国立情報学研究所

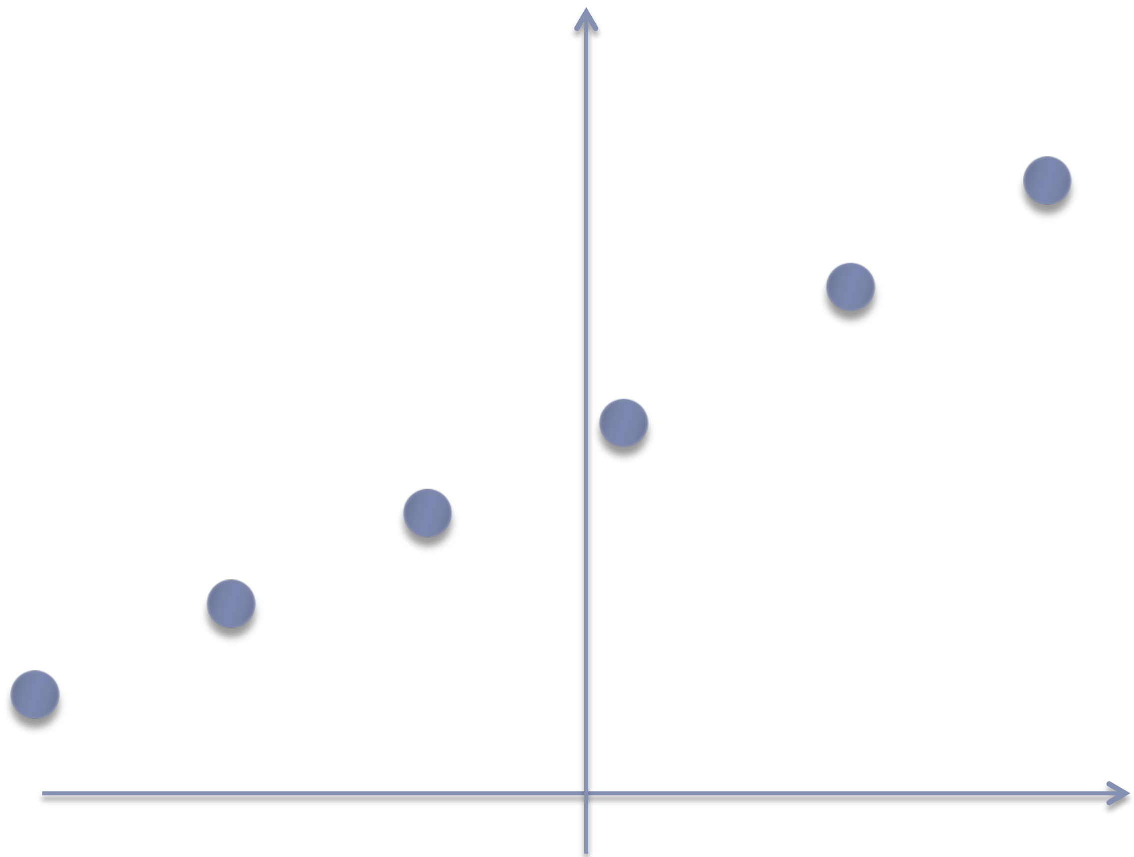
今日からできる スパースモデリング

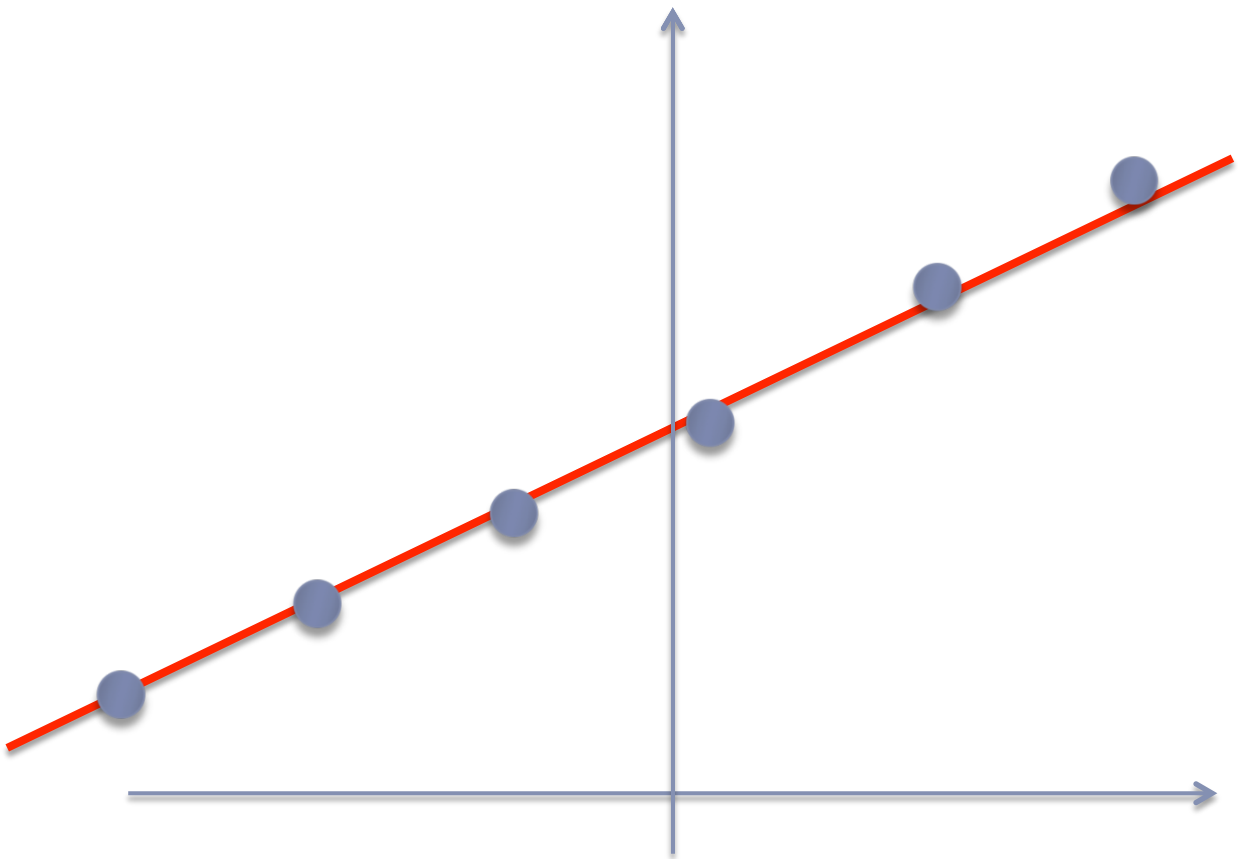
京都大学大学院情報学研究科システム科学専攻
大関　真之

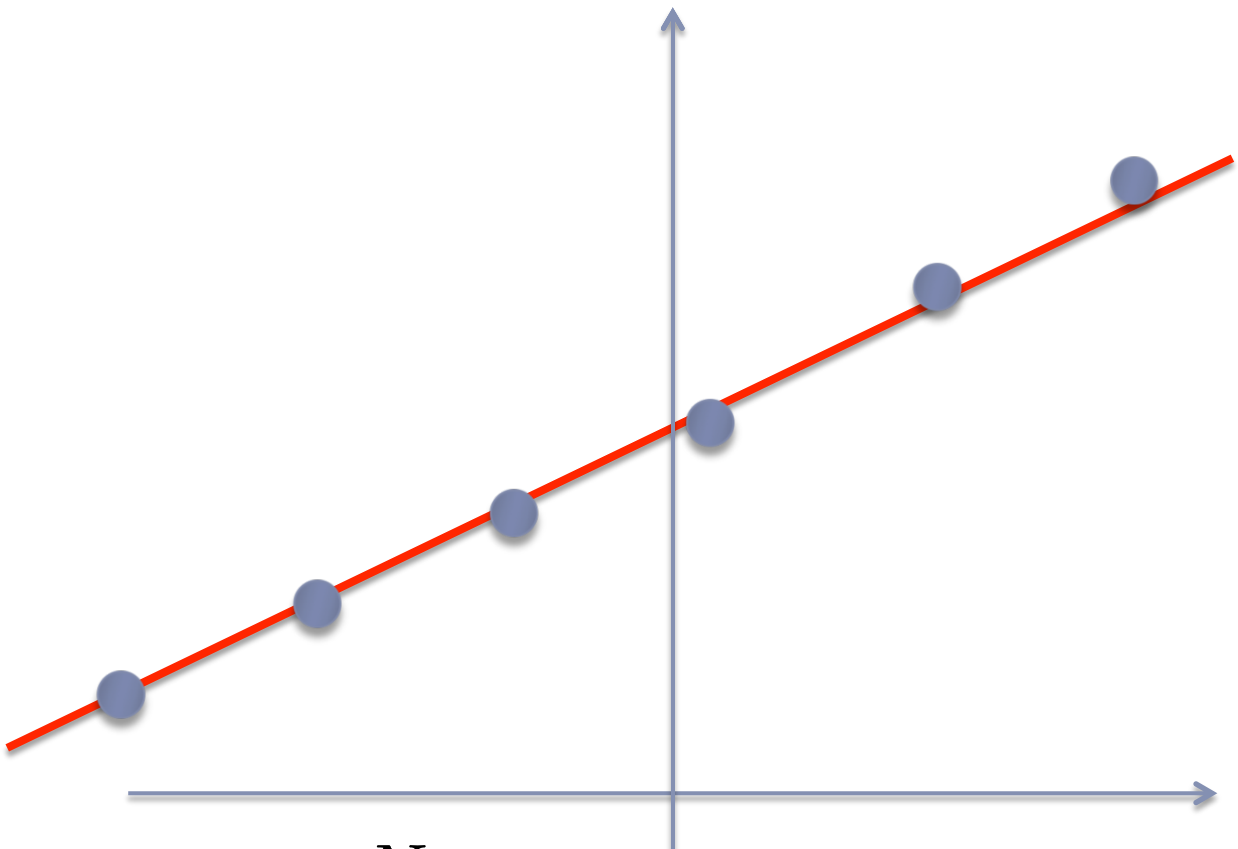
ボルツマン機械学習編

機械学習とは

データに対する関数の自律的フィッティング







