東京大学・大学院新領域創成科学研究科
基盤科学研究系
高次元データ駆動科学教育プログラム
データ駆動科学入門I No1/2

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

自己紹介(理論物理学)

- 大阪市立大学理学部物理学科 (1981 - 1985)
  - アモルファスシリンコンの成長と構造解析
- 大阪大学大学院理学研究科(金森研) (1985 – 1987)
  - 希土類元素の光励起スペクトルの理論
- 三菱電機 北伊丹製作所(量産エンジニア) (1987 - 1989)
  - 化合物半導体(半導体レーザー)のエピタキシャル結晶成長
- 大阪大学大学院基礎工学研究科生物工学 (1989 - 1996)
  - ニューラルネットワーク(脳型人工知能)
  - 福島先生は畳み込み深層ニューラルネットワークの提案者
- JST ERATO 川人学習動態脳プロジェクト (1996 - 2001)
  - 計算論的神経科学
- 理化学研究所 脳科学総合研究センター 甘利チーム (2001 - 04/06)
  - 情報統計力学
  - ベイズ推論, 機械学習, データ駆動科学
- 東京大学・大学院新領域創成科学研究科 複雑理工学専攻
  - データ駆動科学、ベイズ推論、スパースモデリング (2004/07 –)

内容(1/3)

- データ駆動科学入門Iの紹介とゴール
  線形回帰$y=ax+b$の解析計算とスペクトル分解
- 高次元次元データ駆動科学教育プログラムの紹介
  プログラムの目的と設計指針
- 自然科学とデータ駆動科学
  自然記述の二つの戦略:
    要素還元主義と階層的自然観
    階層的自然観とデータ駆動科学
- データ駆動科学の二つの情報数理基盤
  スパースモデリング(SpM)
  ベイズ推論からベイズ計測へ

内容(2/3)

- 計算論的神経科学とデータ駆動科学
  David Marrの三つのレベル
  データ駆動科学の三つのレベル
  連立方程式とデータ駆動科学
- 直線回帰$y=ax+b$の解析手計算
  - 最小二乗法の復習
  - ベイズ推論の導入とベイズ計測の創成
  - パラメータの確率分布、ノイズ分散推定、モデル選択

内容(3/3)

- 阪大理物修士課程での研究からベイズ計測へ
- スペクトル分解
  従来手法の破綻
  ベイズ計測の導入
  レプリカ交換法の適用
  モデル選択
  計測限界の理論的取り扱い
- ベイズ計測の展開: 多様な計測への系統的展開
  NMR、メスバウアー分光、小角散乱、比熱と磁化率のベイズ統合、XPSとXASのベイズ統合、ベイズ計測の適用例
- ベイズ計測の展開
  - Spring-8全ビームラインベイズ化計画
- まとめ今後の展望

内容(1/3)

- データ駆動科学入門Iの紹介とゴール
  線形回帰$y=ax+b$の解析計算とスペクトル分解
- 高次元次元データ駆動科学教育プログラムの紹介
  プログラムの目的と設計指針
- 自然科学とデータ駆動科学
  自然記述の二つの戦略:
  - 要素還元主義と階層的自然観
  - 階層的自然観とデータ駆動科学
- データ駆動科学の二つの情報数理基盤
  スパースモデリング(SpM)
  ベイズ推論からベイズ計測へ

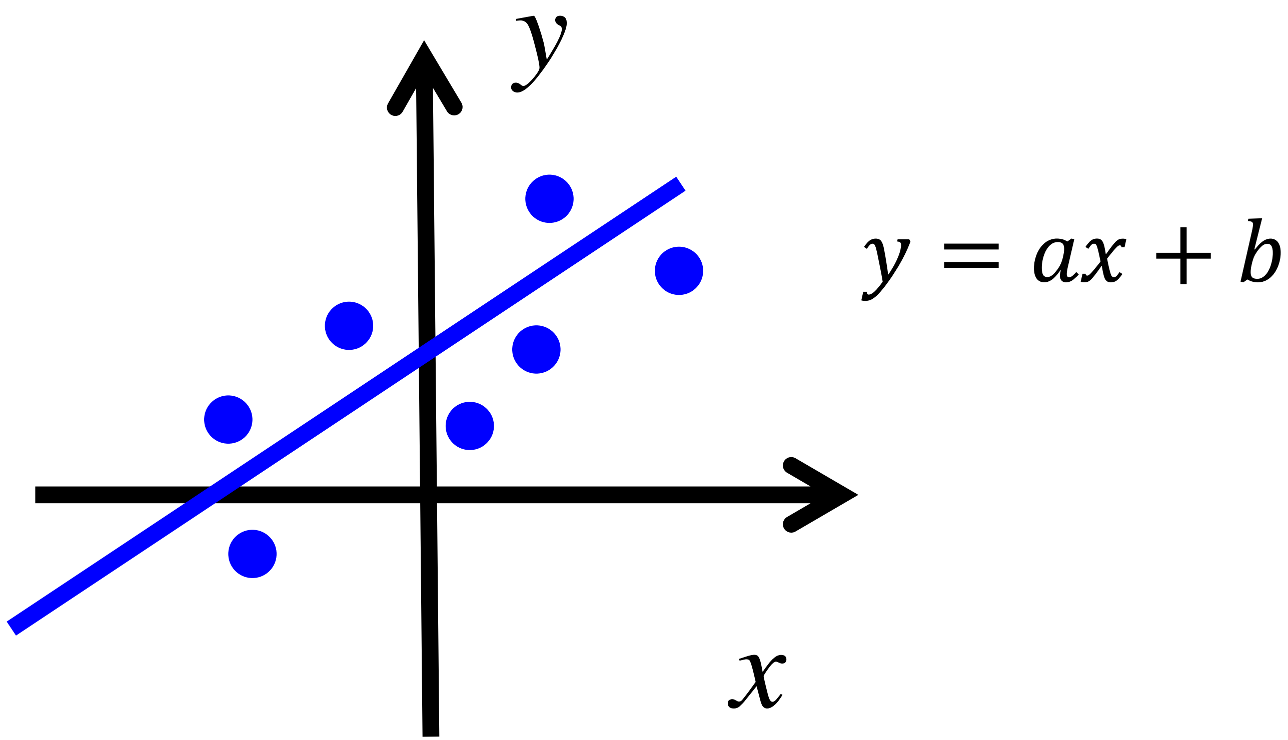
データ駆動科学入門Iの紹介とゴール

- 線形回帰$y=ax+b$の解析計算とスペクトル分解を二日間で講義する。
- データ駆動科学とはなにか? 機械学習や通常の統計学とは何が違うかなどの背景の理解を重視。これにより、聴講者は、単なる計算技術やプログラミング技術の習得だけでなく、得てして実験の付加物/ポストプロセスと考えられてきたデータ解析を学問まで昇華し、脳の中にデータ駆動科学の有機的な体系を構築することができ、あらゆる状況に対応できるフレキシブルな知性を獲得できる。

y=ax+bのベイズ計測 (1/4)

現実の実験/計測で用いられている最も簡単な例

ベイズ計測で、精密な解析的理論が提案されている稀な系

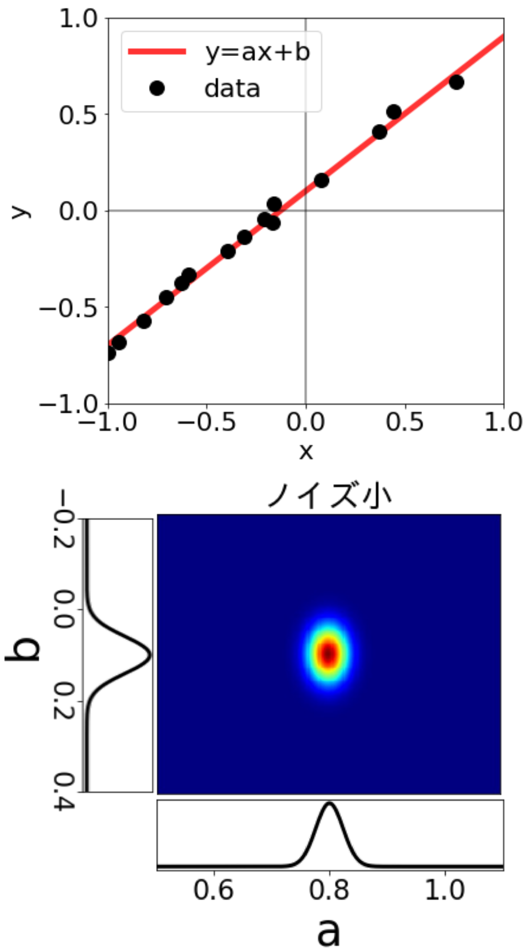


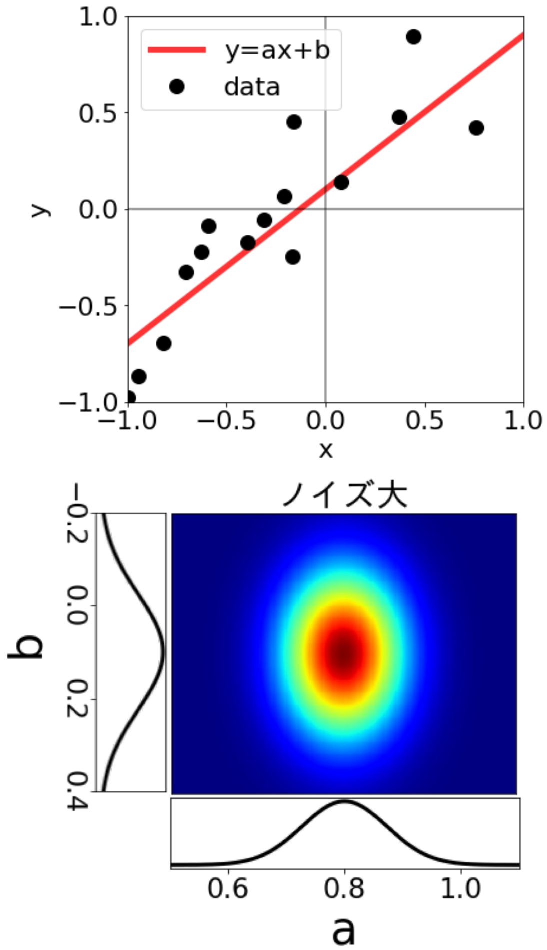
傾き a： 系の線形応答、バネ定数、電気伝導度、誘電率

Katakami, Kashiwamura, Nagata, Mizumaki and Okada
Mesoscopic Bayesian Inference by Solvable Models
https://arxiv.org/abs/2406.02869

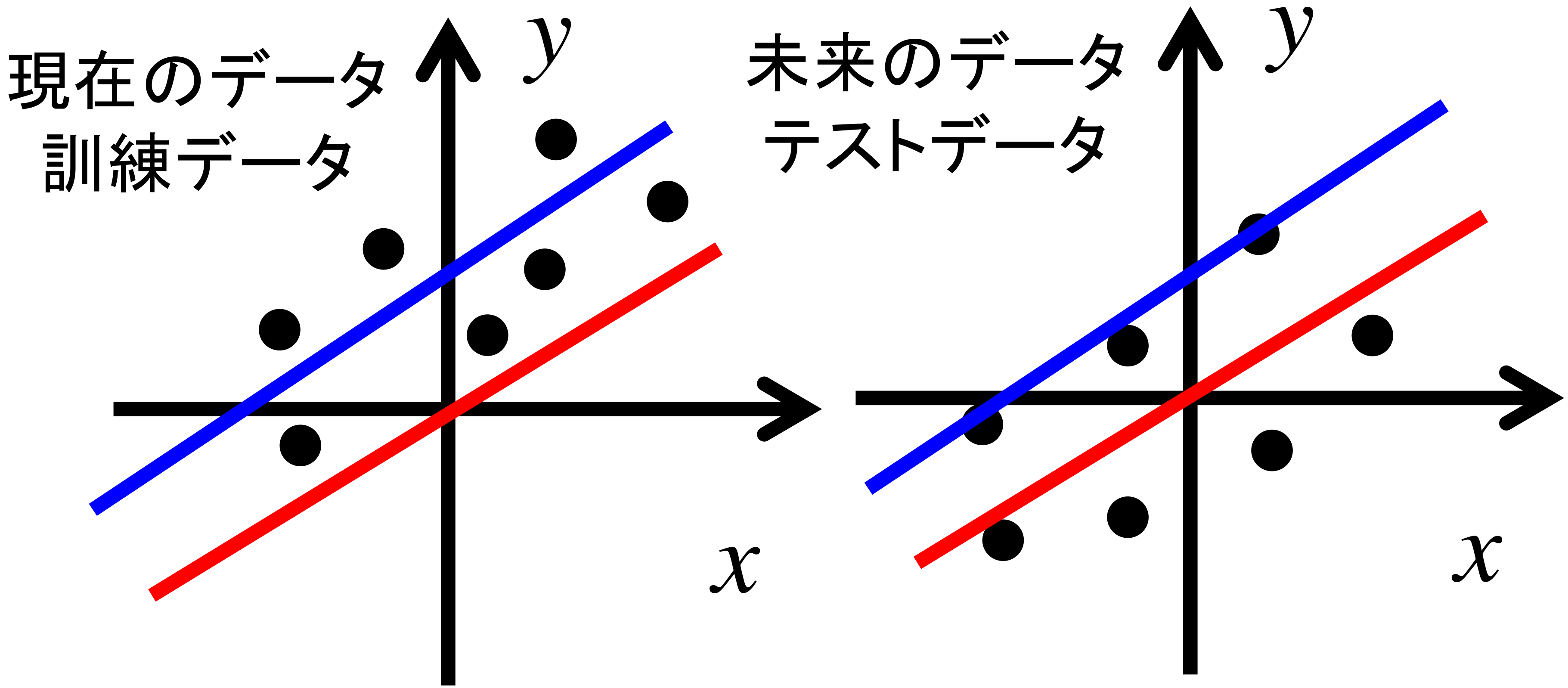
$y=ax+b$のベイズ計測(2/4)

