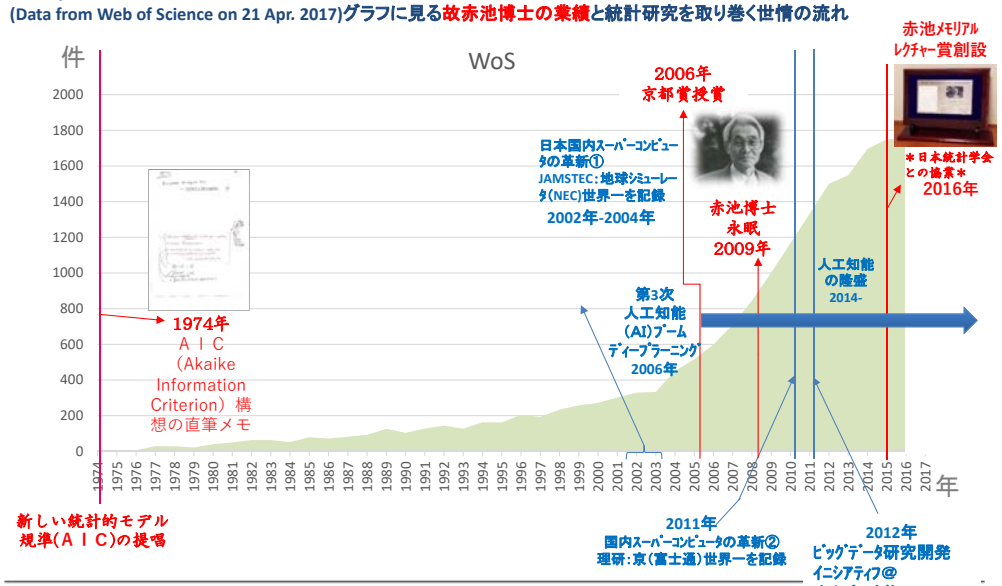
$$y_j = H \cdot x_j + w_j,\ w_j \sim N(0, \sigma^2 I_{m\times m})$$

・計算法：ランダムな要素で埋めたWとHを初期値として、XとWHの差が小さくなるように、WとHを交互に逐次更新。（rの選択が肝）

・y_j　がヒストグラムである事例への応用が多い。他には、パワースペクトル、画素値など。

統計数理の特質 ⇒ 評価されるまでに時間がかかるが、一度評価されると長期間に渡り利用される

Year by Year Citation counts of "NEW LOOK AT STATISTICAL-MODEL IDENTIFICATION" (IEEEac, 1974) の被引用件数
(Data from Web of Science on 21 Apr. 2017)グラフに見る故赤池博士の業績と統計研究を取り巻く世情の流れ



深層学習の強み：人工知能の「度量衡」

長さを計る基準を「度」，体積は「量」，
重さは「衡」と定め，…

2018年 人工知能学会誌　特集「AIとデータ」に招待論文を寄稿

① パラメータ学習がバックプロパゲーションと確率的勾配降下法で統一

② 計算プラットホームが汎化

a. 専用計算機の整備（比較的廉価）
   - GPGPUクラスタ

b. オープン・無償な開発プラットフォームの整備
   - オープンソース：　TensorFlow, Caffe, Keras, Chainer

③ オープンかつリアルタイムな成果公開
   - ✓ 査読プロセスをとおさず、プレプリサーバ（arXiv）にアップ

理論的には第二次AIブームから大きな発展はこれから

Back Propagation

コスト関数

$$E=\sum_n E_n=\sum_n\sum_l\left(y_l^{[n]}-x_l^{(last)}\right)^2$$

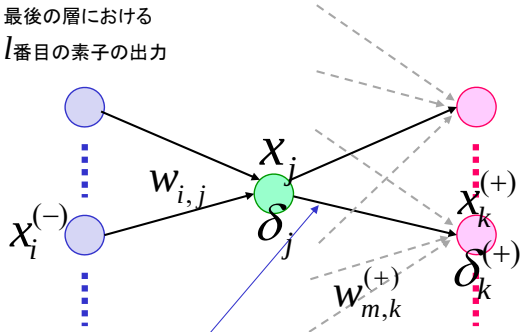
n番目のデータの最後の層におけるl番目の素子の出力

線形結合

$$a_j=\sum_i w_{i,j}x_i^{(-)}+b_j$$

活性化関数

$$x_j=f(a_j)\qquad \frac{\partial a_j}{\partial w_{i,j}}=x_i^{(-)}$$



微分の連鎖率

$$\frac{\partial E_n}{\partial w_{i,j}}=\frac{\partial E_n}{\partial a_j}\cdot\frac{\partial a_j}{\partial w_{i,j}}=x_i^{(-)}\cdot\delta_j\qquad \delta_j=\frac{\partial E_n}{\partial a_j}$$