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

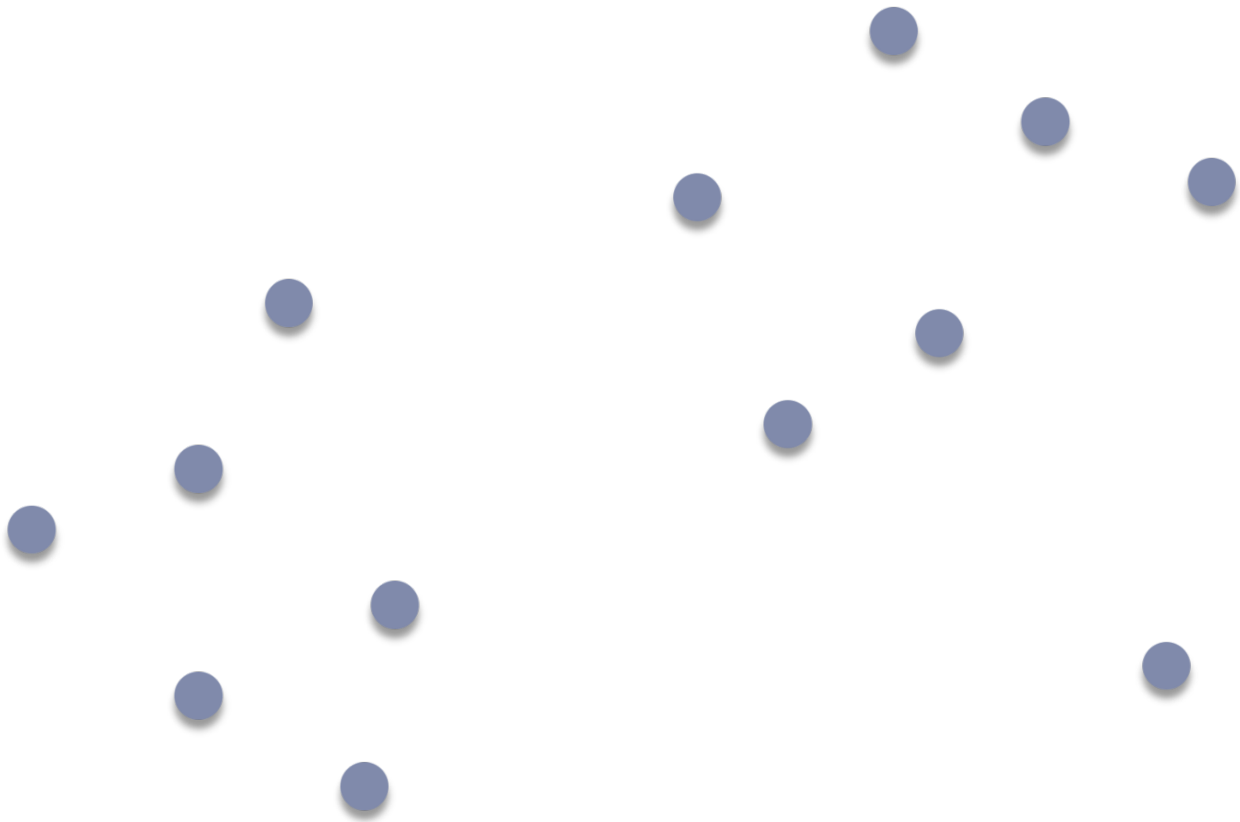
多くのデータから回帰する

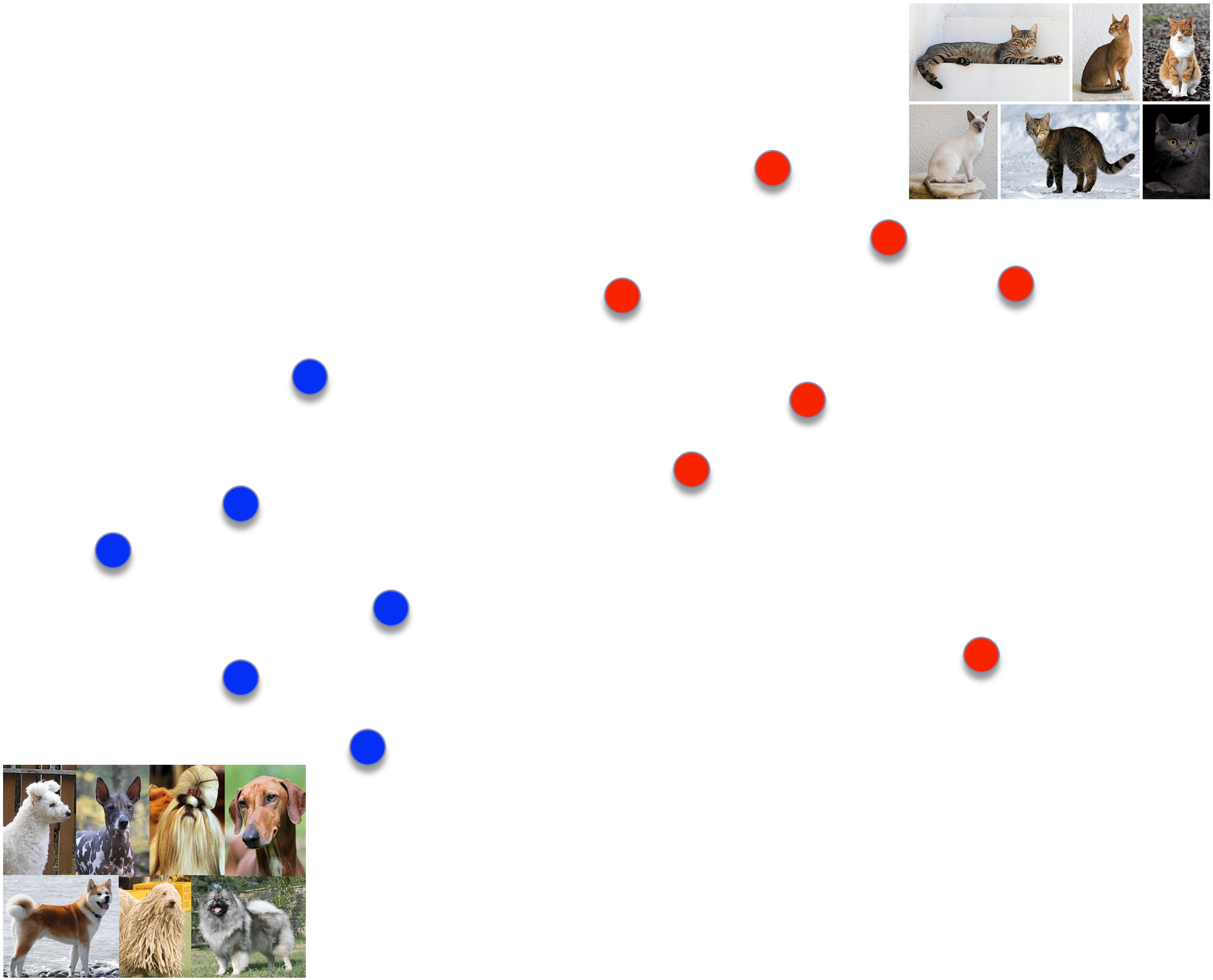
ここでも最適化問題

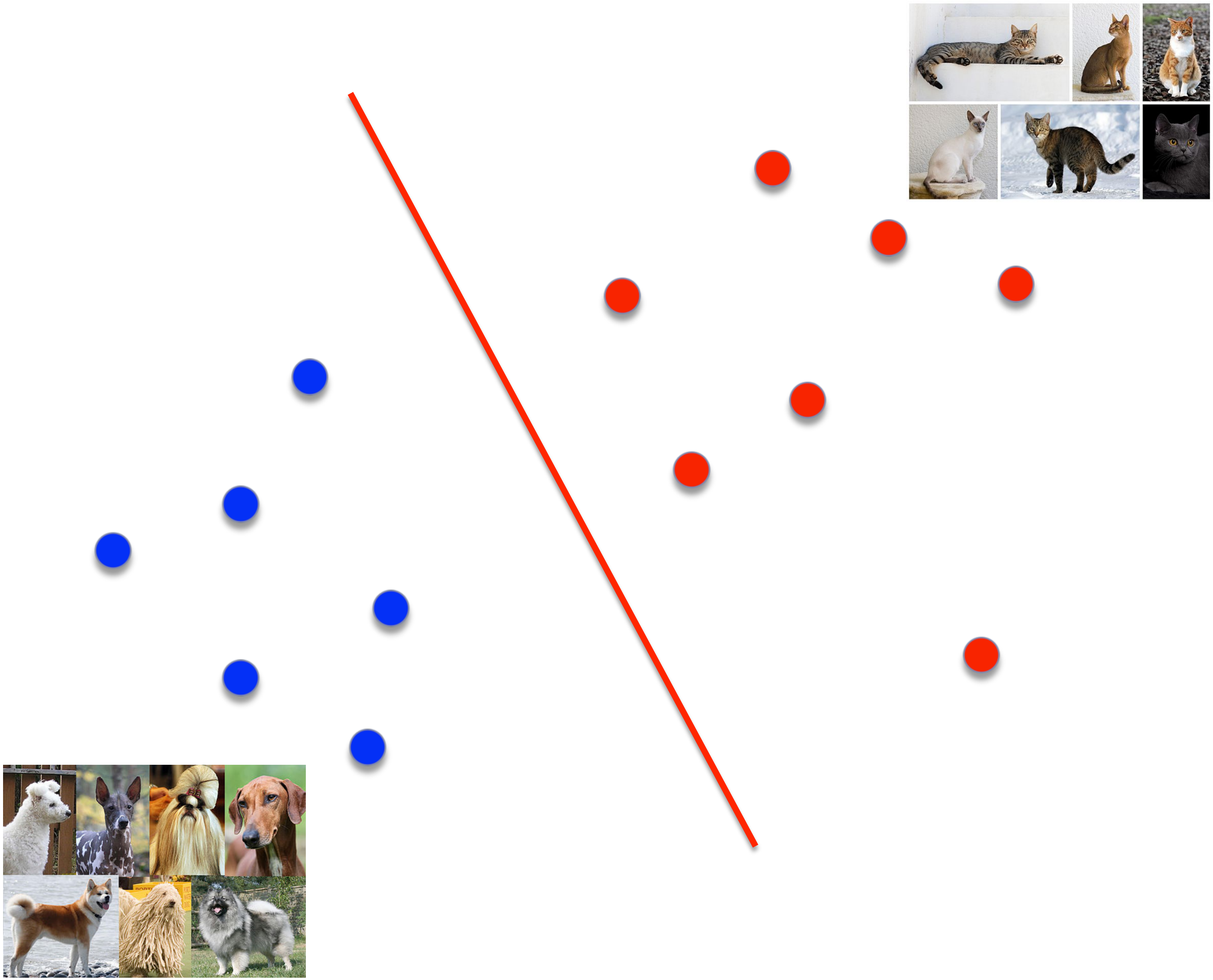
$$\min_{a,b} \sum_{i=1}^{N} \{y_i - (ax_i + b)\}^2$$