結論: 神器1 パラメータの事後確率推定





y=ax+bのベイズ計測(3/4)



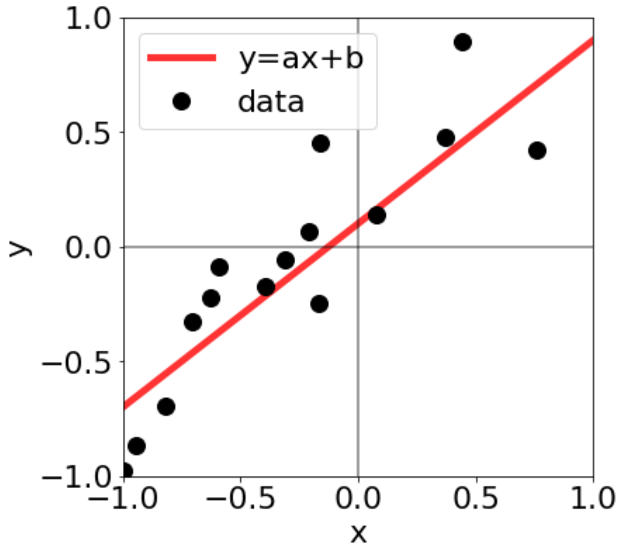
y=ax+b: 訓練誤差小　　y=ax: 訓練誤差小

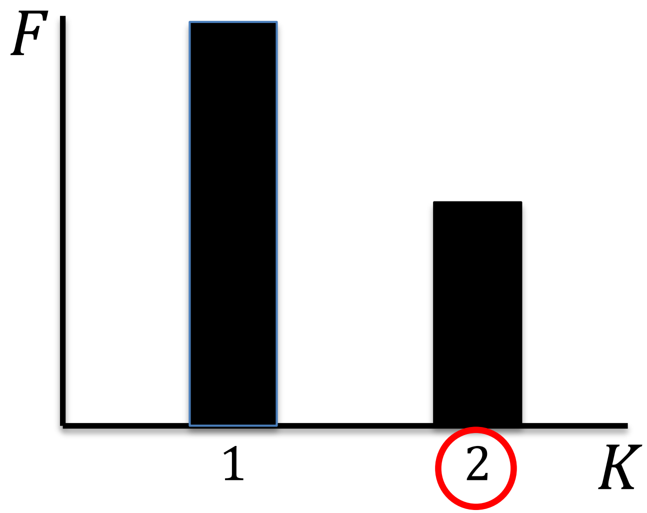
訓練誤差　　汎化誤差

ノイズに過学習

モデル選択できる理由:汎化誤差は観測ノイズに依存する

y=ax+bのベイズ計測(4/4)





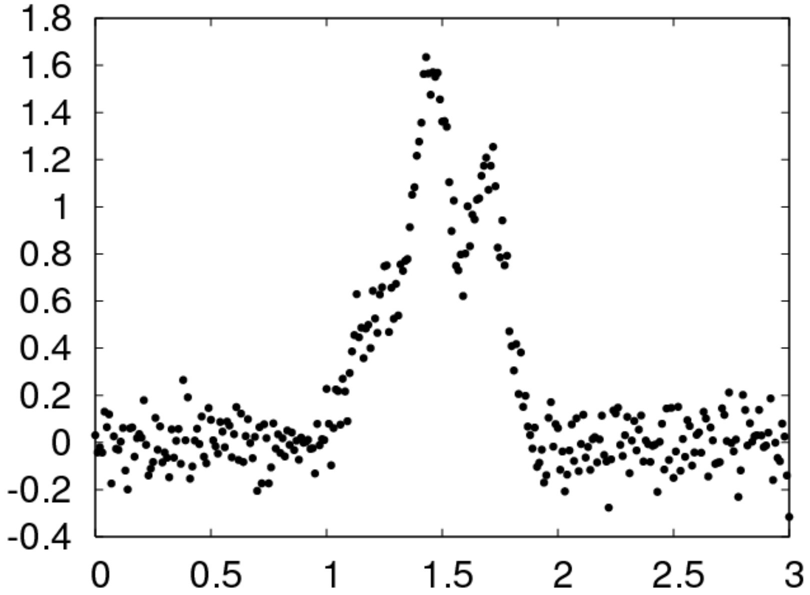
- $K = 1 : y = ax$
- $K = 2 : y = ax + b$

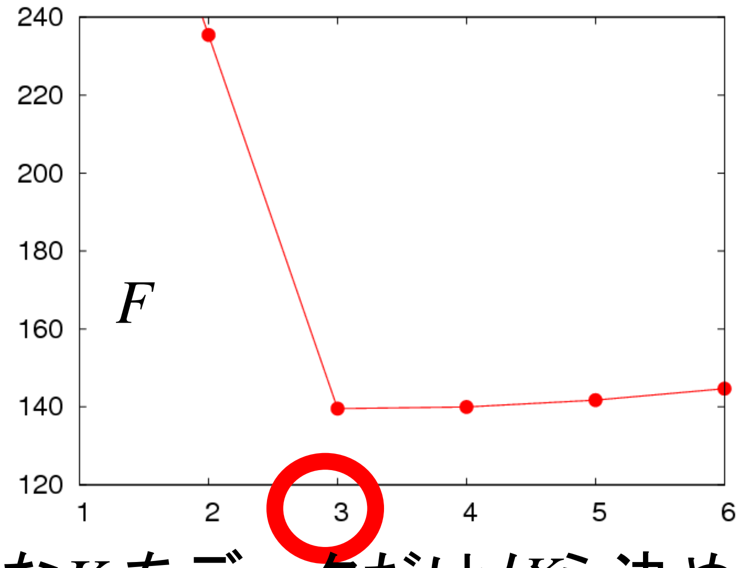
$$F(K=1) = N\left\{\frac{1}{\sigma^2}E(a_0) + \frac{\log N}{2N}\right\}$$