$$a_k^{(+)}=\sum_m w_{m,k}^{(+)}f(a_m)+b_k^{(+)}$$

$$x_k^{(+)}=f(a_k^{(+)})$$

$$\delta_j=\sum_k\frac{\partial E_n}{\partial a_k^{(+)}}\cdot\frac{\partial a_k^{(+)}}{\partial a_j}=\sum_k\delta_k^{(+)}\cdot w_{j,k}^{(+)}f'(a_j)=f'(a_j)\sum_k\delta_k^{(+)}\cdot w_{j,k}^{(+)}$$

$j\to k$しか残らない。灰色矢印の部分はゼロ。

4 dimensional Variational method

Optimize only for initial state vector

$$p(\boldsymbol{x}_0^{[k]}\mid\boldsymbol{y}_{1:T})\propto p(\boldsymbol{y}_{1:T}\mid\boldsymbol{x}_0^{[k]})p(\boldsymbol{x}_0^{[k]}\mid\boldsymbol{y}_{*:0})$$

$$\boldsymbol{x}_0^{[k+1]}\leftarrow\boldsymbol{x}_0^{[k]}$$

Dimension is huge, so we need an elegant fast calculation scheme

$$\left.\frac{\partial\log p(\boldsymbol{y}_{1:T}\mid\boldsymbol{x}_0)}{\partial\boldsymbol{x}'_0}\right|_{\boldsymbol{x}_0=\boldsymbol{x}_0^{[k]}}$$

Calculation method for efficiently deriving differential vector

$\boldsymbol{x}'_0$ is a transposition of $\boldsymbol{x}_0$

- Chain law of differential

$$\frac{\partial\log p(\boldsymbol{y}_{1:T}\mid\boldsymbol{x}_0)}{\partial\boldsymbol{x}'_0}=\frac{\partial\log p(\boldsymbol{y}_{1:T}\mid\boldsymbol{x}_0)}{\partial\boldsymbol{x}'_T}\cdot\frac{\partial\boldsymbol{x}_T}{\partial\boldsymbol{x}'_{T-1}}\cdot\frac{\partial\boldsymbol{x}_{T-1}}{\partial\boldsymbol{x}'_{T-2}}\cdots\frac{\partial\boldsymbol{x}_1}{\partial\boldsymbol{x}'_0}$$

System model $\boldsymbol{x}_t=f(\boldsymbol{x}_{t-1})$

- Decomposition formula of likelihood function

$$\log p(\boldsymbol{y}_{1:T}\mid\boldsymbol{x}_0)=\sum_{t=1}^{T}\log p(\boldsymbol{y}_t\mid\boldsymbol{y}_{1:t-1},\boldsymbol{x}_0)=\sum_{t=1}^{T}J_t=J_{1:T}\qquad J_t=\log p(\boldsymbol{y}_t\mid\boldsymbol{y}_{1:t-1},\boldsymbol{x}_0)$$

$$J_1=\log p(\boldsymbol{y}_1\mid\boldsymbol{x}_0)$$

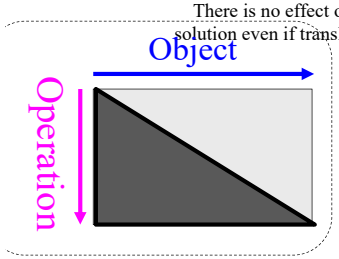
Top-down automatic differentiation→Ajoint method

Back propagation in NN is a special type of top-down automatic differentiation.

$$\frac{\partial J_{1:T}}{\partial\boldsymbol{x}'_0}=\frac{\partial J_0}{\partial\boldsymbol{x}'_0}+\left(\frac{\partial J_1}{\partial\boldsymbol{x}'_1}+\left(\cdots+\left(\frac{\partial J_{T-2}}{\partial\boldsymbol{x}'_{T-2}}+\left(\frac{\partial J_{T-1}}{\partial\boldsymbol{x}'_{T-1}}+\frac{\partial J_T}{\partial\boldsymbol{x}'_T}\frac{\partial\boldsymbol{x}_T}{\partial\boldsymbol{x}'_{T-1}}\right)\frac{\partial\boldsymbol{x}_{T-1}}{\partial\boldsymbol{x}'_{T-2}}\right)\cdots\frac{\partial\boldsymbol{x}_2}{\partial\boldsymbol{x}'_1}\right)\frac{\partial\boldsymbol{x}_1}{\partial\boldsymbol{x}'_0}\right)$$

$$\boldsymbol{d}_T=\frac{\partial J_T}{\partial\boldsymbol{x}'_T},\qquad \boldsymbol{d}_{T-1}=\frac{\partial J_{T-1}}{\partial\boldsymbol{x}'_{T-1}}+\boldsymbol{d}_T\frac{\partial\boldsymbol{x}_T}{\partial\boldsymbol{x}'_{T-1}},\qquad \boldsymbol{d}_{T-2}=\frac{\partial J_{T-2}}{\partial\boldsymbol{x}'_{T-2}}+\boldsymbol{d}_{T-1}\frac{\partial\boldsymbol{x}_{T-1}}{\partial\boldsymbol{x}'_{T-2}}$$