多くのデータから識別をする

パーセプトロン・サポートベクターマシーン







多くのデータから識別をする
パーセプトロン・サポートベクターマシーン
ここでも最適化問題

多くのデータからその素性を明らかにする営み

機械学習

多くの実験データから関数関係を明らかにする

実験レポートでやらされたこと

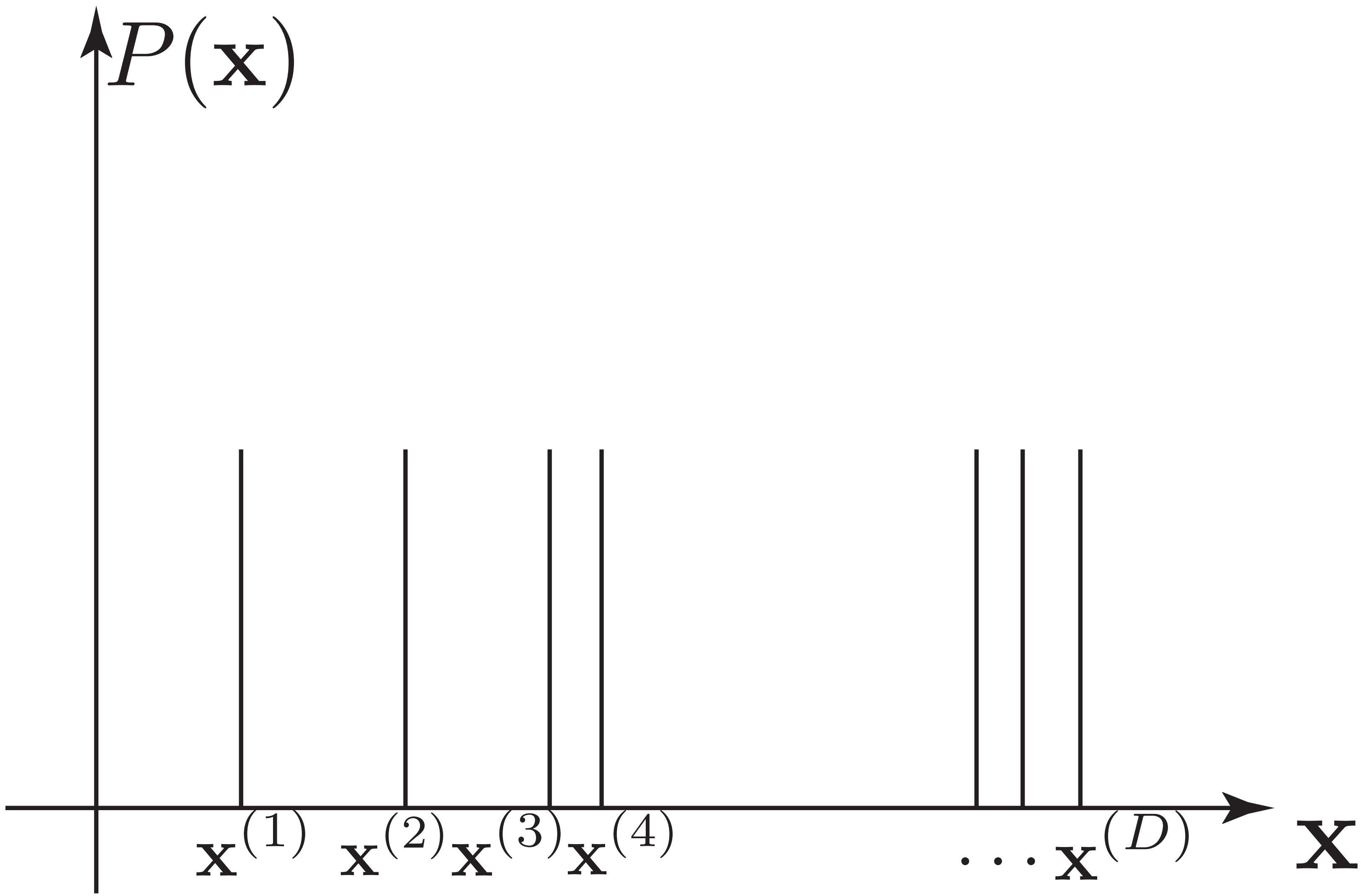
多くの購買データから顧客の嗜好を明らかにする

Am◯zon

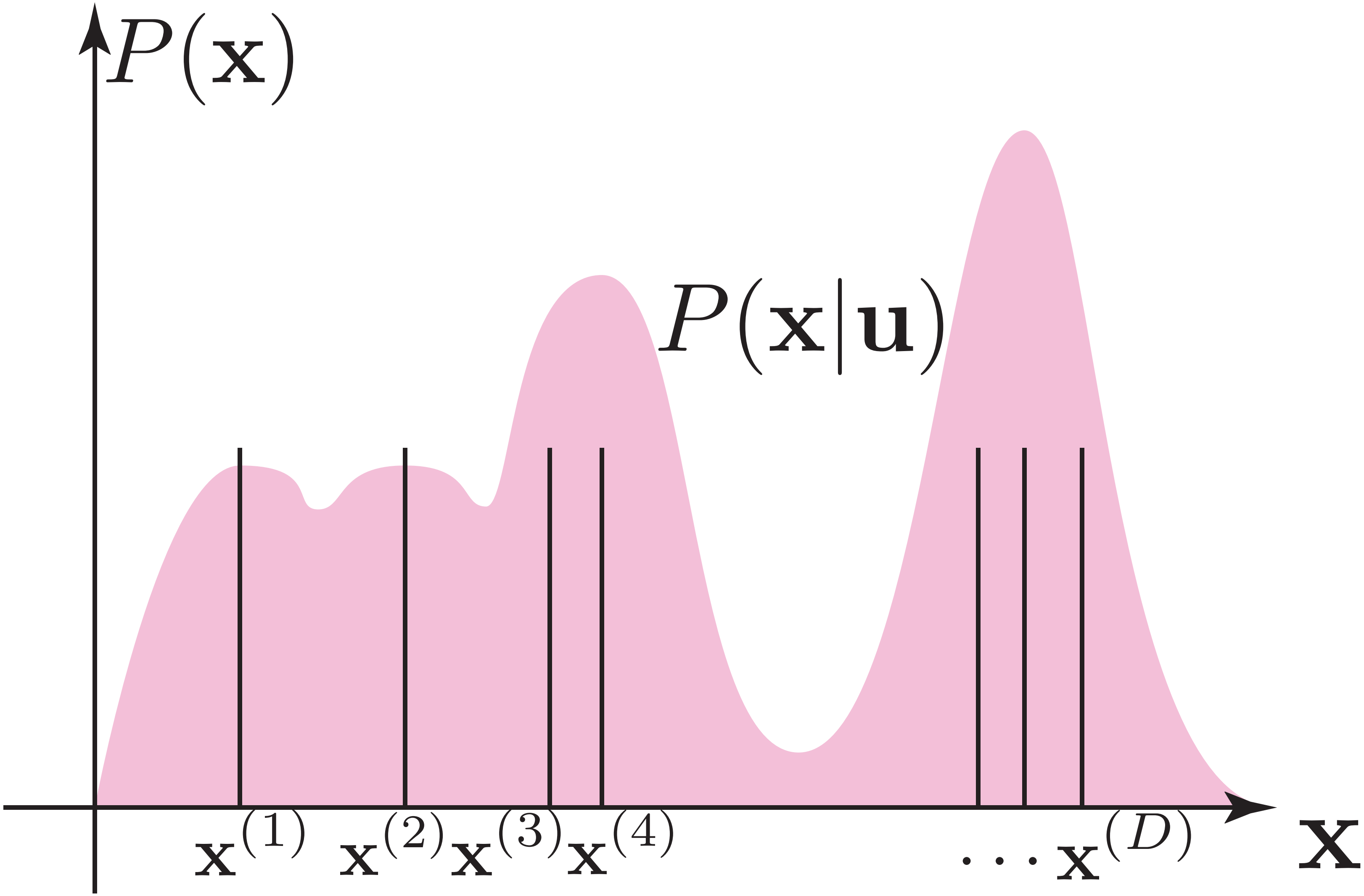
多くのWeb訪問履歴から使用者の意思を明らかにする

Go◯gle

多くのデータから確率分布を明らかにする
ボルツマン機械学習



$P(\mathbf{x})$
$\mathbf{x}^{(1)}$
$\mathbf{x}^{(2)}$
$\mathbf{x}^{(3)}$
$\mathbf{x}^{(4)}$
$\dots \mathbf{x}^{(D)}$
$\mathbf{X}$



$P(\mathbf{x})$
$P(\mathbf{x}|\mathbf{u})$
$\mathbf{x}^{(1)}$
$\mathbf{x}^{(2)}$
$\mathbf{x}^{(3)}$
$\mathbf{x}^{(4)}$
$\dots \mathbf{x}^{(D)}$
$\mathbf{x}$

統計的性質

平均？分散？

統計的性質

平均？分散？

分布を知りたい

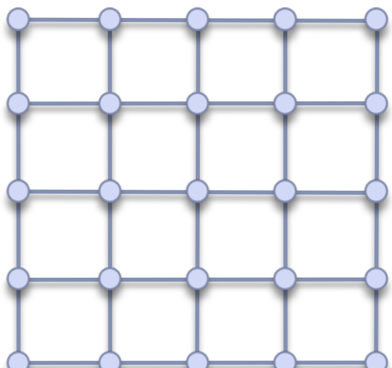
ボルツマン機械学習

- 多数の数値列、データからの分布の推定
  - 生成モデルの存在の仮定

$$P(\mathbf{x}|\mathbf{u}) = \frac{1}{Z(\mathbf{u})} \exp\left\{-E(\mathbf{x}|\mathbf{u})\right\}$$

- 例：Ising模型（グラフは任意）

$$E(\mathbf{x}|\mathbf{u}) = \sum_{i=1}^{N} \sum_{j \in \partial i} J_{ij} x_i x_j + \sum_{i=1}^{N} h_i x_i$$



QUANTUM ANNEALING MACHINE LEARNING Sparse Modeling CREST

ボルツマン機械学習

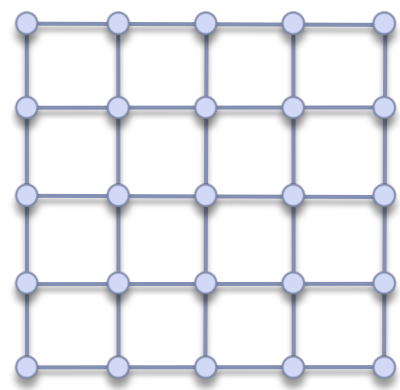
- 多数の数値列、データからの分布の推定
  - 生成モデルの存在の仮定

$$P(\mathbf{x}|\mathbf{u}) = \frac{1}{Z(\mathbf{u})}\exp\{-E(\mathbf{x}|\mathbf{u})\}$$

  - 例：Ising模型（グラフは任意）

$$E(\mathbf{x}|\mathbf{u}) = \sum_{i=1}^{N}\sum_{j\in\partial i} J_{ij}x_i x_j + \sum_{i=1}^{N} h_i x_i$$