$$F(K=2) = N\left\{\frac{1}{\sigma^2}E(a_0, b_0) + \frac{\log N}{N}\right\}$$

データのみからモデルを選択できる

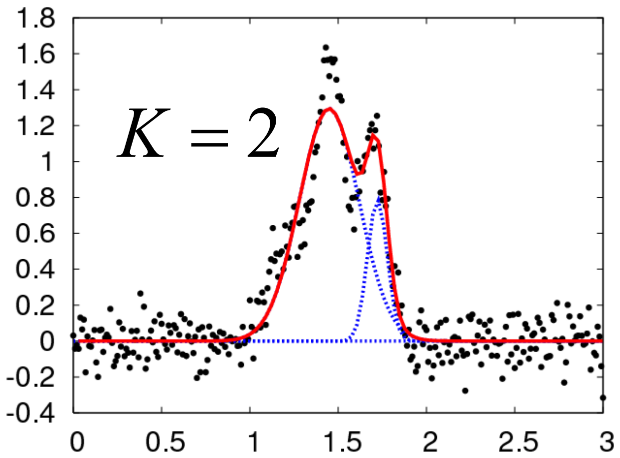
スペクトル分解

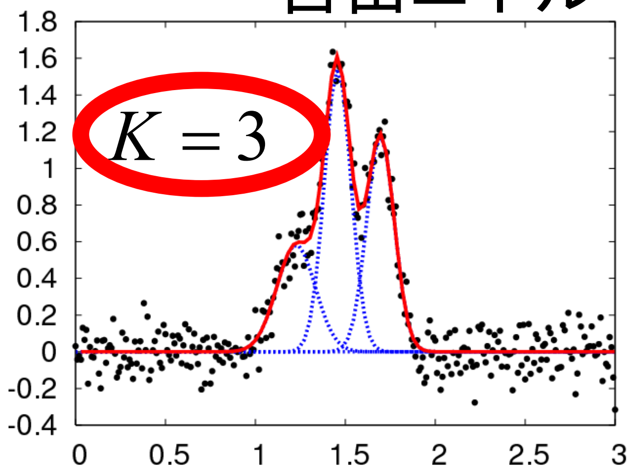


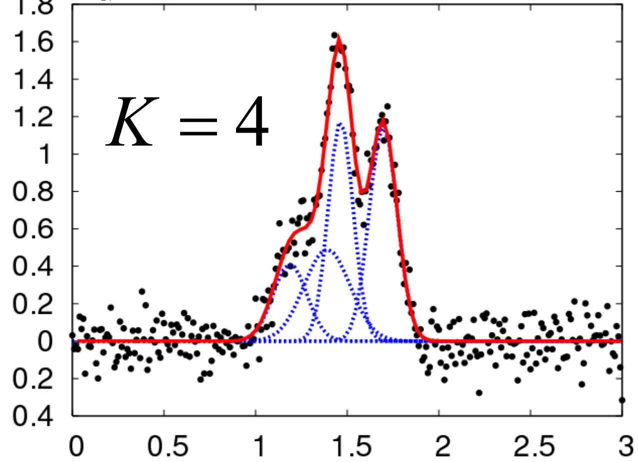


最適なKをデータだけから決める

自由エネルギー最小化



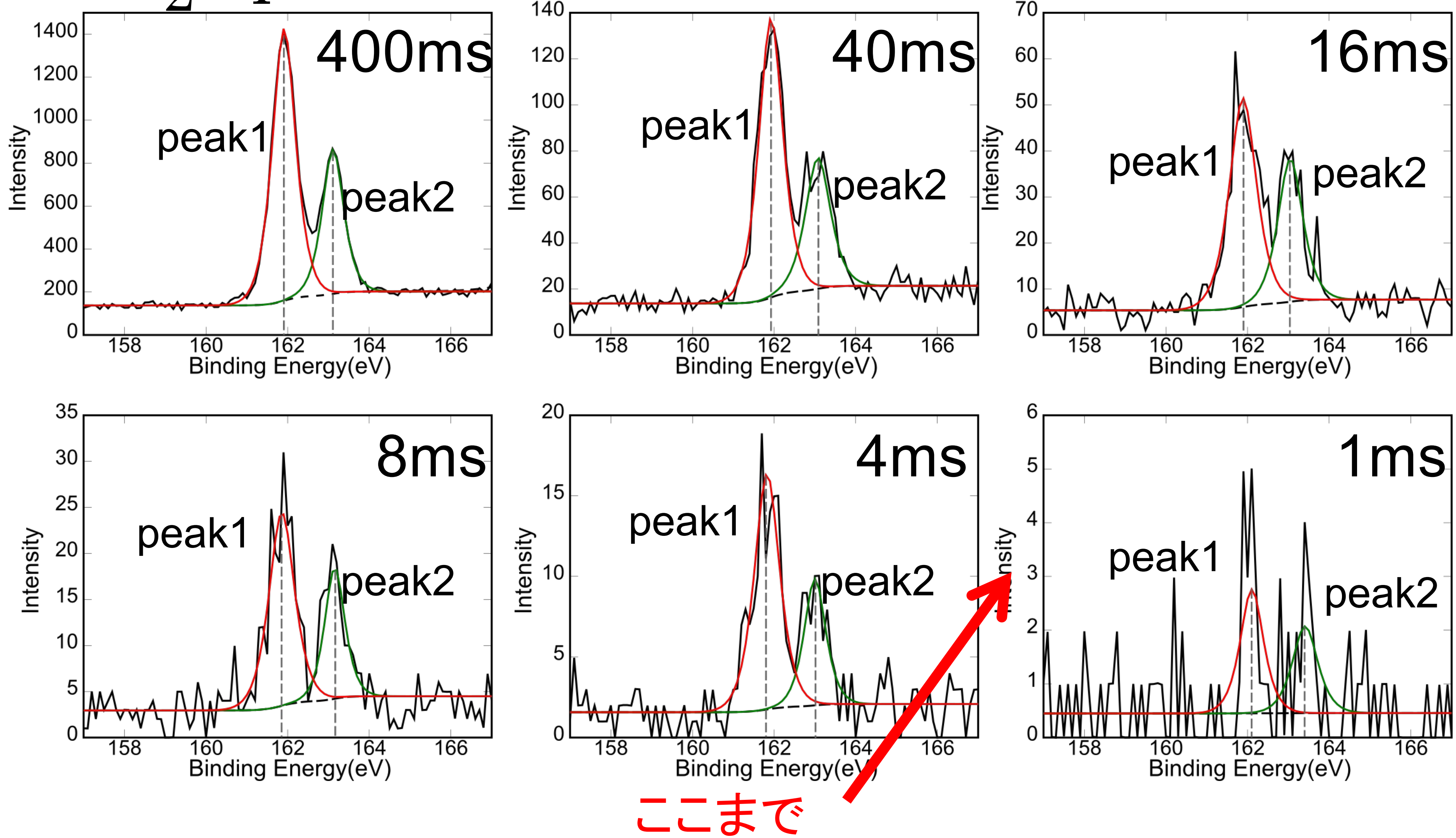




Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

スペクトル分解の計測限界

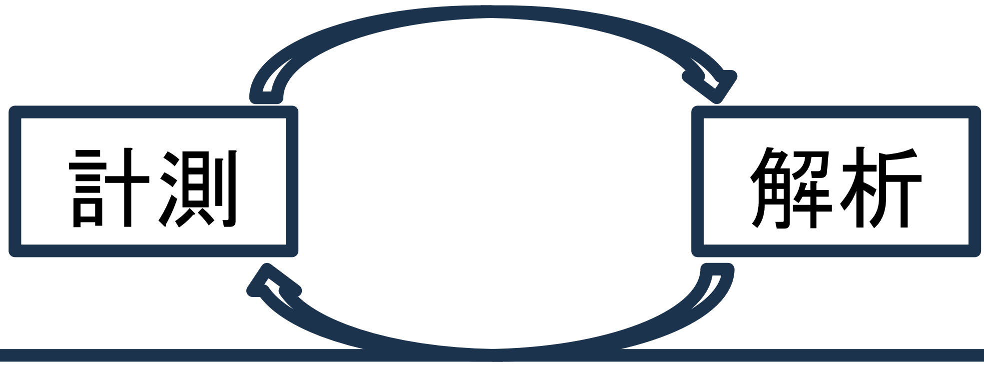
MoS_2 2p



ベイズ計測による
計測と解析の双方向相互作用

従来法: ポストプロセスとしての解析

提案法: 計測と解析の双方向的相互作用



内容(1/3)

- データ駆動科学入門Iの紹介とゴール
  線形回帰$y=ax+b$の解析計算とスペクトル分解
- 高次元次元データ駆動科学教育プログラムの紹介
  プログラムの目的と設計指針
- 自然科学とデータ駆動科学
  自然記述の二つの戦略:
    要素還元主義と階層的自然観
    階層的自然観とデータ駆動科学
- データ駆動科学の二つの情報数理基盤
  スパースモデリング(SpM)
  ベイズ推論からベイズ計測へ

東京大学　大学院新領域創成科学研究科
基盤科学系
高次元データ駆動科学教育プログラム

- 計測、計算、解析のミニマムから構成される、集中講義と既存の専攻の該当科目から構成
- 計測、計算、解析に対して、各2科目1単位で二日間の集中講義から構成される。
- 専攻の既存のカリキュラムと干渉しないように、夏休みと冬休みに集中講義を行う。
- 専攻の既存の講義を含めた所定の単位を取得することにより、研究科長からプログラム修了書をもらえる。

一芸と多芸

- 一芸に秀でる者は多芸に通ず
- Try to learn something about everything and everything about something. (Thomas H. Huxley)
- 民間企業のキャリアパス

  半導体の平社員→半導体の部長(1/20)→洗濯機事業部の部長(1/10)
- でもHuxleyはeverythingを先に書いている?
- 実は一芸と多芸を双方向的に習得すると効率的かも
- 1/200じゃない人を1/200にする秘策!!かもしれない. . .

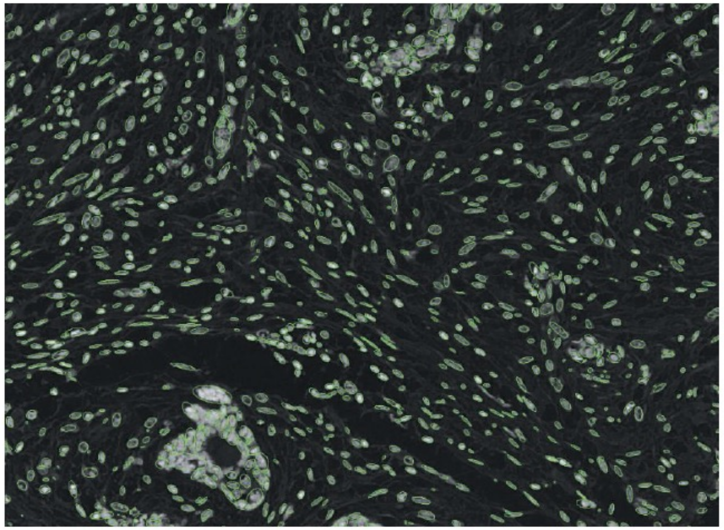
一芸に秀でる ⇔ 多芸を俯瞰し普遍性を抽出

研究室活動 HD3

内容(1/3)

- データ駆動科学入門Iの紹介とゴール
  線形回帰$y=ax+b$の解析計算とスペクトル分解
- 高次元次元データ駆動科学教育プログラムの紹介
  プログラムの目的と設計指針
- 自然科学とデータ駆動科学
  自然記述の二つの戦略:
    要素還元主義と階層的自然観
    階層的自然観とデータ駆動科学
- データ駆動科学の二つの情報数理基盤
  スパースモデリング(SpM)
  ベイズ推論からベイズ計測へ

天文学における高次元データ解析手法が，全く対象とスケールの異なる生命科学でも有効に働く



NEWS

Is There an Astronomer in the House?

With biomedical researchers analyzing stars and astronomers tackling cancer, two unlikely collaborations creatively solve data problems

pathologists look for different biomarkers—specific proteins—in the patient's cancerous

[*Science*, Feb. 2011]

と単純に喜んで良いのか?!　⇒　必要なことは

多様な視点の導入による革新的展開

普遍的な視点による分野を越えたアナロジー／普遍性への探究心

普遍的な原理にもとづく新しい解析法の発展

データ駆動科学とは

- 機械学習などの人工知能を使い，各学問分野の問題を解いていくというアプローチ
- 実験/計測/計算データの背後にある潜在的構造の抽出に関して，データが対象とする学問に依存しない普遍的な学問体系
- 同じアルゴリズムがスケールや対象を超えて，有用であることが多いという経験的事実を背景として，その理由を問い，背後にある普遍性から，データ解析自体を学問的対象とする枠組み.

要素還元主義と階層的自然観

- 要素還元主義
  - 物理学の法則の獲得をよりミクロな方向に進め、得られた基礎方程式からの演繹で、世界を理解する言う立場
- 階層的自然観
  - 自然を各階層に分離して、階層ごとに研究することで、世界を理解する立場
- 以下の状況から、階層的自然観が正しいと言わざるを得ない
  - 第一原理計算の実情
  - 愛情などの高度な認知機能の解明

要素還元主義の破綻

- 要素還元主義: ミクロな基礎方程式からの演繹で、自然を理解できるという立場
- 要素還元主義はとても美しい立場であるが、これまでの科学の以下の発展の事例から考えて、要素還元主義は破綻していると言わざるを得ない。
- 事例1: 量子力学と電磁気学で第一原理的な方程式が書き下せる物性物理学において、強相関電子系などの多対効果が無視できない系においては、計算機を使った第一原理計算でも、系の性質を記述することができないという事実がある。
- このような場合は、実験計測使って、系の性質を決めるパラメータをデータ駆動的に決めざるを得ない。
- 事例2: 例えば恋愛感情のような高度な認知機能のメカニズが量子力学や電磁気学から演繹できるとは誰も思わない。

内容(1/3)

- データ駆動科学入門Iの紹介とゴール
  線形回帰$y=ax+b$の解析計算とスペクトル分解
- 高次元次元データ駆動科学教育プログラムの紹介
  プログラムの目的と設計指針
- 自然科学とデータ駆動科学
  自然記述の二つの戦略:
    要素還元主義と階層的自然観
    階層的自然観とデータ駆動科学
- データ駆動科学の二つの情報数理基盤
  スパースモデリング(SpM)
  ベイズ推論からベイズ計測へ

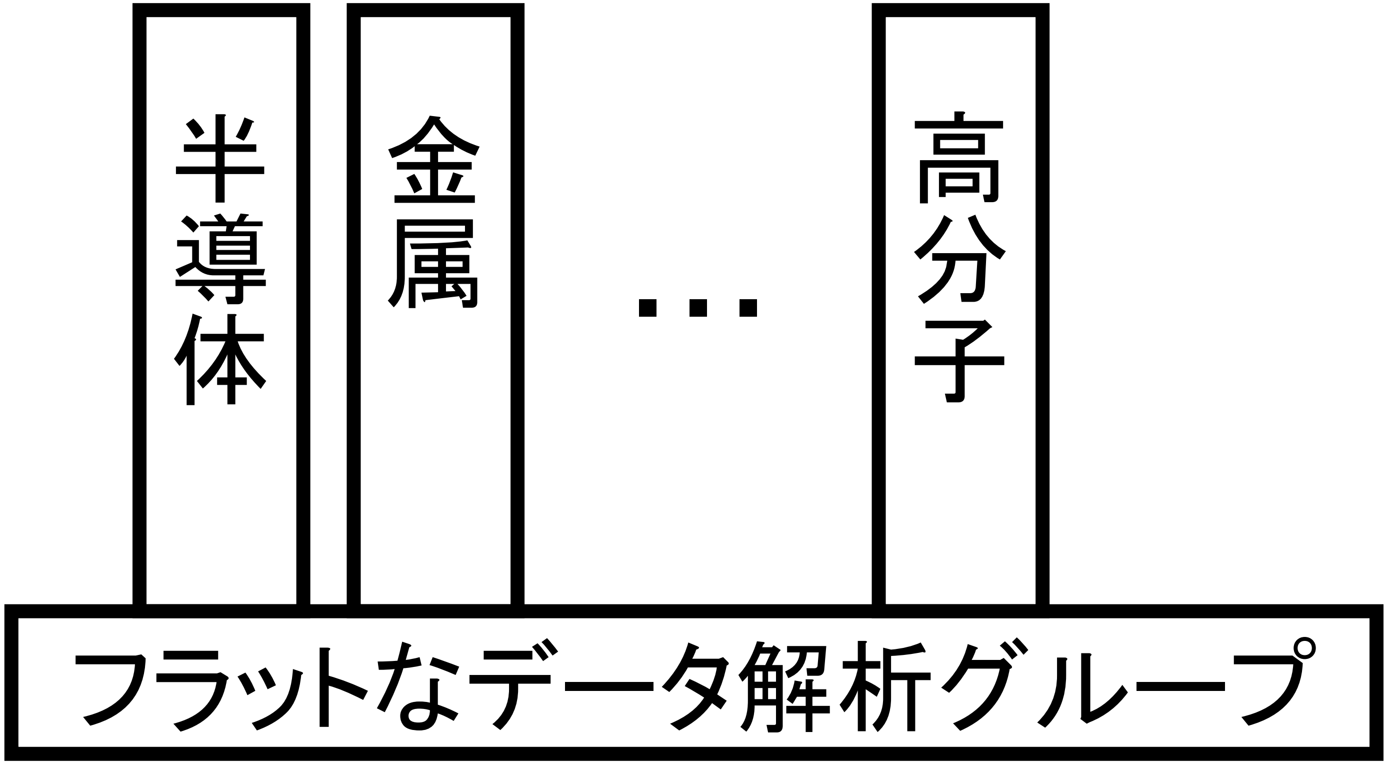
複雑理工学専攻とデータ駆動科学

- 複雑理工学専攻での研究教育活動の成果もあり、階層が異なる場合でも、階層の数理モデルを導出するための方法論としての、データ駆動科学に普遍性があることがわかった。
- 複雑理工学専攻では、ナノから宇宙までのスケールが異なる対象を同一専攻で取り扱っているために、この事実に気づくことができた。
- データ駆動科学の記述の普遍性により、基礎科学のだけでなく、民間企業での研究開発にも、データ駆動科学が普遍的に用いることができる。