There is no effect on the solution even if translating



$$\boldsymbol{d}_T=\boldsymbol{0},\qquad \boldsymbol{d}_{t-1}=\frac{\partial J_{t-1}}{\partial\boldsymbol{x}'_{t-1}}+\boldsymbol{d}_t\frac{\partial\boldsymbol{x}_t}{\partial\boldsymbol{x}'_{t-1}}\quad(t=T,\ldots,1)$$

Recurrence formula going back in time (Recurrence backward formula)

$$\boldsymbol{d}_0=\frac{\partial J_0}{\partial\boldsymbol{x}'_0}+\boldsymbol{d}_1\frac{\partial\boldsymbol{x}_1}{\partial\boldsymbol{x}'_0}=\frac{\partial J_{1:T}}{\partial\boldsymbol{x}'_0}\qquad J_0=\log p(\boldsymbol{y}_0\mid\boldsymbol{x}_0)=0$$

Updated by descent method $\boldsymbol{x}_0^{[k+1]}=\boldsymbol{x}_0^{[k]}+s\cdot\boldsymbol{d}_0^{[k+1]}\quad(k=0,1,\ldots)$

S: scaling factor

確率的勾配降下法（SGD: Stochastic Gradient Decent）

目的（最適化）関数

$$L(\boldsymbol{w})=\sum_{i=1}^{N}\ell(\boldsymbol{z}_i,\boldsymbol{w})+\psi(\boldsymbol{w})$$

N← 超ビッグ！

データに依存しないペナルティ項 あとで考えれば良い

$\boldsymbol{z}_i\equiv(y_i,\boldsymbol{x}_i)$ ラベル　説明変数ベクトル

パラメータの次元は、億以上

尤度関数部分に相当　事前分布部分に相当

$$\ell(\boldsymbol{z}_i,\boldsymbol{w}):=Q_i(\boldsymbol{w})\qquad \boldsymbol{w}^{(k)}=\boldsymbol{w}^{(k-1)}-\eta\nabla\left(\sum_{i=1}^{N}Q_i(\boldsymbol{w}^{(k)})\right)$$

最急降下法

SGD

$$\boldsymbol{w}^{(k)}=\boldsymbol{w}^{(k-1)}-\eta_k\nabla Q_{i^*}(\boldsymbol{w}^{(k-1)})$$

i^* は、1～Nの中からランダムに選択（乱択）

ミニバッチ：　データ分割法

モーメンタム法

Nestrov加速

$$\tilde{\boldsymbol{w}}^{(k)}=\boldsymbol{w}^{(k-1)}-\eta_k\nabla Q_{i^*}(\boldsymbol{w}^{(k-1)})$$

Yurii Nestero (1983)

$$\boldsymbol{w}^{(k)}=\tilde{\boldsymbol{w}}^{(k)}+\frac{k-2}{k+1}\left(\tilde{\boldsymbol{w}}^{(k)}-\tilde{\boldsymbol{w}}^{(k-1)}\right)$$

https://qiita.com/kenmatsu4/items/d282054ddedbd68fecb0 のデモは分かりやすくて推奨

確率的勾配降下法の改良

機会学習ではオンライン学習と呼ばれる研究分野

確率的分散縮小勾配法（SVRG）

Outer loop： $w^{[m]}$ を更新

Inner loop： $w^{(k)}$ を更新　i^* は、1～Nの中からランダムに選択（乱択）

$$w^{(k)} = w^{(k-1)} - \eta_{m,k}\left[\nabla Q_{i^*}(w^{(k-1)}) + \left\{\nabla Q_{i^*}(w^{[m]}) + \nabla\left(\sum_{i=1}^{N} Q_i(w^{[m]})\right)\right\}\right]$$

アイデア自体はそれほど新規味は感じないが、収束速度の証明は大変

二重加速確率的分散縮小勾配法　理論的には世界最速！（Suzuki）

$$z^{(k)} = \tilde{w}^{(k-1)} + \frac{k-2}{k+1}\left(\tilde{w}^{(k-1)} - \tilde{w}^{(k-2)}\right) + \frac{k-1}{k+2}\left(w^{(k-1)} - \tilde{w}^{(k-2)}\right)$$