- 利用例
  - 相関検出
    - カンニング検出、ニューロンシグナル、因子推定



ボルツマン機械学習

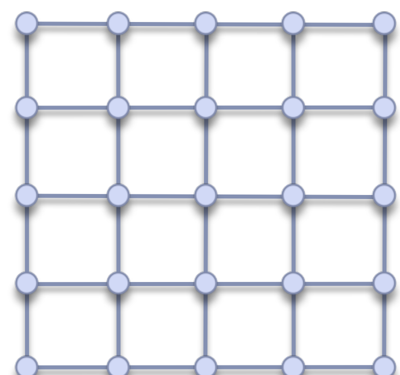
- 多数の数値列、データからの分布の推定
  - 生成モデルの存在の仮定

$$P(\mathbf{x}|\mathbf{u}) = \frac{1}{Z(\mathbf{u})}\exp\left\{-E(\mathbf{x}|\mathbf{u})\right\}$$

    - 例：Ising模型（グラフは任意）

$$E(\mathbf{x}|\mathbf{u}) = \sum_{i=1}^{N}\sum_{j\in\partial i} J_{ij}x_i x_j + \sum_{i=1}^{N} h_i x_i$$

  - 利用例
    - 相関検出
      - カンニング検出、ニューロンシグナル、因子推定
    - ハミルトニアン推定
      - 物理モデルの正当化、パラメータ推定



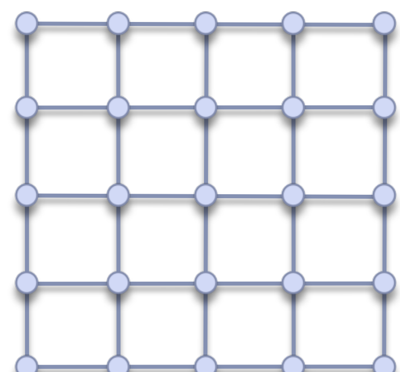
ボルツマン機械学習

- 多数の数値列、データからの分布の推定
  - 生成モデルの存在の仮定

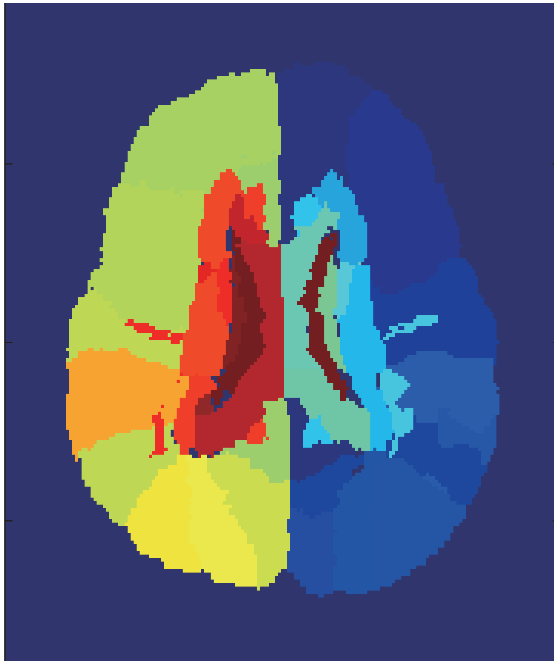
$$P(\mathbf{x}|\mathbf{u}) = \frac{1}{Z(\mathbf{u})}\exp\left\{-E(\mathbf{x}|\mathbf{u})\right\}$$

    - 例：Ising模型（グラフは任意）

$$E(\mathbf{x}|\mathbf{u}) = \sum_{i=1}^{N}\sum_{j\in\partial i} J_{ij}x_i x_j + \sum_{i=1}^{N} h_i x_i$$

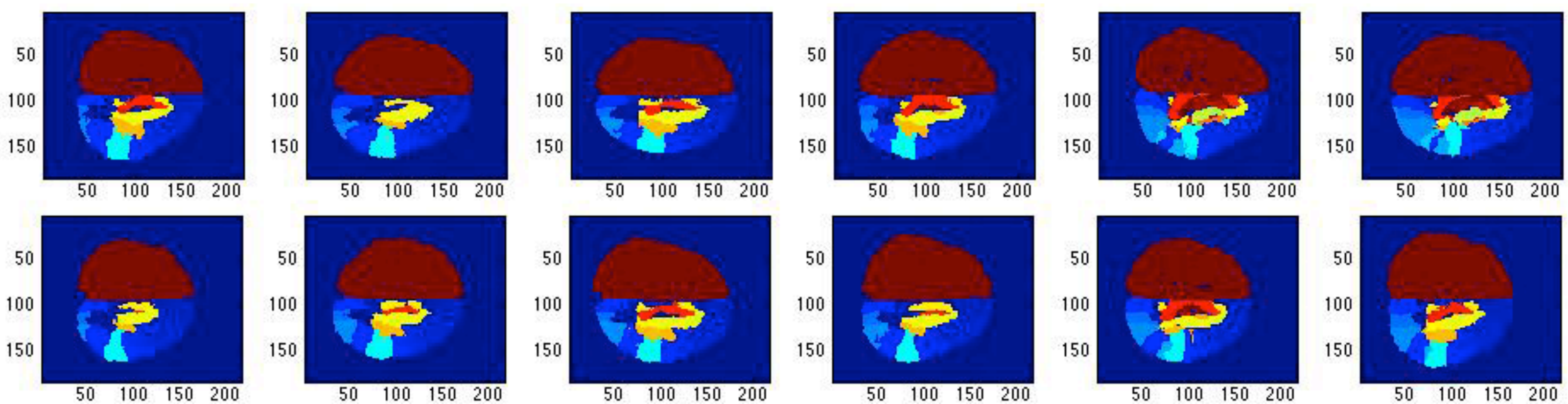


  - 利用例
    - 相関検出
      - カンニング検出、ニューロンシグナル、因子推定
    - ハミルトニアン推定
      - 物理モデルの正当化、パラメータ推定
    - 特徴抽出
      - 職人技の学習、自然の学習



専門的知識の学習

$$E(\mathbf{x}|\mathbf{u}) = \sum_{i=1}^{N} \sum_{j \in \partial i} J_{ij} \delta(x_i - x_j) + \sum_{i=1}^{N} h_i \delta(x_i - \tau_i)$$



脳画像ラベリングの例
Collaboration with U. Yamamoto

ボルツマン機械学習の基礎

- データとモデルがどれだけ一致しているか？
  - Kullback–Leibler divergence

$$D_{\mathrm{KL}}(P|Q) = \int d\mathbf{x} P(\mathbf{x}) \log\left(\frac{P(\mathbf{x})}{Q(\mathbf{x})}\right)$$

    - 非負
    - 三角不等式を満たさない、非対称

QUANTUM ANNEALING MACHINE LEARNING Sparse Modeling CREST

ボルツマン機械学習の基礎

- データとモデルがどれだけ一致しているか？
  - Kullback–Leibler divergence

$$D_{\mathrm{KL}}(P|Q) = \int d\mathbf{x} P(\mathbf{x}) \log\left(\frac{P(\mathbf{x})}{Q(\mathbf{x})}\right)$$

    - 非負
    - 三角不等式を満たさない、非対称

  - データの経験分布とモデルの分布を合わせる.
    - KL情報量の最小化

$$P_D(\mathbf{x}) = \frac{1}{D}\sum_{d=1}^{D}\delta(\mathbf{x}-\mathbf{x}^{(d)})$$

$$D_{\mathrm{KL}}(P_D|P_{\mathbf{u}}) = \int d\mathbf{x} P_D(\mathbf{x}) \log\left(\frac{P_D(\mathbf{x})}{P(\mathbf{x}|\mathbf{u})}\right)$$

ボルツマン機械学習の基礎

- 最尤法
  - データに対するモデルの尤もらしさ（対数尤度関数の経験平均）

$$L(\mathbf{u}) = \frac{1}{D}\sum_{d=1}^{D}\log P(\mathbf{x} = \mathbf{x}^{(d)}|\mathbf{u})$$

ボルツマン機械学習の基礎

- 最尤法
  - データに対するモデルの尤もらしさ（対数尤度関数の経験平均）

$$L(\mathbf{u}) = \frac{1}{D}\sum_{d=1}^{D}\log P(\mathbf{x}=\mathbf{x}^{(d)}|\mathbf{u})$$

  - 統計力学のエントロピー最大原理と等価

$$L(\mathbf{u}) = -\frac{1}{D}\sum_{d=1}^{D}E(\mathbf{x}=\mathbf{x}^{(d)}|\mathbf{u}) - \log Z(\mathbf{u})$$

ボルツマン機械学習の基礎

- 単純な勾配法の適用
  - 勾配方向に更新して最大値を求める

$$\mathbf{u}[t+1] = \mathbf{u}[t] + \eta \frac{\partial L(\mathbf{u})}{\partial \mathbf{u}}$$

  - η:学習係数、更新幅を設定している

- 単純な勾配法の適用
  - 勾配方向に更新して最大値を求める

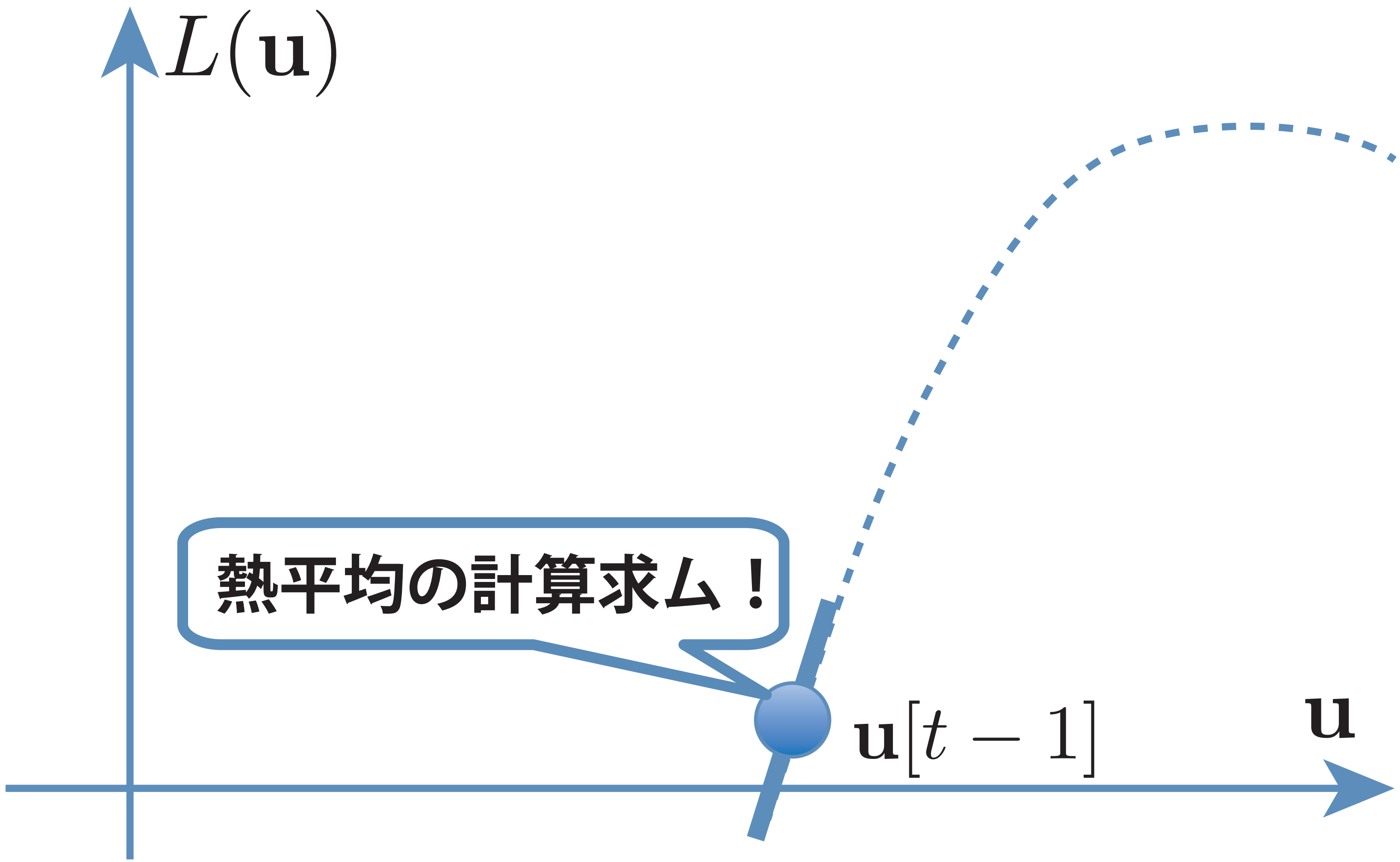
$$\mathbf{u}[t+1] = \mathbf{u}[t] + \eta \frac{\partial L(\mathbf{u})}{\partial \mathbf{u}}$$