データ駆動科学と就活

- データ駆動科学のベイズ推論とスパースモデリングは、民間企業での研究開発にも用いることができるだけでなく、これまで系統的に取り扱ってこられなかった民間企業の研究開発を、システマティックにする新しい枠組みである。
- データ駆動科学を習得した人材は、アカデミアだけでなく、民間企業においても、イニシアチブを取れる人材たる。
- これが、このデータ駆動科学教育プログラム(HD3)を開講した理由である。
- 受講生は、このような背景を理解して、授業に挑むことを期待する。

データ駆動科学による
Ｒ＆Ｄのフラット化と人材流動



ある縦組織に属していても、フラットなデータ解析グループのメンバーと共同作業することで、データ駆動科学の普遍性に接することができ、同じアルゴリズムが他の縦組織でも使用可能なことを感じることができ、その縦組織がリストラでなくなっても、他の縦組織で働くことができ、人材の流動化が加速される。

内容(1/3)

- データ駆動科学入門Iの紹介とゴール
  線形回帰$y=ax+b$の解析計算とスペクトル分解
- 高次元次元データ駆動科学教育プログラムの紹介
  プログラムの目的と設計指針
- 自然科学とデータ駆動科学
  自然記述の二つの戦略:
    要素還元主義と階層的自然観
    階層的自然観とデータ駆動科学
- データ駆動科学の二つの情報数理基盤
  スパースモデリング(SpM)
  ベイズ推論からベイズ計測へ

階層的自然観とデータ駆動科学

- 各階層に依存せずに、研究す能な普遍的枠組みは存在するか？
- その普遍的アプローチがデータ駆動科学にもとづく、スパースモデリングとベイズ推論であるというパラダイムを提案
- 実験計測データを説明する数理モデルが存在する場合としない場合が存在する。
  - 数理モデルが存在しない場合、スパースモデリングを用いて、データそのもの、もしくはデータの特徴量から、知りたい特性を予測する。その際、スパースモデリングで特徴量を刈り込むことにより、データの生成モデルを考察する際の参考にする。
  - データを説明するモデルが存在する場合、ベイズ推論を用いて、数理モデルのパラメータをデータから決定する。さらに、数理モデルが複数ある場合は、ベイズ推論を用いて、データだけから、モデル選択する。

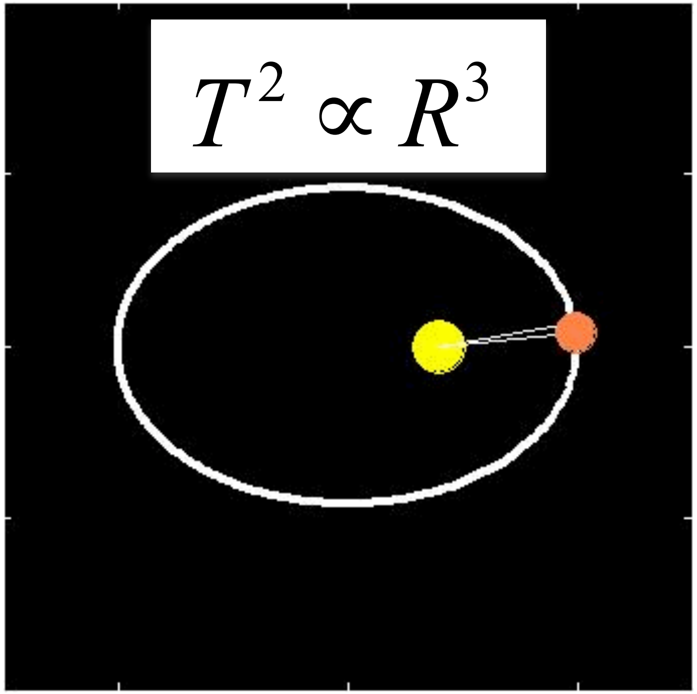
物理学とスパースモデリング(SpM)

- 古典力学や量子力学の前段階で、スパースモデリング(SpM)は活用されている歴史
- ニュートン力学に対するKeplerの法則
  - 公転周期Tと公転半径R
- 前期量子論
  - プランクの輻射の理論、アインシュタインの光量子仮説
- これらは全て、そのレベルを記述する数理モデルがない段階で、実験データから特徴量をヒトが決め、その特徴量を用いて、現象を定量的に記述する現象論である

Keplerの法則

ティコ･ブラーエの
天体観測データ

Keplerの法則



観測データからヒトが直感で特徴量TとRを抽出し
その定量的現象論を提案

新学術領域研究　平成25～29年度
スパースモデリングの深化と高次元データ駆動科学の創成

<u>領域代表岡田の個人的な狙い</u>

世界を系統的に記述したい
その方法論と枠組みを創りたい
ヒトが世界を認識するとは?

実験・計測

モデリング

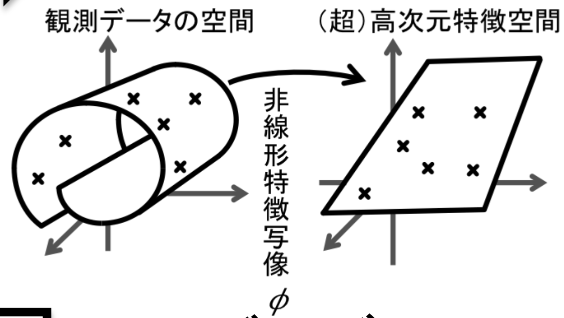
データ　　　　説明変数

$$\boldsymbol{y} = \text{基底}\Phi\,\boldsymbol{x}$$

非0　0

数理・情報

性能評価



アルゴリズムの
高速・高精度化

階層的自然観での普遍的な構造

現在考察している階層

分断!!

ミクロな階層

- 数理モデルのフリーパラメータの決定精度が、そのモデルの正しさ証明つながる
- ミクロなレベルと分断されているため、一般に複数の数理モデルが存在する

ベイズ計測の必然性(パラメータ推定、モデル選択)

内容(2/3)

- 計算論的神経科学とデータ駆動科学
  David Marrの三つのレベル
  データ駆動科学の三つのレベル
  連立方程式とデータ駆動科学
- 直線回帰$y=ax+b$の解析手計算
  - 最小二乗法の復習
  - ベイズ推論の導入とベイズ計測の創成
  - パラメータの確率分布、ノイズ分散推定、モデル選択

David Marrの三つのレベル (1982)

David Marrは複雑な情報処理装置を理解するには以下の三つのレベルが必要であると説いた

計算理論

計算の目的とその適切性を議論し，実行可能な方法の論理を構築

表現・アルゴリズム

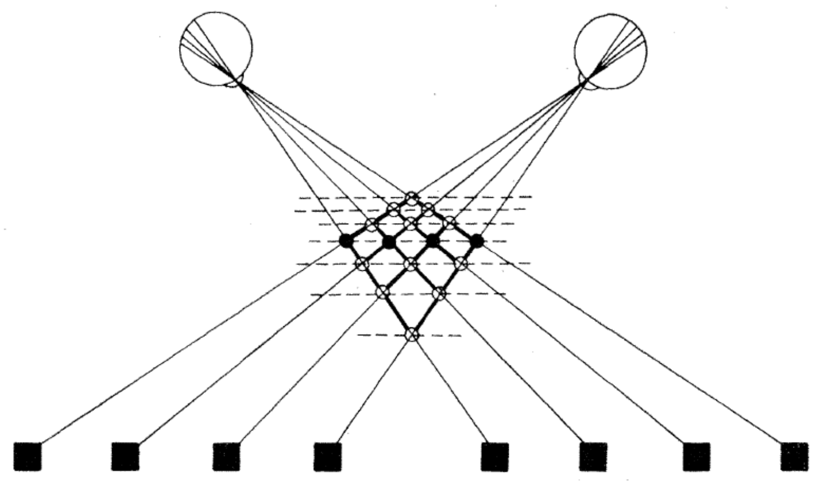
計算理論の実行方法．特にその入力と出力の表現と変換のためのアルゴリズム

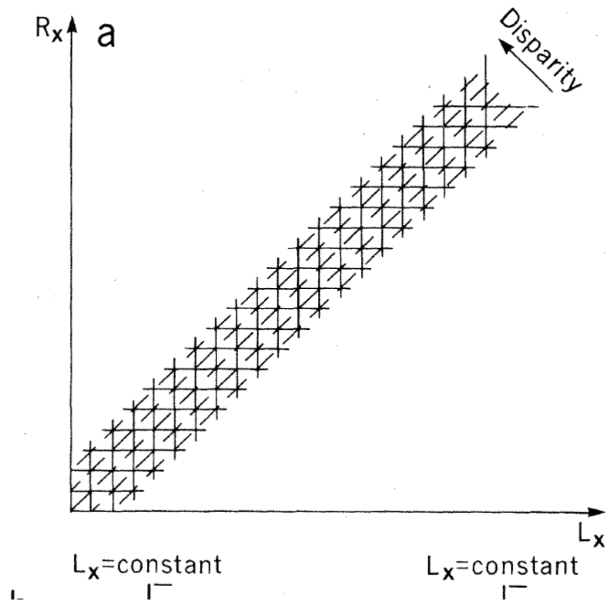
ハードウェア実装

表現とアルゴリズムの物理的な実現: ニューラルネットワーク

David Marr Vision: A Computational Investigation into the Human Representation and Processing of Visual Information (1982)

Marr Poggio: Cooperative Computation of Stereo Disparity *Science*, (1976)





両眼立体視三つのレベル (1976)

計算理論

計算の目的:両眼の対応問題を解く
適切性:解を一意に決めることは出来ない
網膜上で近い点同士の奥行きは近いという拘束条件を導入

表現・アルゴリズム

表現:右と左網膜に対して、それぞれ1次元の場を用意し、その1次元の場で奥行きを表現。1次元の場の間に、近い点同士の相互作用を導入し、計算理論で導入された拘束条件を表現

Marr Poggio: Cooperative Computation of Stereo Disparity *Science*, (1976)

データ駆動科学の三つのレベル (2016)

計算理論(対象の科学，計測科学)

データ解析の目的とその適切性を議論し，実行可能な方法の論理(方略)を構築

モデリング(統計学，理論物理学，数理科学)

計算理論のレベルの目的，適切さ，方略を元に，系をモデル化し，計算理論を数学的に表現する

表現・アルゴリズム(統計学，機械学習，計算科学)

モデリングの結果得られた計算問題を，実行するのためのアルゴリズムを議論する.

Igarashi, Nagata, Kuwatani, Omori, Nakanishi-Ohno and M. Okada, “Three Levels of Data-Driven Science”, *Journal of Physics: Conference Series*, 699, 012001, 2016.