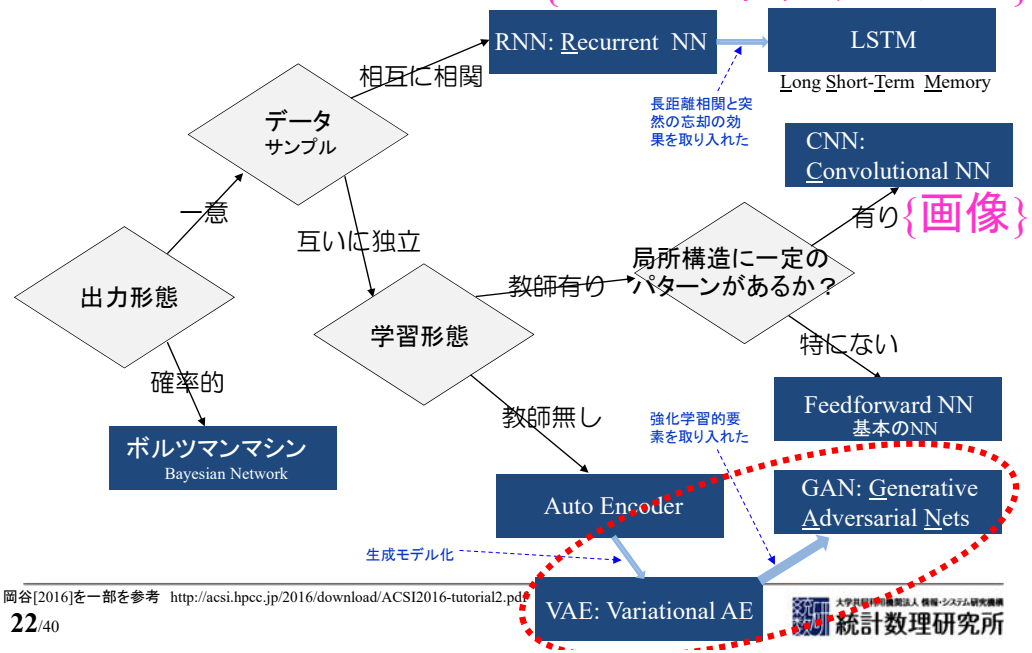
Nestrov加速

$$\left(\tilde{w}^{(k)}, w^{(k)}\right) = \text{inner loop}\left(z^{(k)}, \tilde{w}^{(k-1)}\right)$$

勾配をSVRGと加速法で計算

DNNの分類

{系列データ：音声、テキスト}



岡谷[2016]を一部を参考　http://acsi.hpcc.jp/2016/download/ACSI2016-tutorial2.pdf

生成モデルと識別モデル（関数）

Generative Model vs. **D**iscriminative Model

生成モデル（広義）　$p(y \mid \theta, z)$

データ　パラメータ　説明変数（アスペクト）

クラス分類問題　例：C=1 or 0

生成モデルからベイズの定理を経由して条件付き確率を計算　$p(C = C_i \mid y, \theta) = \dfrac{p(y, C = C_i \mid \theta)}{\sum_i p(y, C = C_i \mid \theta)}$