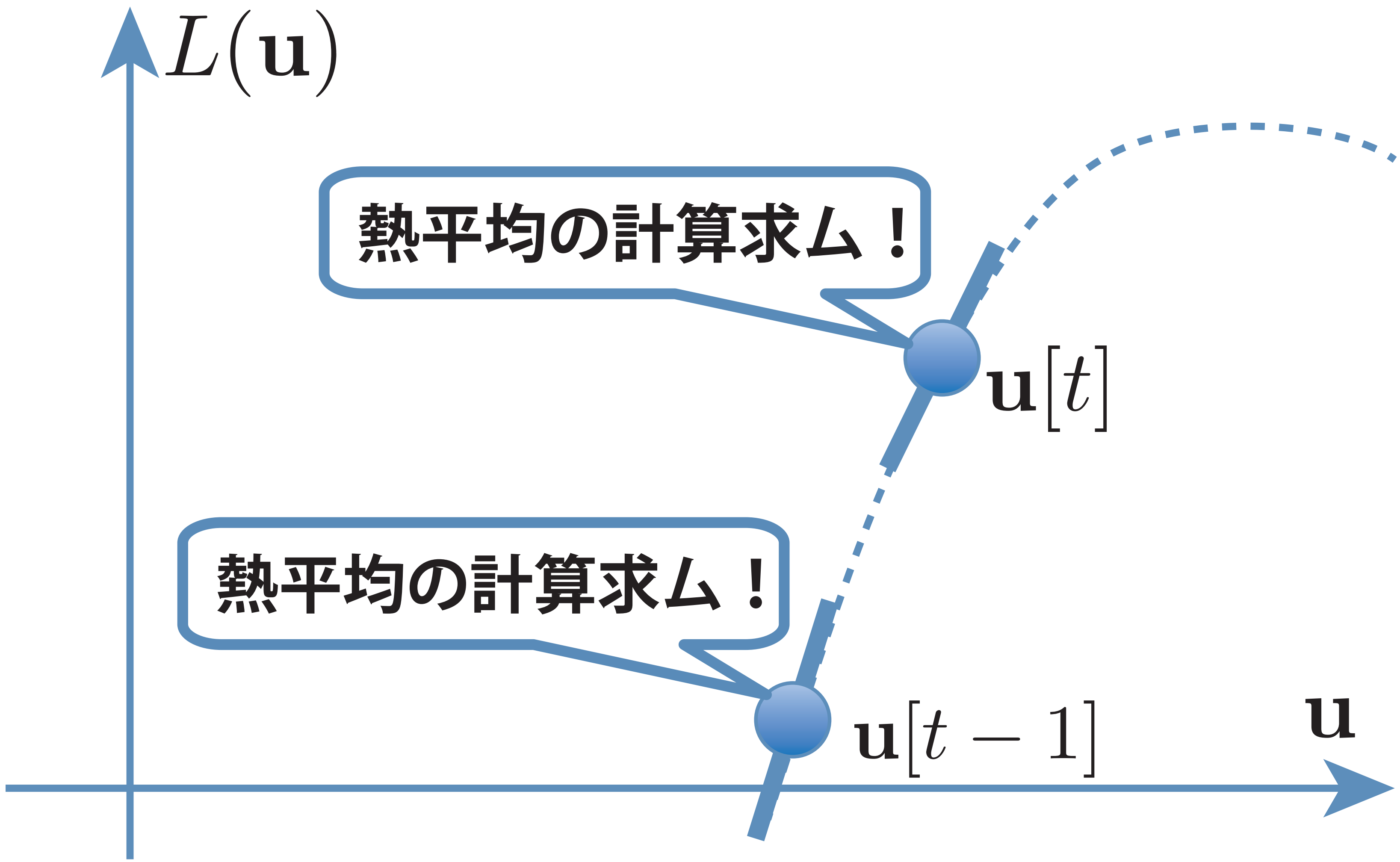
  - η:学習係数、更新幅を設定している

- じゃあ勾配を計算してみよう！

$$\frac{\partial L(\mathbf{u})}{\partial \mathbf{u}} = -\frac{1}{D}\sum_{d=1}^{D} \frac{\partial E(\mathbf{x}=\mathbf{x}^{(d)}|\mathbf{u})}{\partial \mathbf{u}} + \left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial \mathbf{u}} \right\rangle_{\mathbf{u}}$$

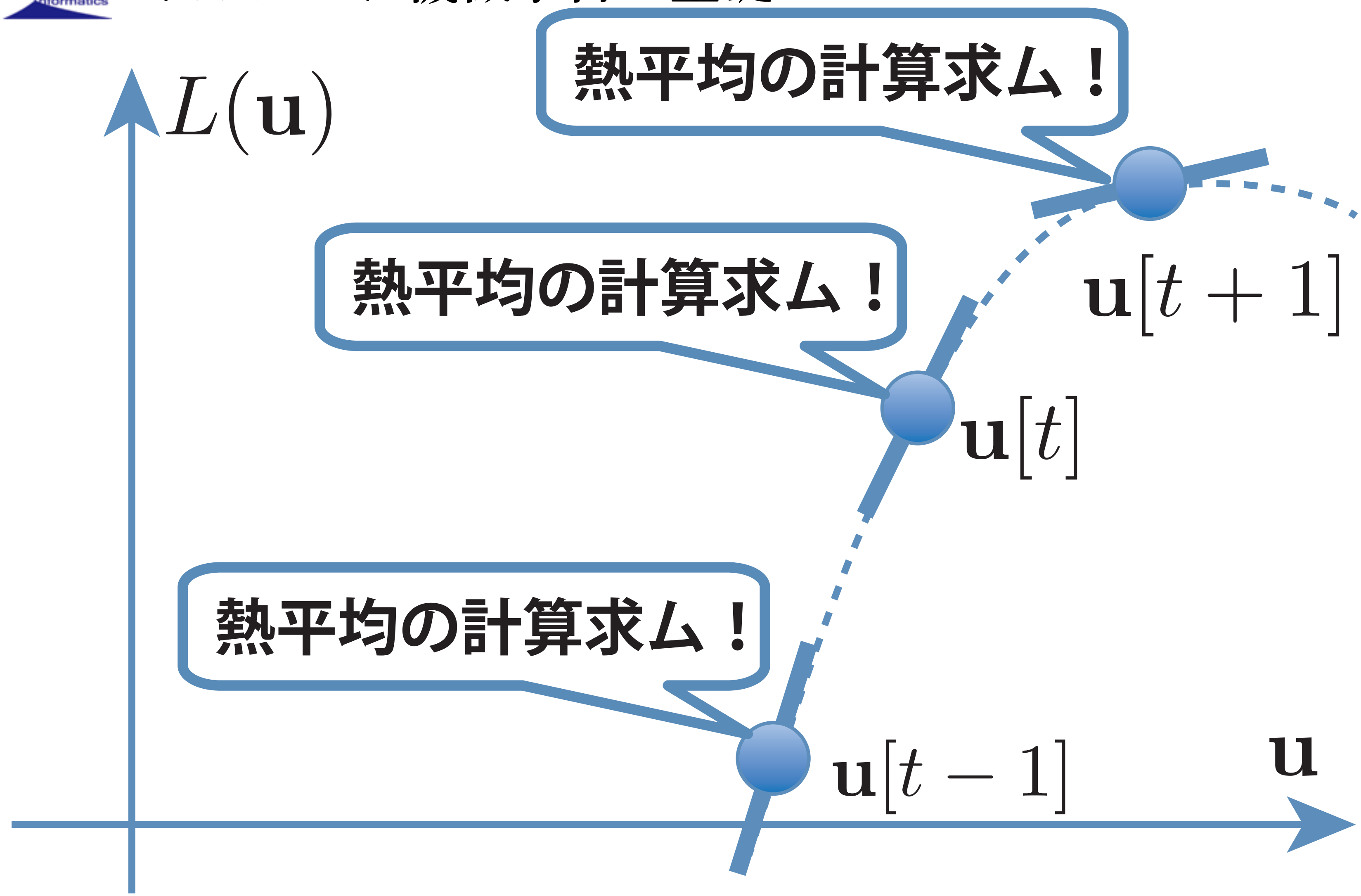
ボルツマン機械学習の基礎





$L(\mathbf{u})$
熱平均の計算求ム！
$\mathbf{u}[t]$
熱平均の計算求ム！
$\mathbf{u}[t-1]$
$\mathbf{u}$

ボルツマン機械学習の基礎



QUANTUM ANNEALING
MACHINE LEARNING
Sparse Modeling
CREST

$$\left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial \mathbf{u}} \right\rangle_{\mathbf{u}}$$