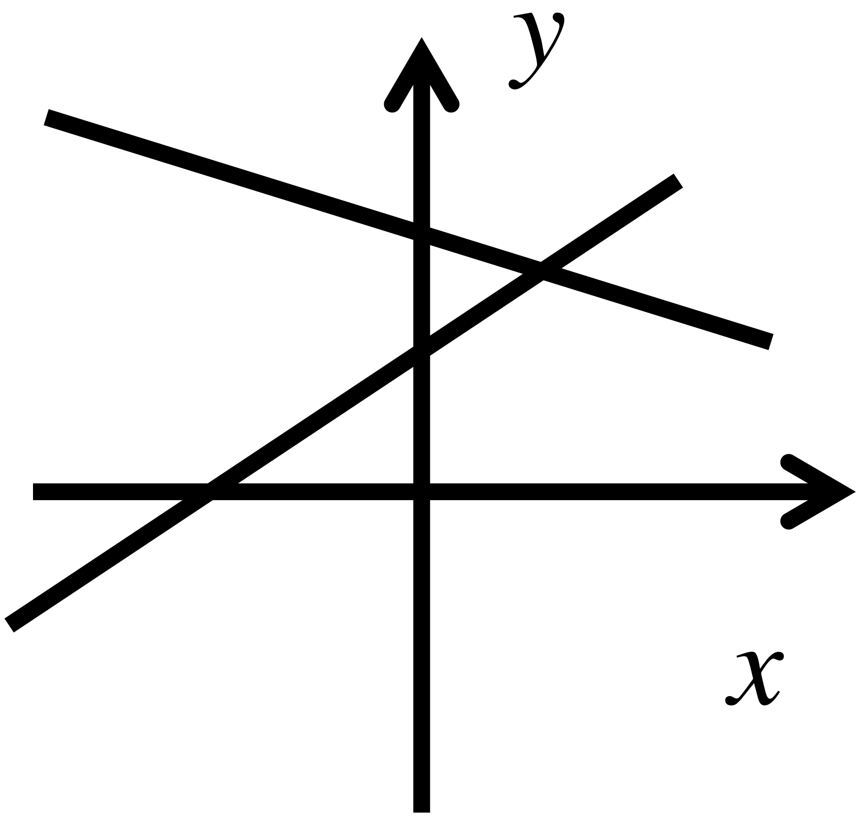
連立方程式とデータ駆動科学

連立方程式とその応用

鶴亀算　食塩水　寝坊して追いかける問題

連立方程式への変換

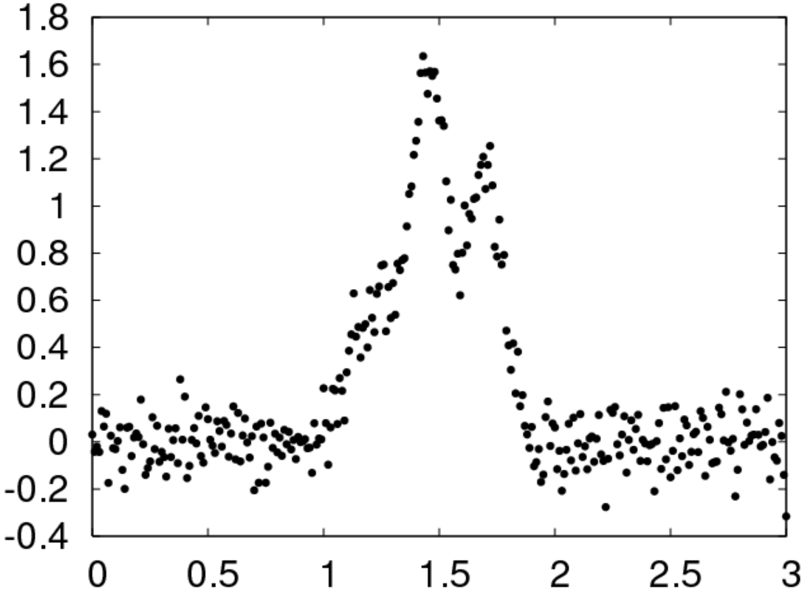
加減法, 代入法

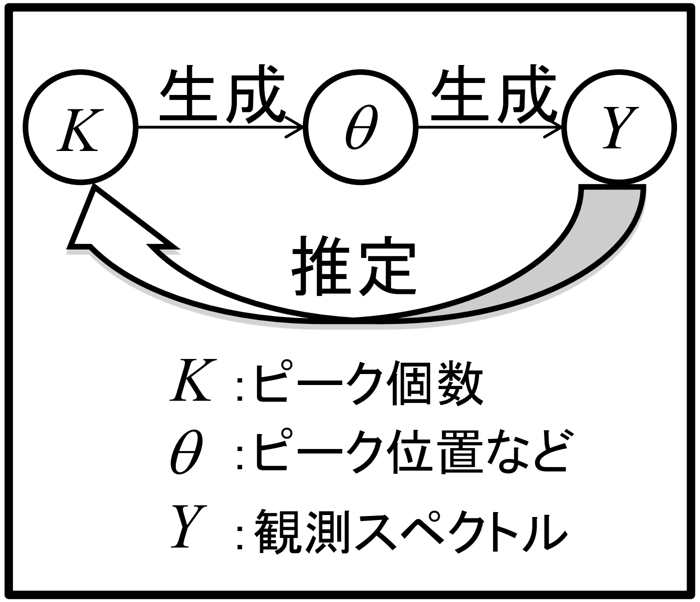


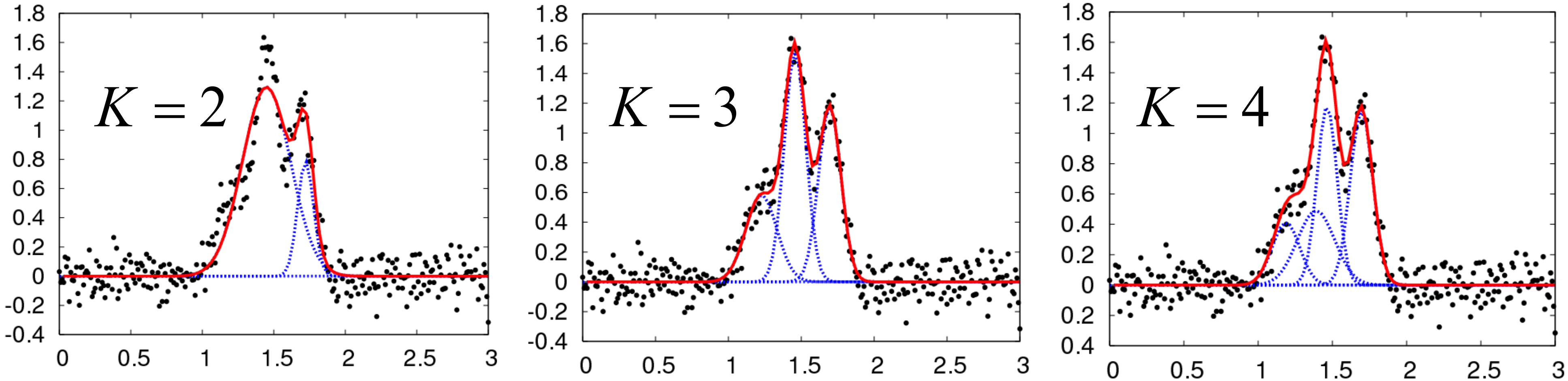
一つの方程式が一本の線
二本の線の交点が解になる

（五十嵐, 竹中, 永田, 岡田, *応用統計学*, 2016）

ベイス的スペクトル分解: Kをどう選ぶか







Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

スペクトル分解の三つのレベル (1/2)

スペクトル分解の計算理論

データ解析の目的: 多峰性スペクトルから背後にある離散
　　電子のエネルギー準位を推定する
データ解析の適切さ:　多峰性スペクトルを単峰性関数の線
　　形和で表し、その単峰性関数の個数を推定する。
　　誤差関数の最小化では、単峰性関数が多い方が誤差
　　関数は小さい。そこで統計学の交差検証誤差やベイズ
的モデル選択で単峰性関数の数Kを決める。

スペクトル分解のモデリング

多峰性スペクトルを単峰性関数の線形和に観測ノイズが付
加されれて生成するとモデリングする

スペクトル分解の三つのレベル (2/2)

スペクトル分解の表現･アルゴリズム

多峰性スペクトルを単峰性関数の線形和に観測ノイズが付加されれて生成するとモデリングし、ベイズ推論を適用することで、K個の単峰性関数の大きさ、位置、幅の事後確率を求める。各Kに対して、ベイズ的自由エネルギーを求め、ベイズ的自由エネルギーを最小にするKを求める。そのK個の単峰性関数の位置を、電子のエネルギー準位とする。

Igarashi, Nagata, Kuwatani, Omori, Nakanishi-Ohno and M. Okada, “Three Levels of Data-Driven Science”, *Journal of Physics: Conference Series*, 699, 012001, 2016.

Nagata, Sugita and M. Okada, “Bayesian spectral deconvolution with the exchange Monte Carlo method”, *Neural Networks*, 28, 82-89 2012.

内容(2/3)

- 計算論的神経科学とデータ駆動科学

  David Marrの三つのレベル

  データ駆動科学の三つのレベル

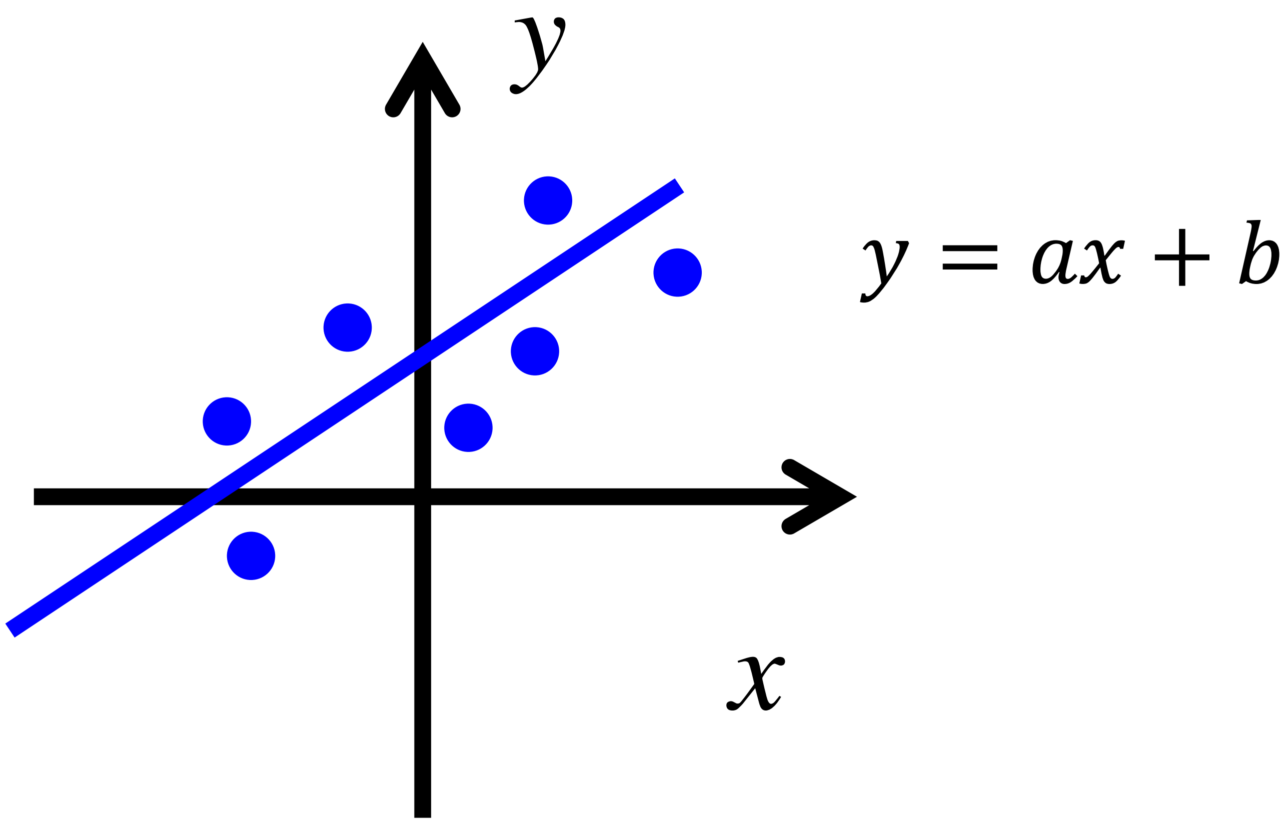
  連立方程式とデータ駆動科学
- 直線回帰$y=ax+b$の解析手計算
  - 最小二乗法の復習
  - ベイズ推論の導入とベイズ計測の創成
  - パラメータの確率分布、ノイズ分散推定、モデル選択

内容(2/3)

- 計算論的神経科学とデータ駆動科学
  David Marrの三つのレベル
  データ駆動科学の三つのレベル
  連立方程式とデータ駆動科学
- 直線回帰$y=ax+b$の解析手計算
  - 最小二乗法の復習
  - ベイズ推論の導入とベイズ計測の創成
  - パラメータの確率分布、ノイズ分散推定、モデル選択

早わかりベイズ計測の導入の効果
y=ax+bの取り扱いを通じて (1/4)