生成モデル（狭義）　$p(y, C \mid \theta)$　同時分布

データ　クラス

$p(y \mid C, \theta_2) \cdot p(C \mid \theta_1)$

識別モデル　$p(C \mid y, \theta)$　条件付き分布

識別関数：　y を C に写像する関数

統計学の強みは生成モデルの取り扱いに慣れていること

シミュレーションと機械学習の融合

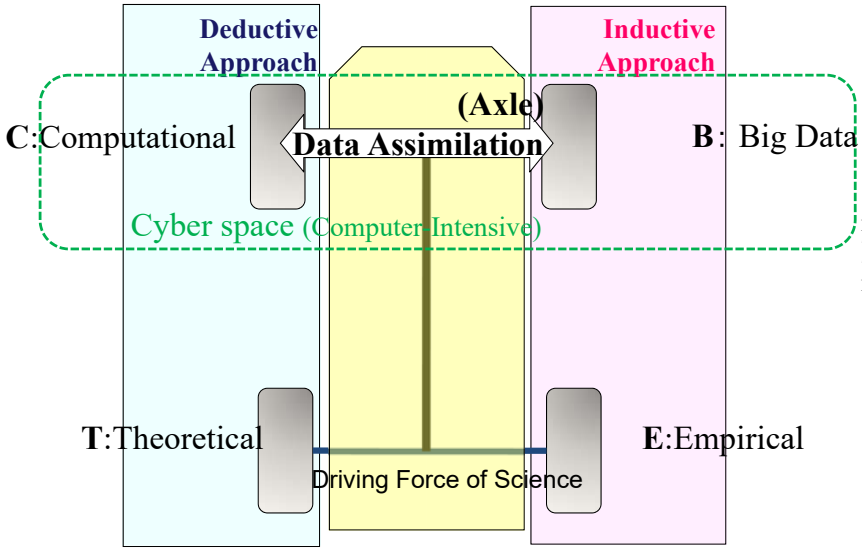
Smart Simulation

① Emulation
- ✓ Sparse modeling

② Neo experimental design
- ✓ Bayesian optimization

③ Automation of making a generative model
- ✓ Reversal formula by Bayese Theorem
- ✓ GDAE,VAE, GAN

Science Paradigms: Four Kind of Methodology of Science



By *Tony Hey, Stewart Tansley, and Kristin Tolle, Eds.* Microsoft Research, 2009.

統計数理研究所

① Emulation 気候変動の確率地図

（上野玄太）
文科省創生プログラム「気候変動予測データの統計学的解析手法の開発」

複数の全球気候モデルのシミュレーション結果を用い、気候変動を確率的に評価する手法を開発し、東アジア及び日本における気温の将来変化の確率地図を初めて作成した。

21世紀後半の気温上昇の確率地図

観測値　シミュレーション

スパース回帰モデル

$$J\left(\beta_0,\beta_1,\cdots,\beta_N\right)=\sum_{t=1}^{T}\left(y_t-\beta_0-\sum_{n=1}^{N}\beta_n x_t^{(n)}\right)^2+\gamma\left[(1-\alpha)\sum_{n=0}^{N}\beta_n^2+\alpha\sum_{n=0}^{N}\left|\beta_n\right|\right]$$

(例) 東京の気温
白: 観測値
灰: シミュレーション
緑: アンサンブル平均
赤: 推定した確率分布

2℃以上昇温する確率

東アジア及び日本の多くの地域で2度以上月平均気温が上昇する確率が70～80%以上。

統計数理研究所

① Emulation エミュレーション：スパースモデリング

地球科学では

Simulation Result（ 10^4～10^7）

Climate Prediction：Climate **Emulator**

For given $x_{0,i}$

Sample

$$\left\{\left\{x_t\right\}_{t=1}^{T},\ x_{0,i}\right\}_{i=1}^{N}$$

Scalar Variable: 50年後の地球全体の平均気温

Sparse regression

$$y_i = H x_{0,i} + e_i$$

Kernel Regression, Deep Learning, ..

Sparse Learning

0

Prediction

$$x_i^{+} \Rightarrow y_i^{+}$$

New input from a simulation system

物質科学では

Data production

$$\hat{\mathbf{H}}\cdot\mathrm{Y}(\mathbf{r})=\mathbf{E}\cdot\mathrm{Y}(\mathbf{r})$$

Schrödinger equation

$$\longrightarrow \quad \left\{Y_i,S_i \middle| i=1,\cdots,n\right\} \quad \longrightarrow$$