ボルツマン機械学習の危機

$$\log Z(\mathbf{u})$$

ボルツマン機械学習

- 様々な手法（統計力学）が存在
  - 期待値計算に統計力学由来の手法
    - 高速だけど近似法
      - 平均場近似やその発展法＝周囲の確率変数を期待値で置き換える

$$P(\mathbf{x}|\mathbf{u}) \approx \prod_{i=1}^{N} P(x_i|\mathbf{u})$$

QUANTUM ANNEALING MACHINE LEARNING Sparse Modeling CREST

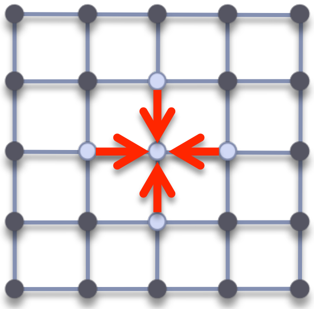
ボルツマン機械学習

- 様々な手法（統計力学）が存在
  - 期待値計算に統計力学由来の手法
    - 高速だけど近似法
      - 平均場近似やその発展法＝周囲の確率変数を期待値で置き換える

$$P(\mathbf{x}|\mathbf{u}) \approx \prod_{i=1}^{N} P(x_i|\mathbf{u})$$

      - 例：Ising模型

$$J_{ij}x_ix_j = J_{ij}x_i m_j$$



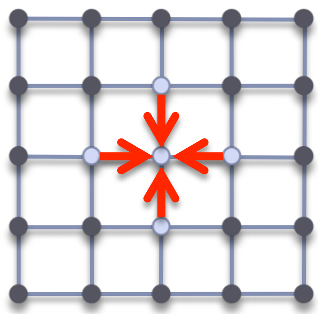
ボルツマン機械学習

- 様々な手法（統計力学）が存在
  - 期待値計算に統計力学由来の手法
    - 高速だけど近似法
      - 平均場近似やその発展法＝周囲の確率変数を期待値で置き換える

$$P(\mathbf{x}|\mathbf{u}) \approx \prod_{i=1}^{N} P(x_i|\mathbf{u})$$

      - 例：Ising模型

$$J_{ij}x_ix_j = J_{ij}x_i m_j$$



    - 低速だけど厳密
      - 分配関数計算、マルコフ連鎖モンテカルロ法

$$P(\mathbf{x}|\mathbf{u}) \approx \prod_{t=0}^{T-1} \{P_{\mathbf{u}}(\mathbf{x}_{t+1}|\mathbf{x}_t)\} P(\mathbf{x}_0)$$

統計力学の敗北

ボルツマン機械学習＝データ駆動型統計力学

- 様々な手法（統計力学）が存在＋計測技術（データ）の向上
  - 統計的機械学習
    - 疑似最尤法＝平均場近似?＋データの利用
      - 周囲を期待値ではなく、データの値そのものを入れる

$$P(\mathbf{x}|\mathbf{u}) \approx \prod_{i=1}^{N} P(x_i|\mathbf{x}_{/i}, \mathbf{u})$$

      - 例：Ising模型

$$J_{ij}x_ix_j = J_{ij}x_ix_j^{(d)}$$

    - 最小確率流法

＝確率過程＋データの利用：ちょっと動かして定常性を探る

$$\min_{\mathbf{u}} \left\{ \sum_{\mathbf{x}} P_D(\mathbf{x}) \log \frac{Q_{\mathbf{u}}(\mathbf{x})}{P_D(\mathbf{x})} \right\} \quad Q_{\mathbf{u}}(\mathbf{x}') = \sum_{\mathbf{x}} P_{\mathbf{u}}(\mathbf{x}'|\mathbf{x}) P_D(\mathbf{x})$$

機械学習

統計力学＋データ

機械学習

統計力学＋データ

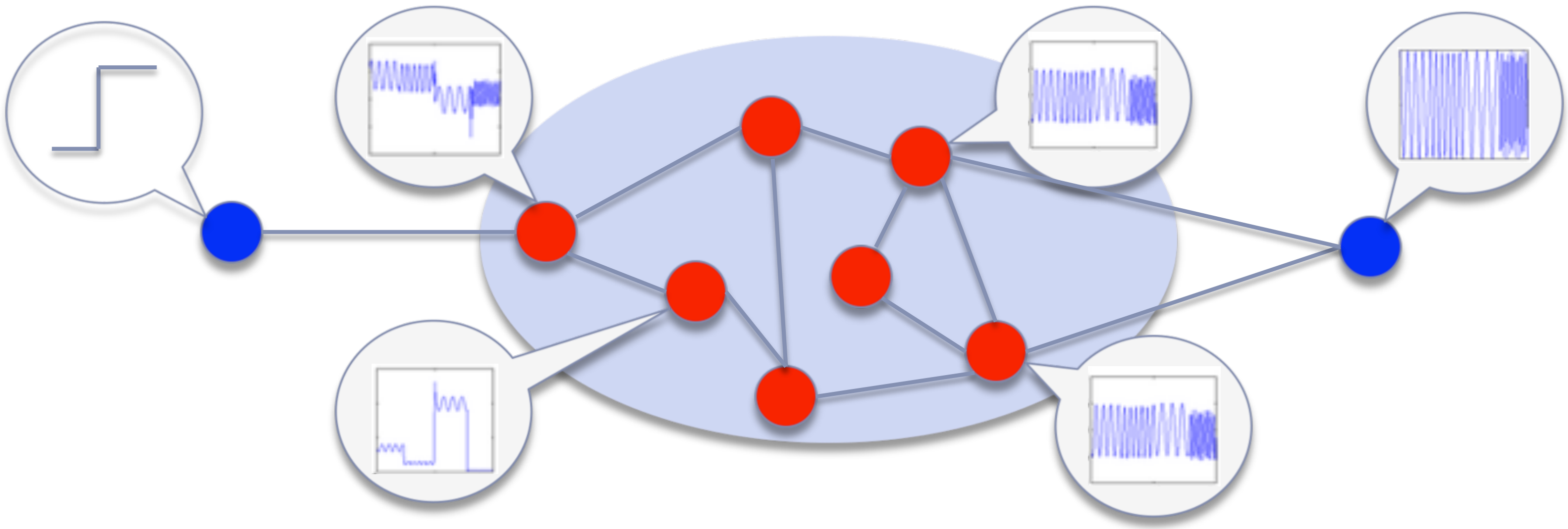
但し大量のデータが必要

非線形写像でデータを散らす

リザバーコンピューティング

リザバーコンピューティング

- 複雑だけど適応的学習はしないフレームワーク
  - リカレント（再帰型）ニューラルネットワークの利用
    - 重みはランダムに与えたままで固定（中間層の学習はしない）



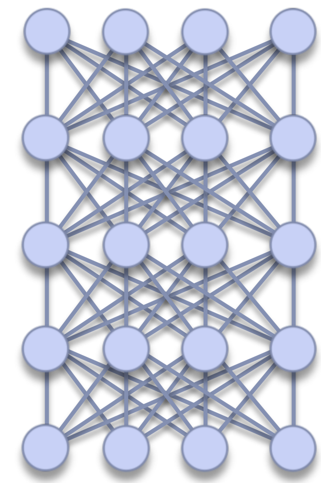
- 既存の非線形入出力関係を持つハードウェアの利用が可能
- 時系列の予測・判別
- データの増大をした上での線形識別

更なる複雑度への挑戦

深層学習

深層学習への挑戦

- 簡単に計算できるけど複雑なモデルを利用する
  - 隠れ変数有りボルツマン機械学習
    - それぞれの層の変数間には相互作用がない（条件付き独立性）



      - 例：Ising模型 $\sum_{k\in\partial i} J_{ik} v_i h_k = J_{ij}^{(\mathbf{h})} x_i x_j$
    - 物理の言葉で言うと実空間繰り込みで実効相互作用をもつ
    - コントラスティヴ・ダイヴァージェンス

      ＝（独立な）マルコフ連鎖モンテカルロ法＋データの利用
  - 大量の高次元データの利用=経験平均ですら大変
    - 確率勾配法＝確率過程＋データの利用

確率勾配法：最近の発展

M. Ohzeki: to appear soon in arXiv

- 大量の高次元データに対する勾配計算をさぼる
  - 確率勾配法（最尤法の近似）[Robbins and Monro (1951)]
    - ランダムに選択されたデータのみによる勾配を利用

$$\frac{\partial L(\mathbf{u})}{\partial \mathbf{u}} = -\frac{1}{d}\sum_{k=1}^{d}\frac{\partial E(\mathbf{x}=\mathbf{x}^{(k)}|\mathbf{u})}{\partial \mathbf{u}} + \left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial \mathbf{u}} \right\rangle_{\mathbf{u}}$$

  - 確率勾配Langevin法（サンプリング法へ） [Welling and The(2011)]
    - Langevin方程式を利用したノイズでパラメータ空間での平衡状態を生成

$$\mathbf{u}[t+1] = \mathbf{u}[t] + \eta\frac{\partial L(\mathbf{u})}{\partial \mathbf{u}} + dW$$

    - 詳細釣り合いの破れを利用して高速サンプリング法へと発展

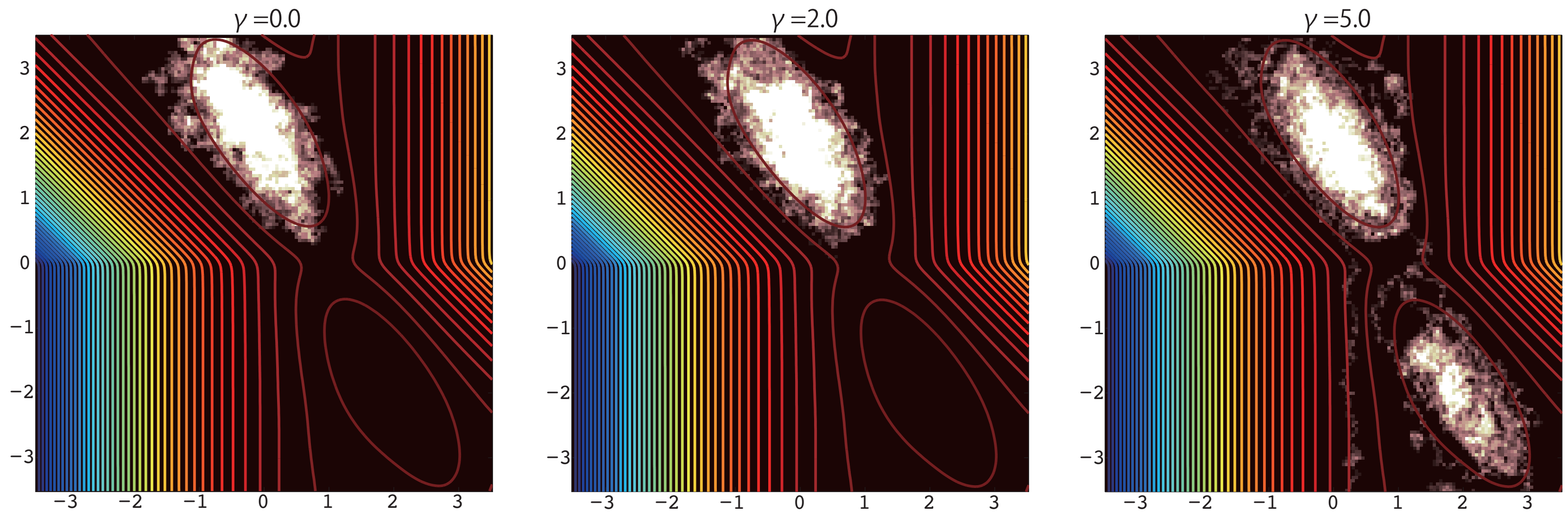
[Ohzeki and Ichiki (2015)]