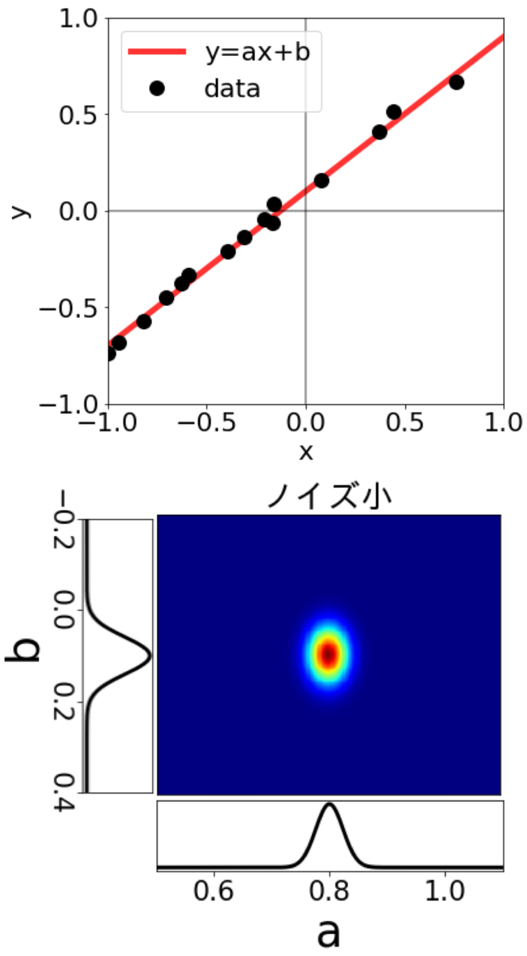
現状でも用いられている最も簡単な例

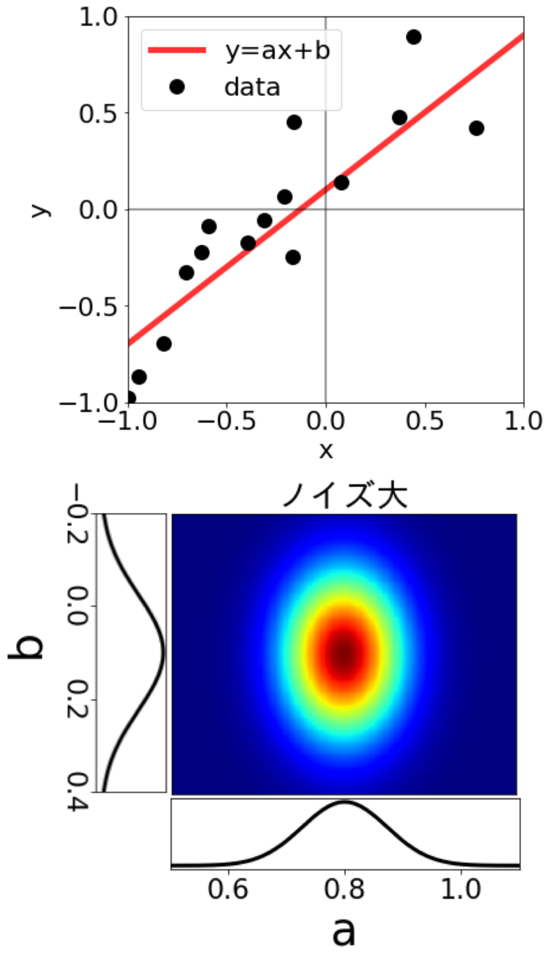


傾き a： 系の線形応答、バネ定数、電気伝導度、誘電率

早わかりベイズ計測の導入の効果
$y=ax+b$の取り扱いを通じて (2/4)

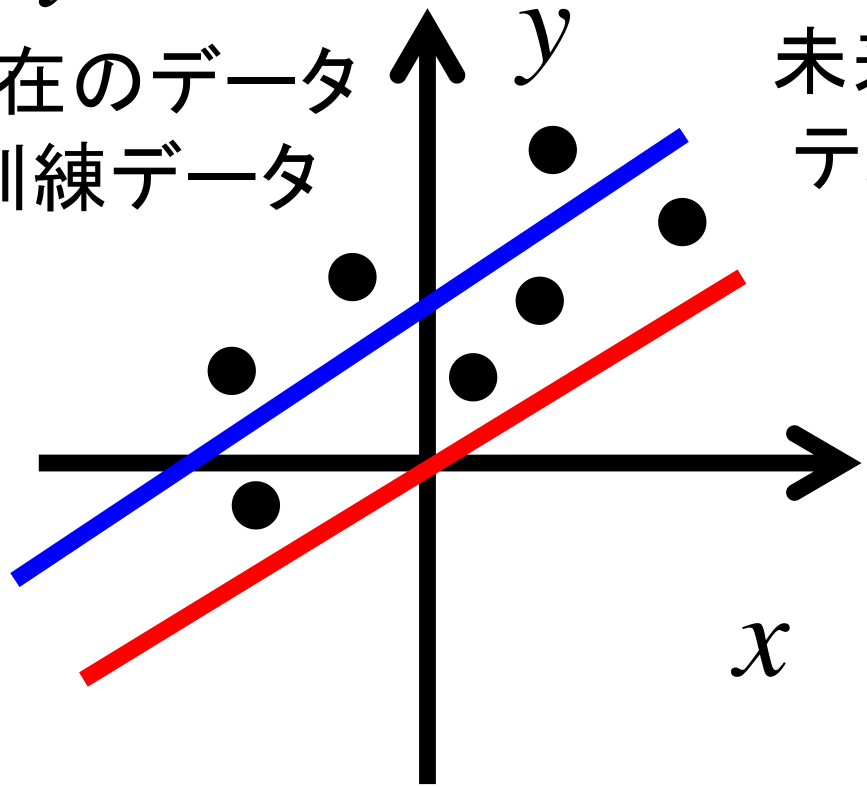
結論: 神器1 パラメータの事後確率推定





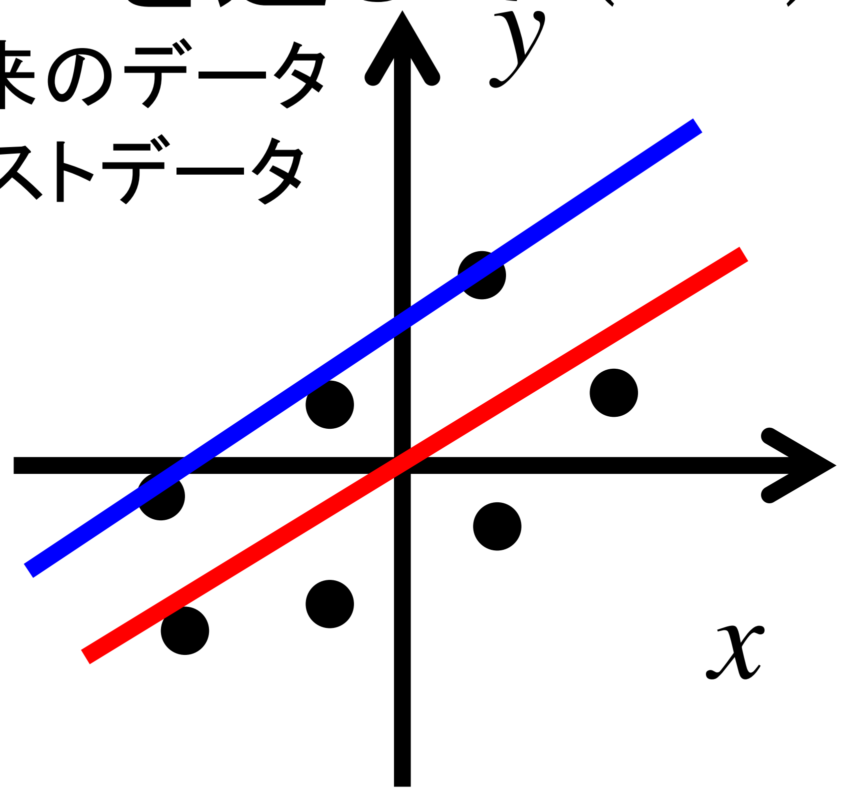
早わかりベイズ計測の導入の効果
$y=ax+b$の取り扱いを通じて (3/4)

現在のデータ
訓練データ



$y=ax+b$: 訓練誤差小
訓練誤差
ノイズに過学習

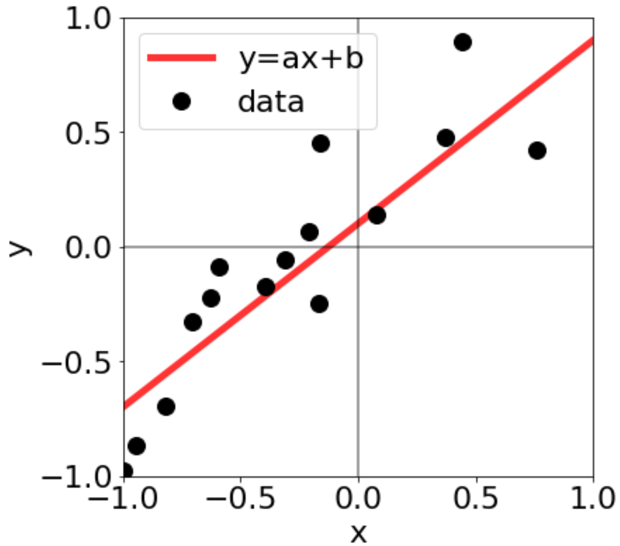
未来のデータ
テストデータ

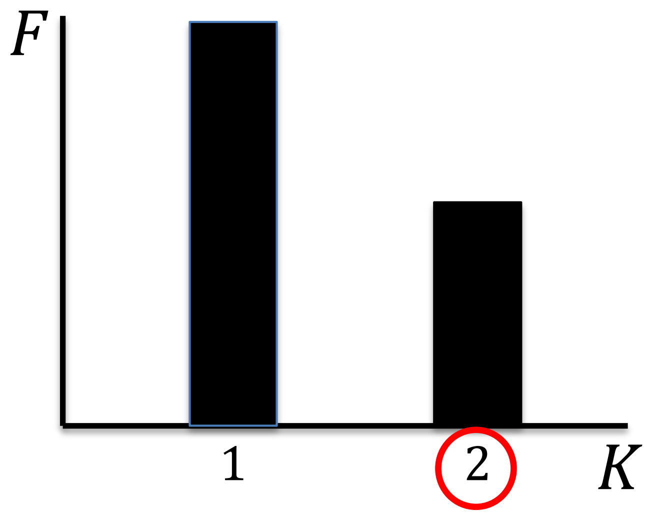


$y=ax$: 訓練誤差小
汎化誤差

モデル選択できる理由: 汎化誤差は観測ノイズに依存する

早わかりベイズ計測の導入の効果 $y=ax+b$の取り扱いを通じて (4/4)





- $K=1: y=ax$
- $K=2: y=ax+b$

$$F(K=1) = N\left\{\frac{1}{\sigma^2}E(a_0)+\frac{\log N}{2N}\right\}$$

$$F(K=2) = N\left\{\frac{1}{\sigma^2}E(a_0,b_0)+\frac{\log N}{N}\right\}$$

データのみからモデルを選択できる

内容(2/3)

- 計算論的神経科学とデータ駆動科学
  David Marrの三つのレベル
  データ駆動科学の三つのレベル
  連立方程式とデータ駆動科学
- 直線回帰$y=ax+b$の解析手計算
  - 最小二乗法の復習
  - ベイズ推論の導入とベイズ計測の創成
  - パラメータの確率分布、ノイズ分散推定、モデル選択

ここから直線回帰$y=ax+b$の解析計算の手計算を講義する（1/2）

- 手計算のpdfは配布済みであるが、勉強のために、必ず手計算すること。
- pdfがあるからいいと思って、計算を目で追うだけでは、達人にはなれない。
- 授業の後、何度も何度も、まるで写経のように手計算すると、単なる記号操作の式変形がセグメントされてきて、それぞれが意味を持つようになる。
- こうなれば、締めたものである。

ここから直線回帰$y=ax+b$の解析計算の手計算を講義する（2/2）

- 直線回帰ベイズ計測1.pdf
  - 大学1年で習う最小二乗法の復習
- 直線回帰ベイズ計測2.pdf
  - ベイズ推論の導入。事後確率推定、ノイズ分散推定
- 直線回帰ベイズ計測3.pdf
  - ベイズ的モデル選択: $y=ax$か$y=ax+b$かをデータのみから決める