Structure-property data

Machine Learning

$$Y_s=\hat{f}(S)$$

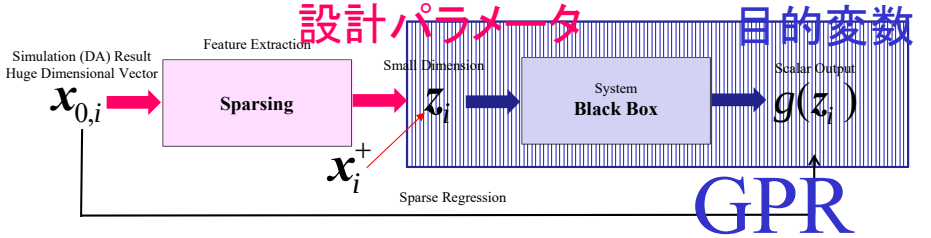
シミュレータのコモディティ化

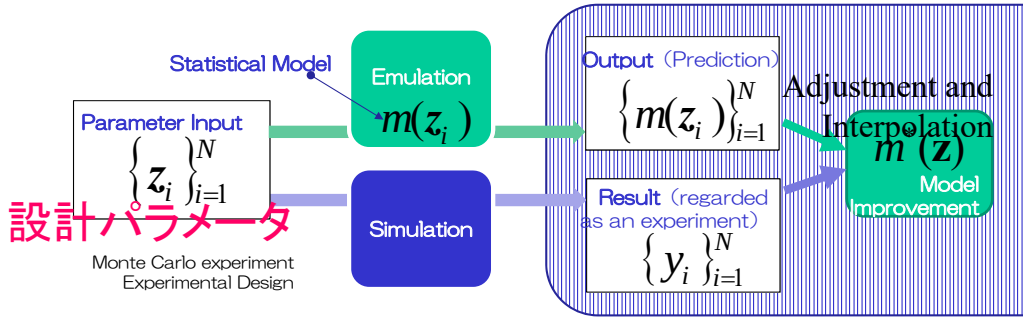
統計数理研究所

Emulator and Emulation

② Bayesian optimization

Emulate an output of simulation given parameters





統計数理研究所

ベイズの反転公式

③ a. Reversal formula by Bayese Theorem

x：興味のある対象

y：データ

順解析：シミュレーション等

ベイズの反転公式

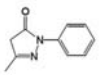
$$p(x \mid y) = \frac{p(y \mid x)p(x)}{\sum p(y \mid x)p(x)}$$

逆解析

ベイズの定理。等号の右側と左側で，赤と青で示した変数部分の縦棒との相対関係が反転していることがわかる。この事実により，ベイズの反転公式と呼ばれる。

非効率（現実には機能しない）探索法

構造 x



$f(x)$ 物理あるいは化学的指数や係数、特性の計算値

機能発現 y その（実験値あるいは経験値）データ

膨大な数のモンテカルロ計算

x の候補を恣意的に多数発生

一つ一つの x に対して順解析（MDなどの第一原理計算）を実施。データとの整合性 $p(y \mid f(x))$ を計算

エキスパートが目と勘で判断

Data Science Center for Creative Design and Manufacturing

established in July, 2017

Prof. Yoshida

■ Virtual screening, QSAR modeling: P(Y|G) **Forward Problem**

Quantitative Structure-Activity(Affinity) Relationship

Develop statistical models to predict biochemical or physiochemical activities Y of an input chemical structure G

■ Chemical design, Inverse-QSAR: P(G|Y=y) **Inverse Problem**

Generate novel chemical structures G achieving desired activities Y=y

Preimage reconstruction of the graph kernel using a MCMC algorithm

（Thanks to Prof. Yoshida; 吉田@統数研から借用、改変）

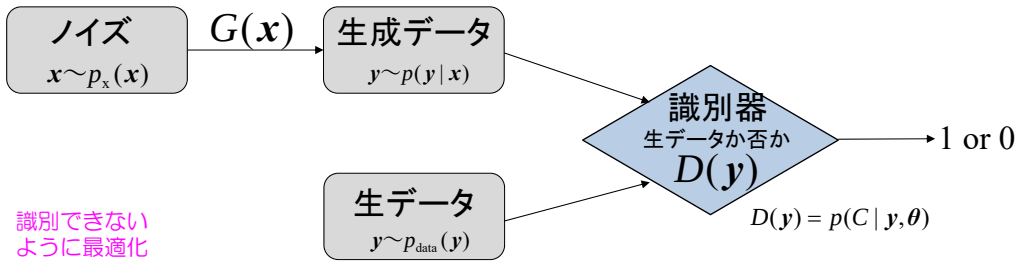
GAN（敵対的生成ネットワーク）の基本アルゴリズム

Generative Adversarial Nets: Goodfellow, Ian *et al*, Advances in Neural Information Processing Systems 27, 2014年

エミュレータ

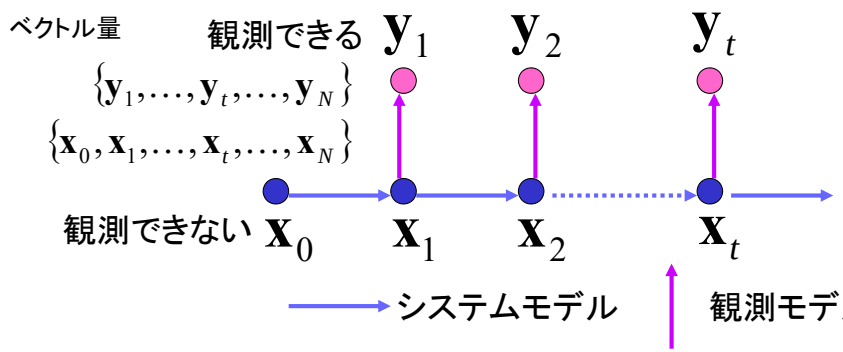
$D(y)$：yが学習データである確率を与える写像（関数）。sigmoidを利用