QUANTUM ANNEALING MACHINE LEARNING Sparse Modeling CREST

確率勾配法：最近の発展

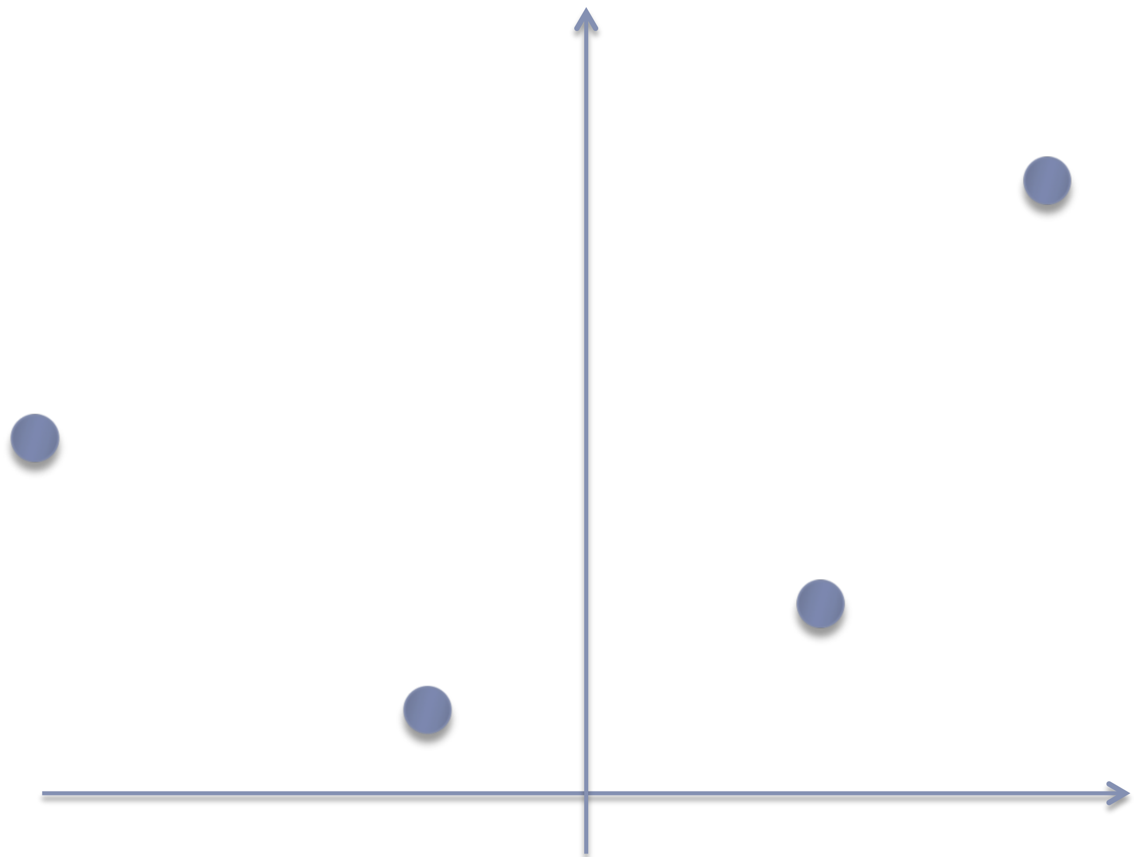
M. Ohzeki: to appear soon in arXiv

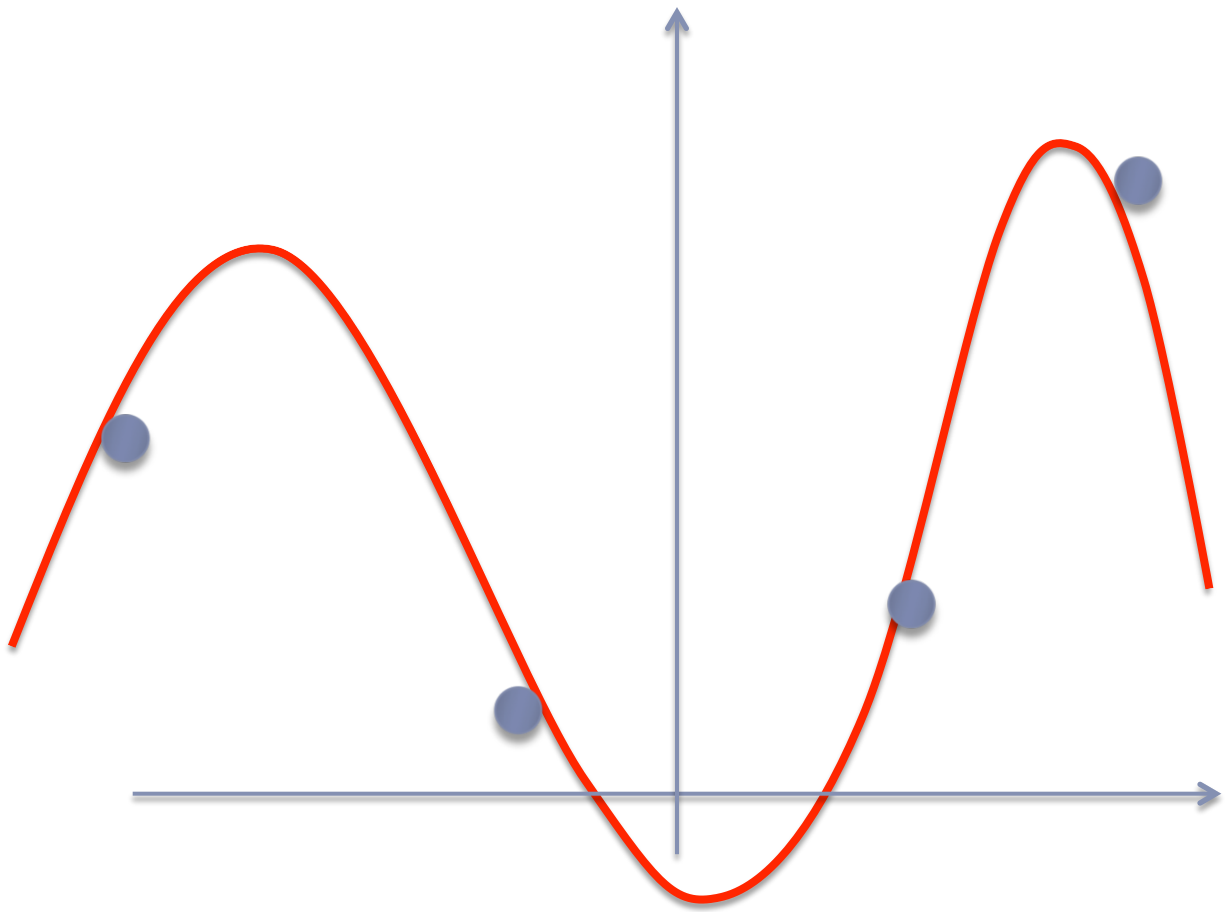
- 利用例：γ ＝詳細釣り合いの破れ度合い



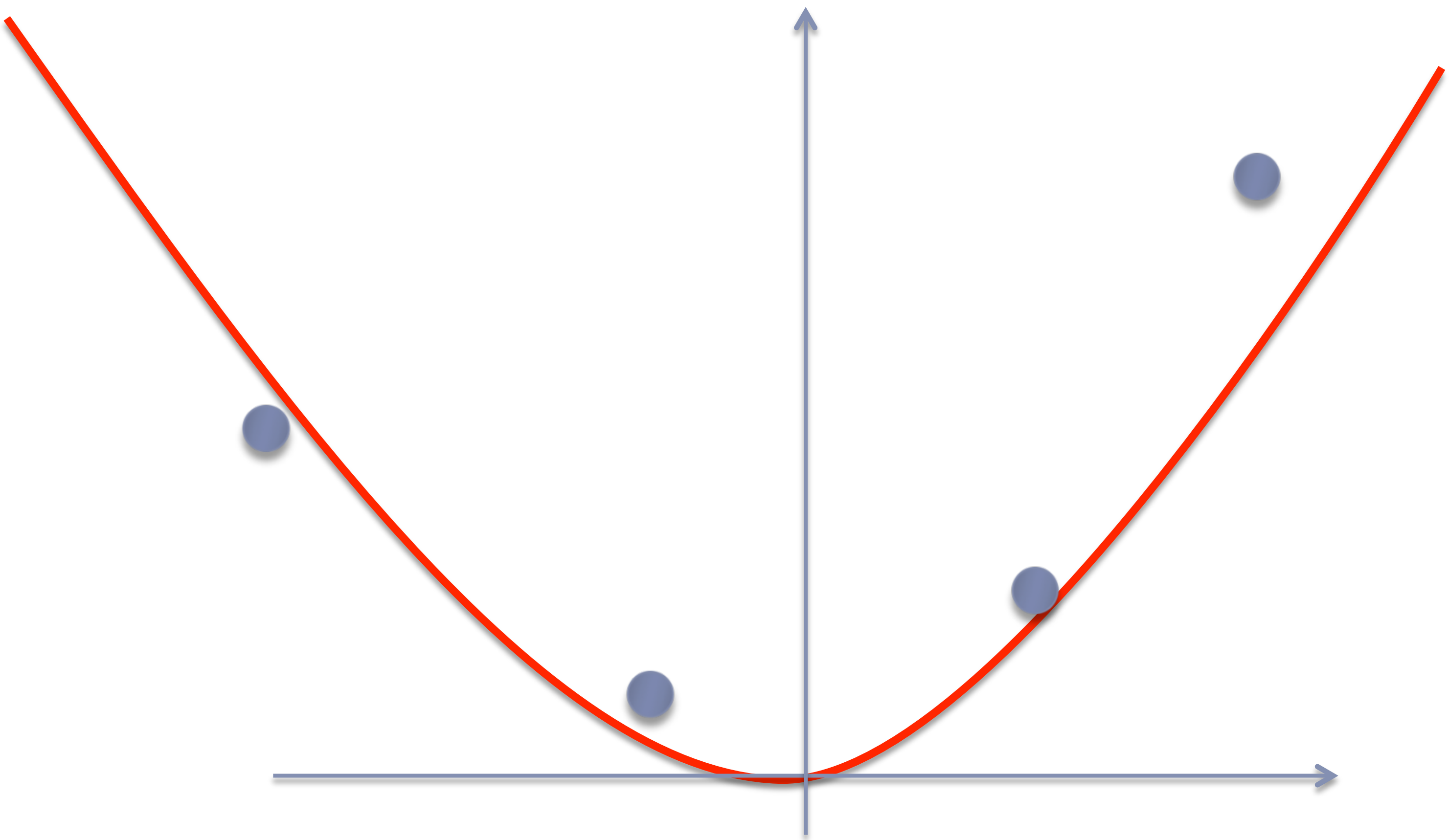
機械学習とは

データに対する関数の自律的フィッティング

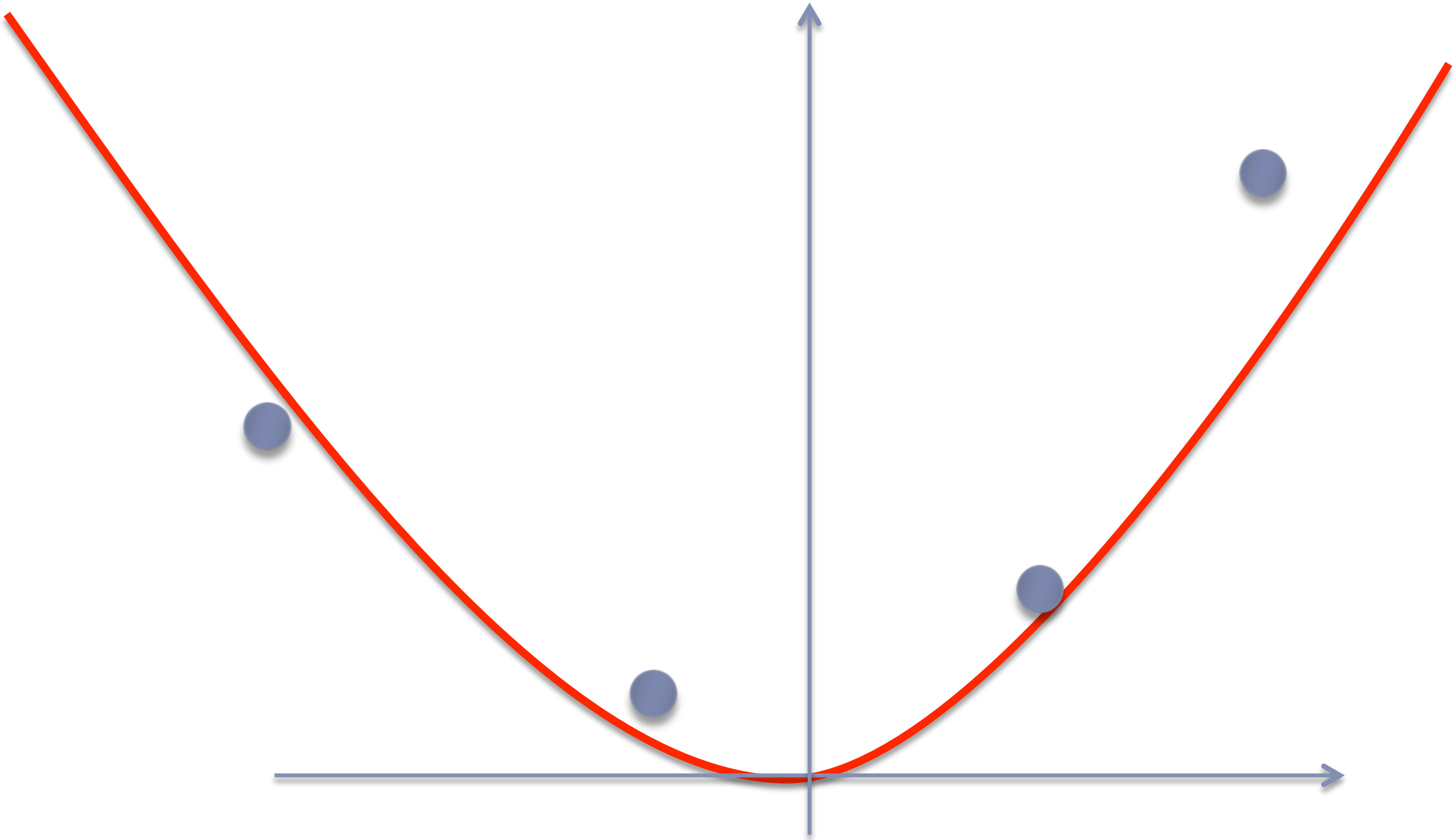




$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_N x^N$$



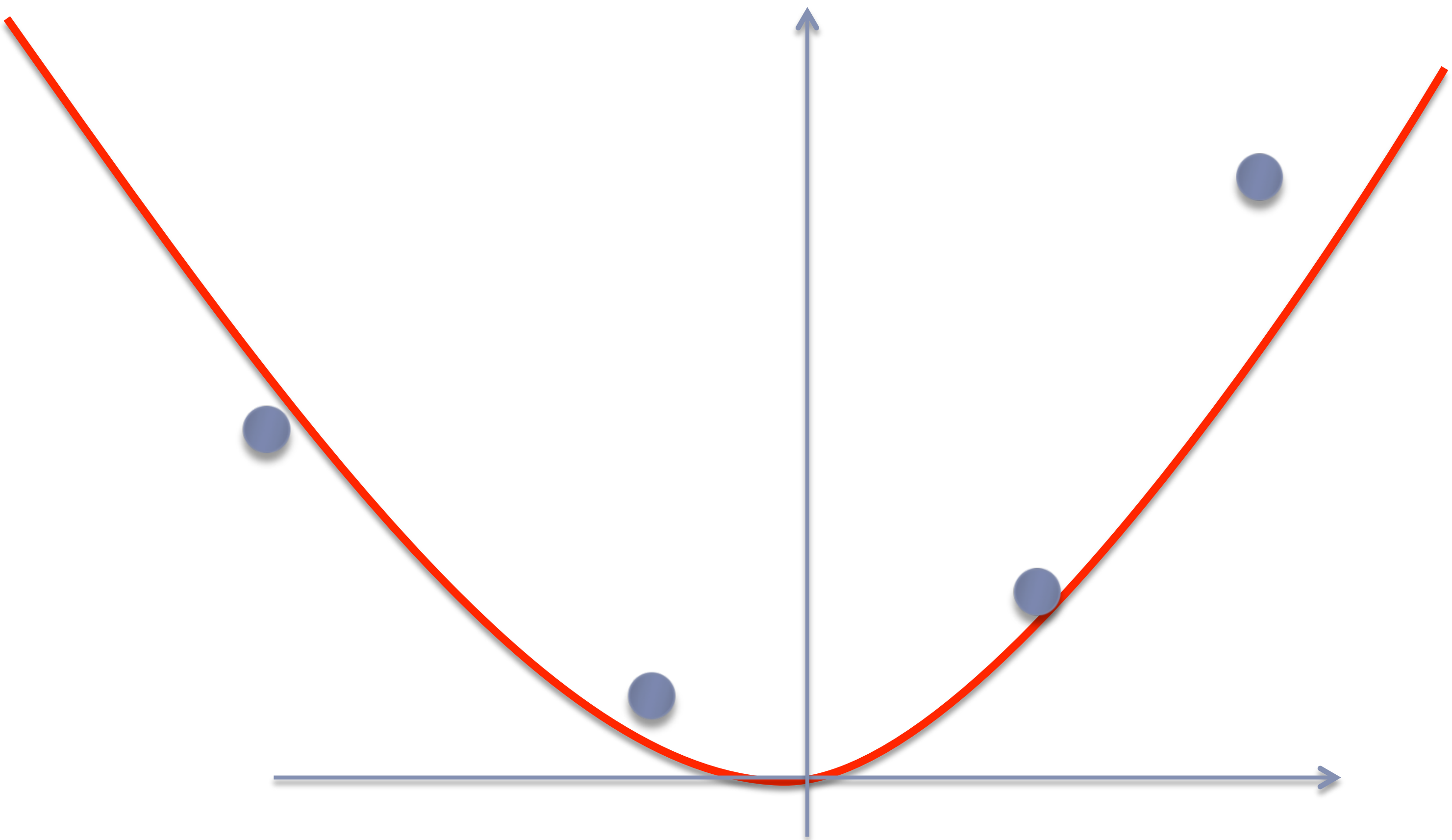
$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_N x^N$$



$$y = 0 + a_1 x + a_2 x^2 + 0 \times x^3 + \cdots$$

本質の抽出

スパースモデリング