$y=ax+b$の最小二乗法の復習（1/3）

$$E(a,b)=\frac{1}{N}\sum_{i=1}^{N}(y_i-(ax_i+b))^2$$

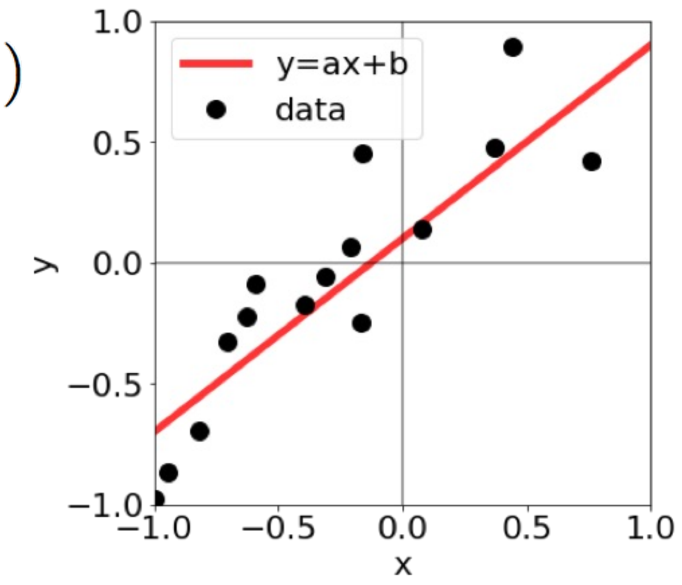
二乗誤差$E(a,b)$を最小にするようにパラメータをフィット（最小二乗法）

$$\overline{x}=\frac{1}{N}\sum_{i=1}^{N}x_i=0$$ とする場合

$$E(a,b)=\frac{1}{2}\left(\overline{x^2}\left(a-\frac{\overline{xy}}{\overline{x^2}}\right)^2+(b-\bar{y})^2-\frac{\overline{xy}^2}{\overline{x^2}}-\overline{y}^2+\overline{y^2}\right)$$

$$a_0=\frac{\overline{xy}}{\overline{x^2}}\quad b_0=\overline{y}$$

$$=\mathcal{E}_a(a)+\mathcal{E}_b(b)+E(a_0,b_0)\geqq E(a_0,b_0)$$



平均：$\bar{y}=\frac{1}{N}\sum_{i=1}^{N}y_i$，分散：$\overline{x^2}=\frac{1}{N}\sum_{i=1}^{N}x_i^2$

$$\overline{xy}=\frac{1}{N}\sum_{i=1}^{N}x_iy_i$$

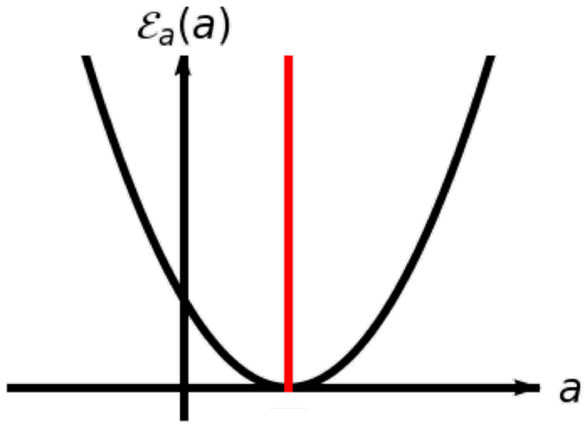
y=ax+bの最小二乗法の復習 (2/3)

$$E(a,b)=\frac{1}{N}\sum_{i=1}^{N}(y_i-(ax_i+b))^2$$

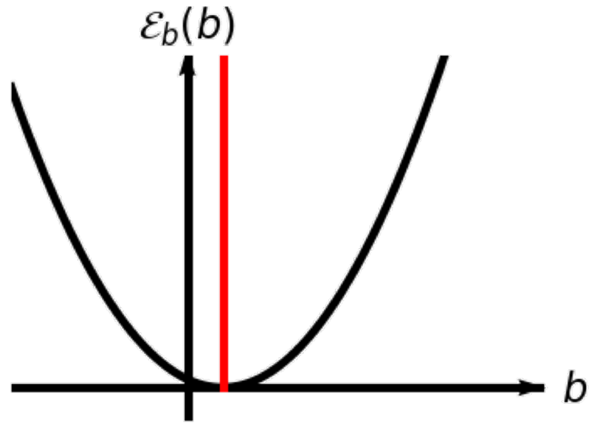
$$E(a,b)=\frac{1}{2}\left(\overline{x^2}\left(a-\frac{\overline{xy}}{\overline{x^2}}\right)^2+(b-\bar{y})^2-\frac{\overline{xy}^2}{\overline{x^2}}-\bar{y}^2+\overline{y^2}\right)$$

$$a_0=\frac{\overline{xy}}{\overline{x^2}}\quad b_0=\bar{y}$$

$$E(a,b)=\mathcal{E}_a(a)+\mathcal{E}_b(b)+E(a_0,b_0)\geqq E(a_0,b_0)$$



$$a_0=\frac{\overline{xy}}{\overline{x^2}}$$



$$b_0=\bar{y}$$

y=ax+bの最小二乗法の復習 (3/3)

$$E(a,b) = \frac{1}{N}\sum_{i=1}^{N}(y_i - (ax_i + b))^2$$

二乗誤差$E(a,b)$を最小にするようにパラメータをフィット（最小二乗法）

$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = 0$ とする場合

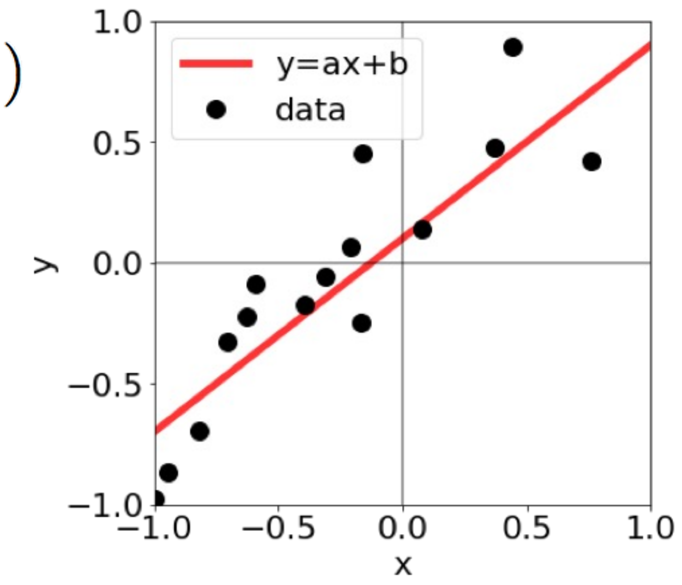
$$E(a,b) = \frac{1}{2}\left(\overline{x^2}\left(a - \frac{\overline{xy}}{\overline{x^2}}\right)^2 + (b-\bar{y})^2 - \frac{\overline{xy}^2}{\overline{x^2}} - \bar{y}^2 + \overline{y^2}\right)$$

$$a_0 = \frac{\overline{xy}}{\overline{x^2}} \quad b_0 = \overline{y}$$

$$= \mathcal{E}_a(a) + \mathcal{E}_b(b) + E(a_0, b_0) \geqq E(a_0, b_0)$$

平均: $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$, 分散: $\overline{x^2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2$

$$\overline{xy} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i$$



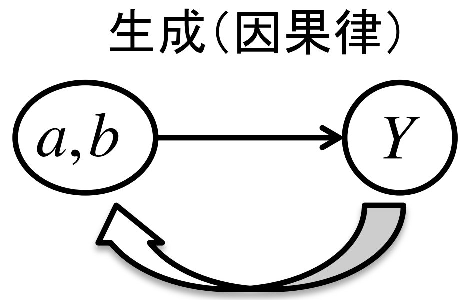
内容(2/3)

- 計算論的神経科学とデータ駆動科学
  David Marrの三つのレベル
  データ駆動科学の三つのレベル
  連立方程式とデータ駆動科学
- 直線回帰$y=ax+b$の解析手計算
  - 最小二乗法の復習
  - ベイズ推論の導入とベイズ計測の創成
  - パラメータの確率分布、ノイズ分散推定、モデル選択

ベイズ推論の導入とベイズ計測の創成

ベイズの定理による
神器1: パラメータの事後確率推定

$$p(Y,a,b) = \underline{p(Y|a,b)\,p(a,b) = p(a,b|Y)\,p(Y)}$$



$$\Downarrow$$

＜ベイズの定理＞

$$p(a,b|Y) = \frac{p(Y|a,b)\,p(a,b)}{p(Y)} \propto \exp\left(-nE(a,b)\right)p(a,b)$$

$p(a,b|Y)$ ：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.

$p(a,b)$ ：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

内容(2/3)

- 計算論的神経科学とデータ駆動科学
  David Marrの三つのレベル
  データ駆動科学の三つのレベル
  連立方程式とデータ駆動科学
- 直線回帰$y=ax+b$の解析手計算
  - 最小二乗法の復習
  - ベイズ推論の導入とベイズ計測の創成
  - パラメータの確率分布、ノイズ分散推定、モデル選択

神器1: パラメータの事後確率推定(1/3)

$$y_i = ax_i + b + n_i$$

$$p(n_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{n_i^2}{2\sigma^2}\right)$$

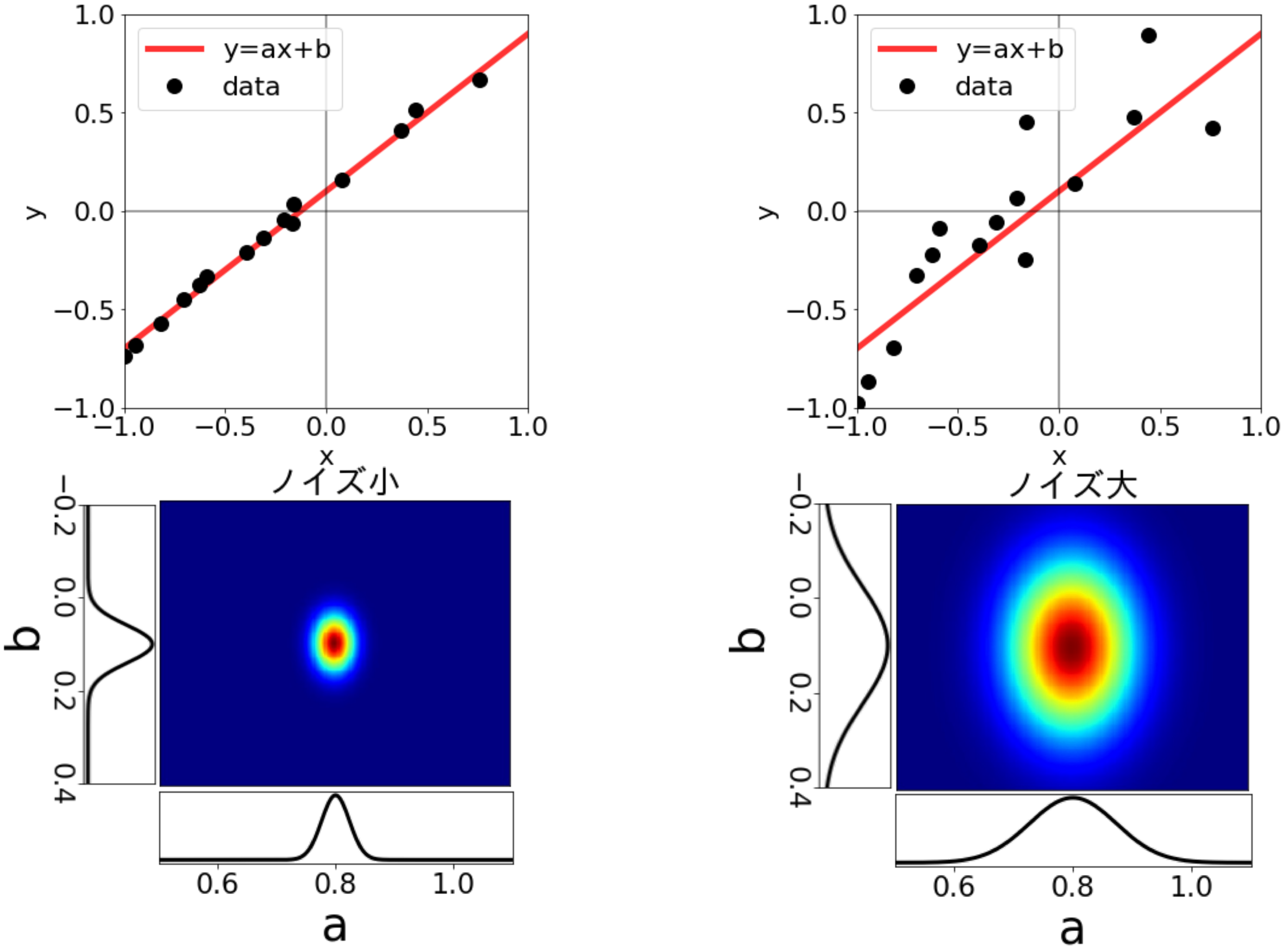
$$p(n_i) = p(y_i|a,b) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - (ax_i + b))^2}{2\sigma^2}\right)$$

$$\begin{aligned} p(Y|a,b) &= \prod_{i=1}^{N} p(y_i|a,b) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left(-\frac{\sum_{i=1}^{N}(y_i - (ax_i + b))^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left(-\frac{N}{\sigma^2}E(a,b)\right) \end{aligned}$$

神器1: パラメータの事後確率推定(2/3)

$$
\begin{aligned}
p(a,b|Y) &= \frac{p(Y|a,b)p(a,b)}{p(Y)} \propto p(Y|a,b) \\
&= \exp\left\{-\frac{N}{\sigma^2}\Big(\mathcal{E}_a(a)+\mathcal{E}_b(b)+E(a_0,b_0)\Big)\right\} \\
&\propto \exp\left\{-\frac{N}{\sigma^2}\Big(\mathcal{E}_a(a)+\mathcal{E}_b(b)\Big)\right\} \\
&= \exp\left\{-\frac{N\overline{x^2}}{2\sigma^2}(a-a_0)^2+\frac{N}{2\sigma^2}(b-b_0)^2\right\}
\end{aligned}
$$