$G(x)$：ノイズあるいは潜在変数xからデータ空間yへの写像（関数）



Discriminative NetworkはGenerative Networkの学習で使うときは訓練しない

Chain Structure Graphical Model

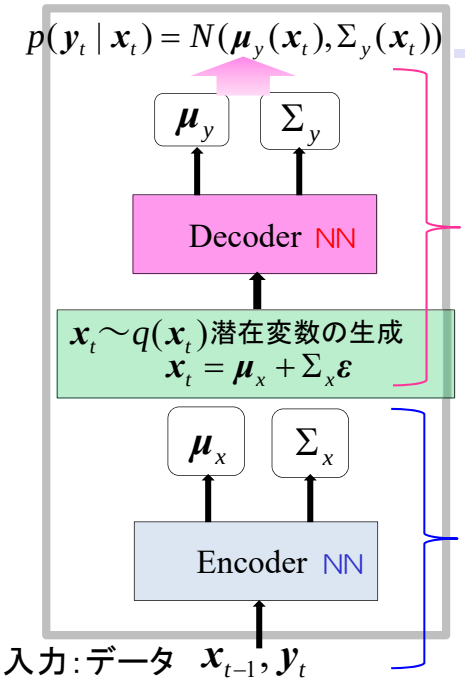


一般型

$$p(\mathbf{y}_{1:T-1},\mathbf{y}_T,\mathbf{x}_{1:T})=\prod_{t=1}^{T}p(\mathbf{y}_t|\mathbf{y}_{1:t-1},\mathbf{x}_{1:t})\cdot p(\mathbf{x}_t|\mathbf{y}_{1:t-1},\mathbf{x}_{1:t-1})$$

$$\Rightarrow\prod_{t=1}^{T}p(\mathbf{y}_t|\mathbf{x}_t)\cdot p(\mathbf{x}_t|\mathbf{x}_{t-1})$$

VAEと粒子フィルタの融合



私のアイデア

Observation model

$$p_t(\mathbf{y}_t|\mathbf{x}_t,\mathbf{x}_{1:t-1},\mathbf{y}_{1:t-1})$$
確率分布のモデル化
$$\equiv N(\boldsymbol{\mu}_y(\mathbf{x}_{1:t},\mathbf{y}_{1:t-1}),\Sigma_y(\mathbf{x}_{1:t},\mathbf{y}_{1:t-1}))$$
$$\approx N(\boldsymbol{\mu}_y(\mathbf{x}_t),\Sigma_y(\mathbf{x}_t))$$
複雑なDNN利用可

$$\boldsymbol{\varepsilon}\sim N(0,I)$$

LSTMの利用も有望

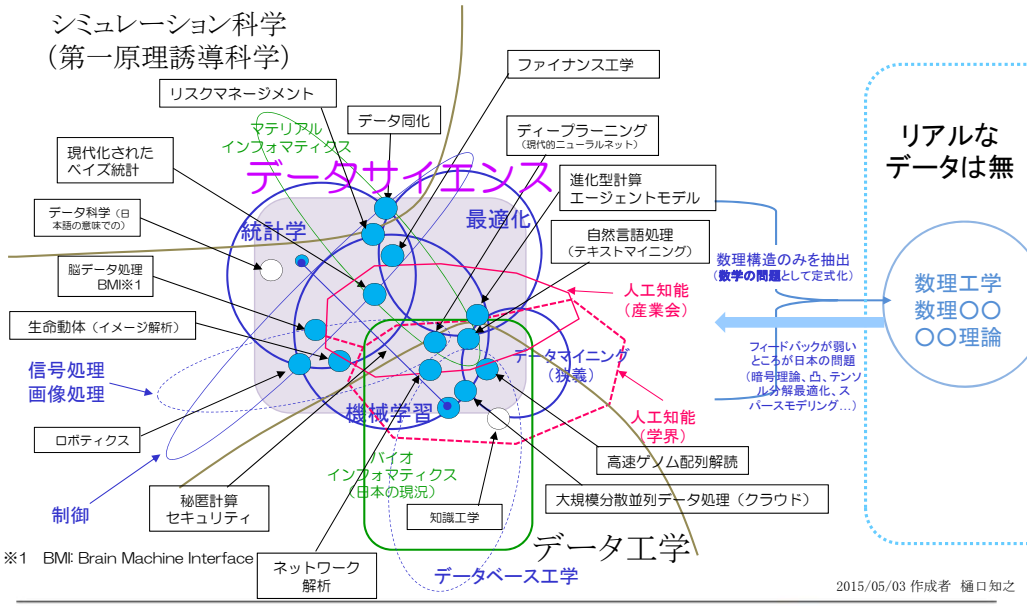
Proporsal distribution

$$q_t(\mathbf{x}_t|\mathbf{x}_{1:t-1},\mathbf{y}_{1:t})$$
確率分布のモデル化
$$\equiv N(\boldsymbol{\mu}_x(\mathbf{x}_{1:t-1},\mathbf{y}_{1:t}),\Sigma_x(\mathbf{x}_{1:t-1},\mathbf{y}_{1:t}))$$
$$\approx N(\boldsymbol{\mu}_x(\mathbf{x}_{t-1},\mathbf{y}_t),\Sigma_x(\mathbf{x}_{t-1},\mathbf{y}_t))$$
複雑なDNN利用可