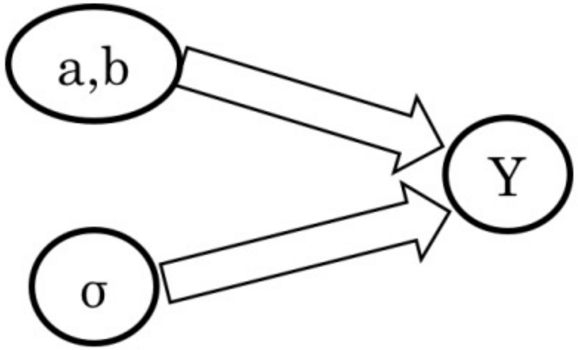
神器1: パラメータの事後確率推定(3/3)



内容(2/3)

- 計算論的神経科学とデータ駆動科学
  David Marrの三つのレベル
  データ駆動科学の三つのレベル
  連立方程式とデータ駆動科学
- 直線回帰$y=ax+b$の解析手計算
  - 最小二乗法の復習
  - ベイズ推論の導入とベイズ計測の創成
  - パラメータの確率分布、ノイズ分散推定、モデル選択

神器1: パラメータの事後確率推定ノイズ分散推定



$$
\begin{aligned}
p(\sigma^2|Y) &\propto \int dadb\, p(Y|a,b,\sigma^2) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \int dadb \exp\left\{-\frac{N}{\sigma^2}E(a,b)\right\} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \left\{\exp\left(-\frac{N}{\sigma^2}E(a_0,b_0)\right) + \int da \exp\left(-\frac{N\overline{x^2}}{2\sigma^2}(a-a_0)^2\right) + \int db \exp\left(-\frac{N}{2\sigma^2}(b-b_0)^2\right)\right\} \\
&= (2\pi\sigma^2)^{-\frac{N-2}{2}}(N^2\overline{x^2})^{\frac{1}{2}} \exp\left(-\frac{N}{\sigma^2}E(a_0,b_0)\right) \qquad (25)
\end{aligned}
$$

$$
\sigma^2 = \frac{NE(a_0,b_0)}{N-2} = \frac{1}{N-2}\sum_{i=1}^{N}\{y_i - (a_0x_i + b_0)\}^2
$$

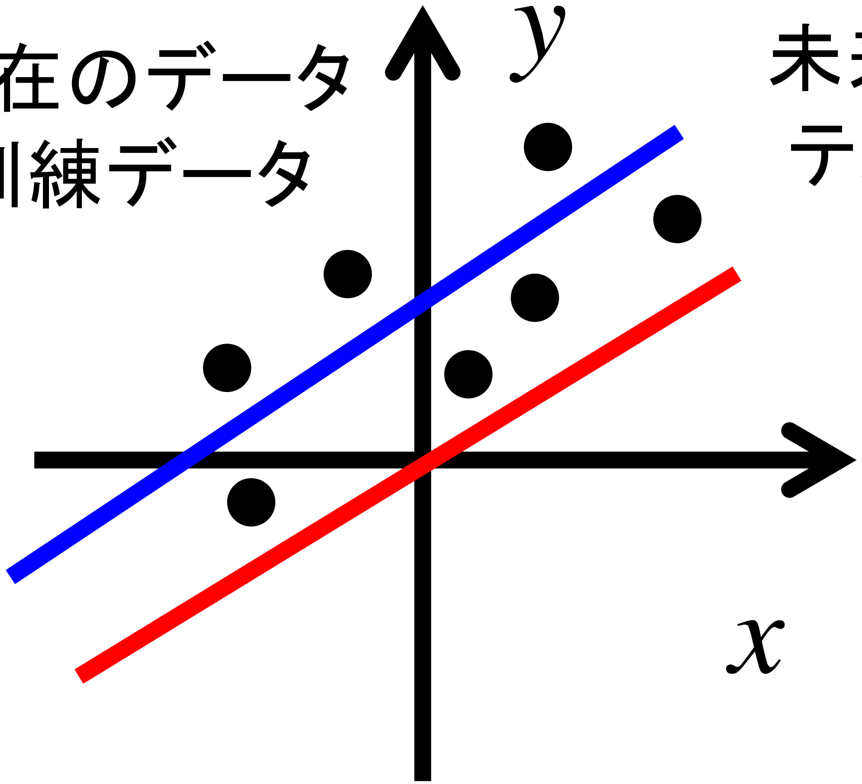
内容(2/3)

- 計算論的神経科学とデータ駆動科学
  David Marrの三つのレベル
  データ駆動科学の三つのレベル
  連立方程式とデータ駆動科学
- 直線回帰$y=ax+b$の解析手計算
  - 最小二乗法の復習
  - ベイズ推論の導入とベイズ計測の創成
  - パラメータの確率分布、ノイズ分散推定、モデル選択

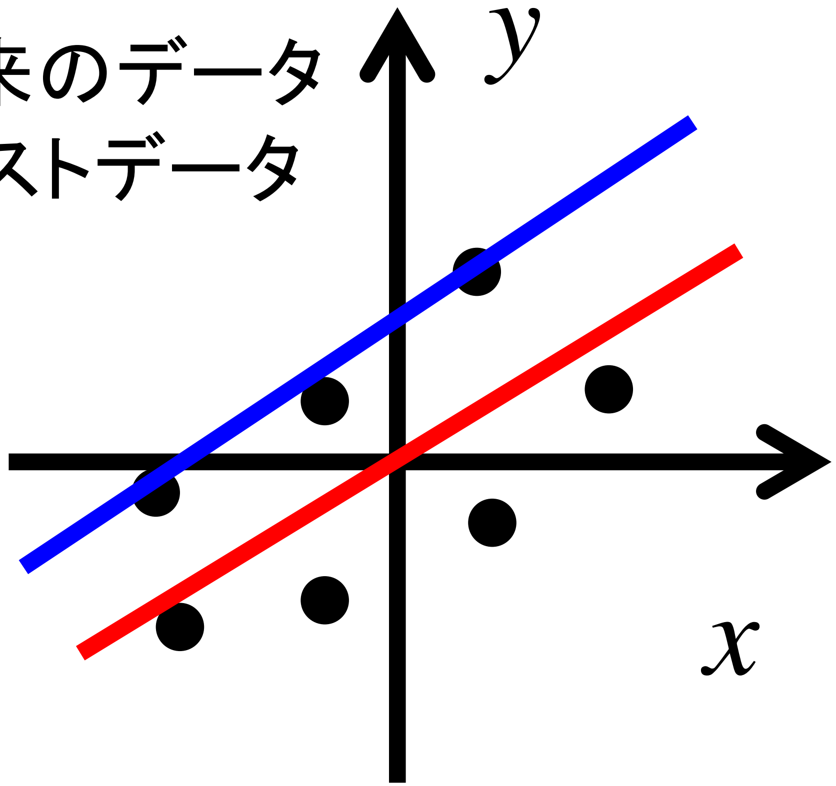
問題意識　神器2: ベイズ的モデル選択 (1/3)

$y=ax$か$y=ax+b$か?

現在のデータ
訓練データ



未来のデータ
テストデータ

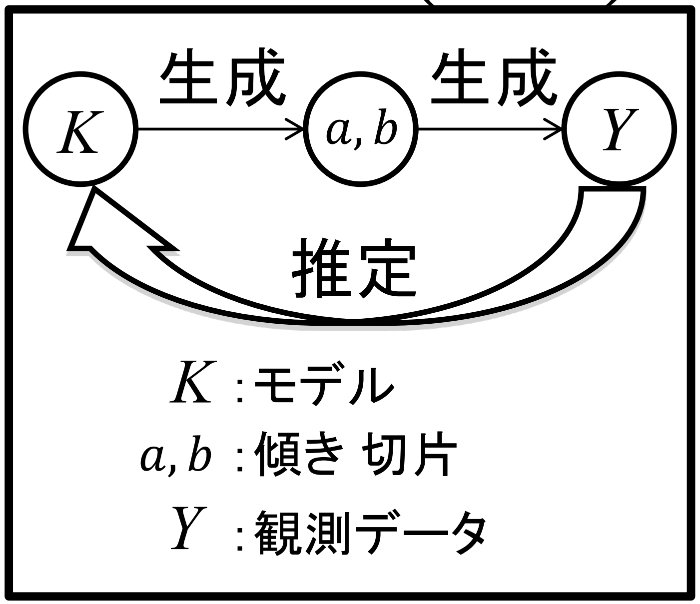


$y=ax+b$: 訓練誤差小
訓練誤差
ノイズに過学習

$y=ax$: 訓練誤差小
汎化誤差

モデル選択できる理由:汎化誤差は観測ノイズに依存する

神器2: ベイズ的モデル選択 (2/3)



1. 欲しいのは　$p(K \mid Y)$

2. θがないぞ

3. $p(K,\theta,Y)$ の存在を仮定

$$p(K,\theta,Y) = p(Y \mid \theta, K)p(K)$$

$$p(Y \mid \theta, K) = \prod_{i=1}^{n} p(y_i \mid \theta) \propto \exp(-nE(\theta))$$

4. 無駄な自由度の系統的消去: 周辺化，分配関数

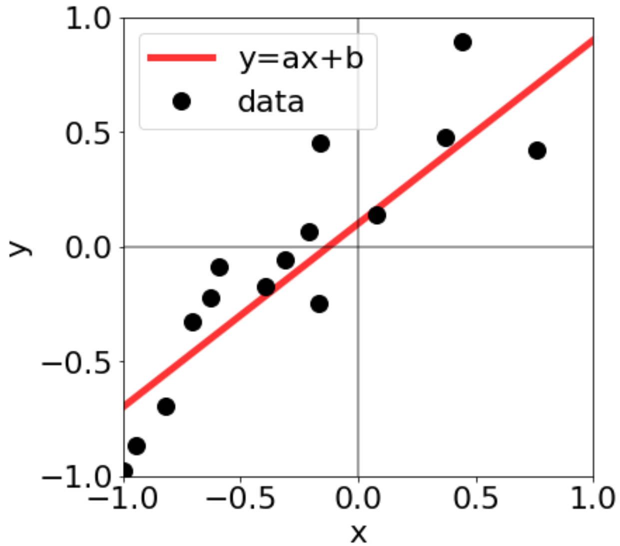
$$p(K,Y) = \int p(K,\theta,Y)d\theta$$

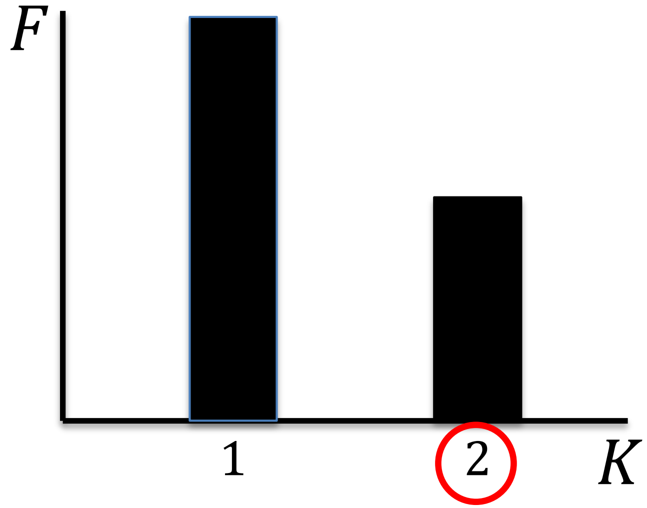
$$p(K \mid Y) = \frac{p(Y \mid K)p(K)}{p(Y)} \propto p(K)\int \exp(-nE(\theta))p(\theta)d\theta$$

$$F(K) = -\log \int \exp(-nE(\theta))p(\theta)d\theta$$

自由エネルギーを最小にするモデル K を求める.

神器2: ベイズ的モデル選択 (3/3)
$\boldsymbol{y = ax}$か$\boldsymbol{y = ax + b}$か?





- $K = 1 : y = ax$
- $K = 2 : y = ax + b$

$$F(K=1) = N\left\{\frac{1}{\sigma^2}E(a_0) + \frac{\log N}{2N}\right\}$$

$$F(K=2) = N\left\{\frac{1}{\sigma^2}E(a_0, b_0) + \frac{\log N}{N}\right\}$$

データのみからモデルを選択できる

まとめ: ベイズ計測三種の神器 $y=ax+b$の解析取り扱いを通じて

- 従来の最小二乗法
  - 1.物理パラメータの点推定
- ベイズ計測

1. 物理パラメータの確率分布推定
2. データからのベイズ的モデル選択
3. ベイズ統合: 今回は説明を省略
   - 水牧先生の基調講演