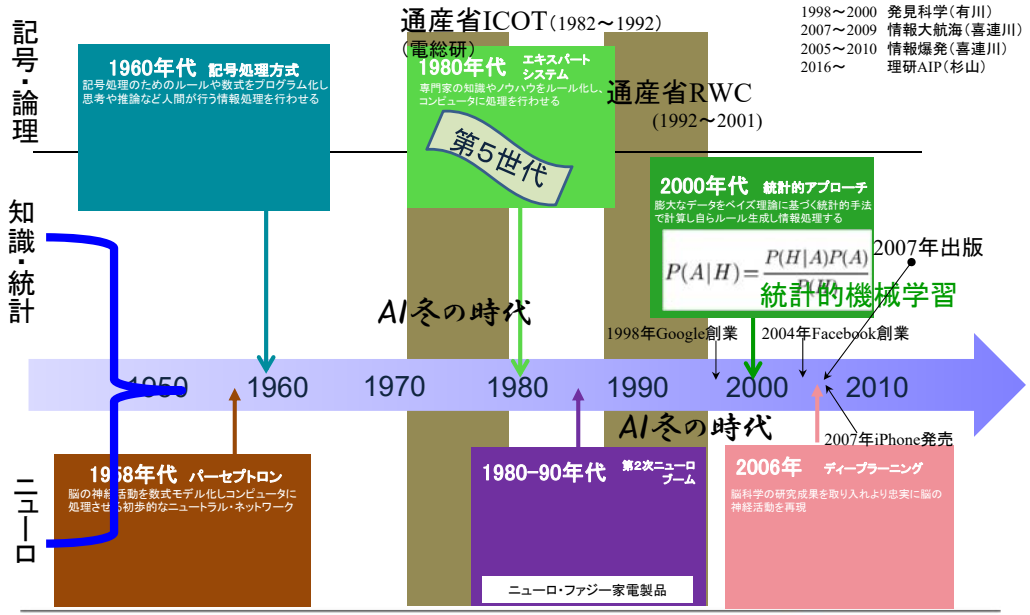
1980年代数理・情報技術の再興

- ニューラルネットワーク→深層学習
- 情報統計力学
- 最適化
  - ✓ イジングモデル→量子アニーリング
  - ✓ シミュレイティドアニーリング
  - 遺伝的アルゴリズム、遺伝的プログラミング→超並列計算による大規模離散最適化
  - 連続最適化(降下法)→確率的勾配降下法、オンライン学習
- 制約付き最適化→スパースモデリング
- 次元圧縮→行列、テンソル分解

データに関連した数理技術の俯瞰図



人工知能研究および研究開発プロジェクトの歴史



＊http://blogs.itmedia.co.jp/itsolutionjuku/2015/07/post_105.html を参考として作成
＊統数研・福水教授のスライド内容を一部借用

再掲： 統計学は対応できているのか？

- 計算量に配慮しているか？：計算量が $O(N)$ では役に立たない
- p次元空間内の分布の表現をどうするのか？
- オンライン学習（推定）法の研究は？
- 統計的モデリングはエキスパートがやるのか？：深層学習とスパースモデリングの勃興
- 可視化技術は人が見るため？：深層学習（とくにCNN：Convolutional NN）の画像処理における圧倒的性能
- 計算プラットフォームの整備は？
- 目的特化型計算機への実装は？

何が肝か？

薄く幅広く習得するのがよい。引き出しをたくさん持っておくこと。

理由： Webでいくらでも自学自習できる。実務でもしっかり体得できる。
「厳密・唯一」から、「近似・アンサンブル」の世界観を身につけさせる。

- 高次元
- 超大サンプル数 （オンライン）
- 非線形
- 最適化
- シミュレーション
- プログラミング

- ビッグデータをベースに、従来、分断されていた教育内容を整理・融合
- 学部前期教育もそれに向けた数学の基礎とデータリテラシーに

人生の分水嶺

ロッキー山脈越え@2006年

- 基礎的な数学を修得するかどうかで、人生の分水嶺は決まる
- 『仕組み』をつくるものが勝利を得る世界
- そこに資するのは数学とデザインセンス！