$$\min_{a_i}\left\{\sum_{d=1}^{D}\left(y_d-\sum_{i=0}^{N}a_i x_d^i\right)^2+\lambda\left\|a_i\right\|_1\right\}$$

L1ノルムによる変数選択

M. Ohzeki: J. Phys. Soc. Jpn. 84, 054801 (2015)

- 基本は勾配法ベースの最尤法
  - コスト関数は様々に選べる
    - ボルツマン機械学習はデータが少ないと精度がでない
      - ビッグデータ時代だから大丈夫？
      - L1ノルムを使った正則化で解を選択

$$\min_{\mathbf{u}} \left\{ -L(\mathbf{u}) + \lambda \sum_{i<j} \| J_{ij} \|_1 \right\}$$

  - メジャライザー最小化で実行可能！

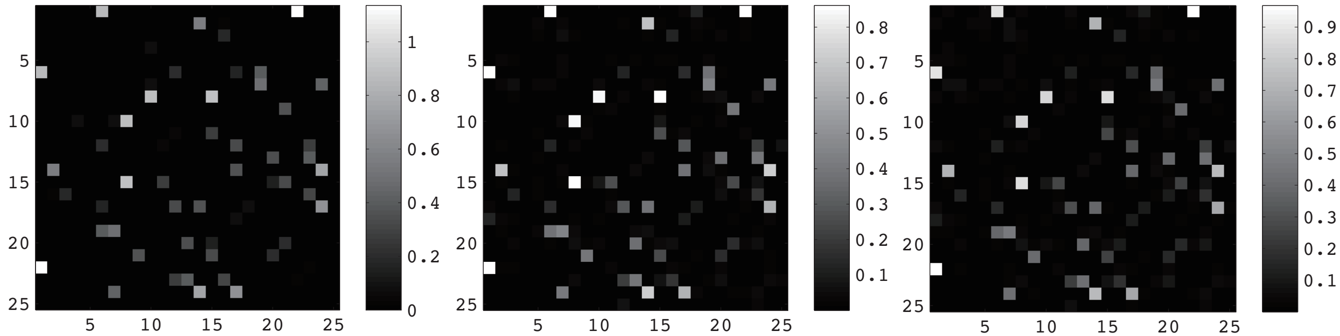
$$f(\mathbf{u}) = \|\mathbf{u}\|_1 \quad g(\mathbf{u}) = -L(\mathbf{u})$$

$$\mathbf{u}[t+1] = \arg\min_{\mathbf{u}} \left\{ f(\mathbf{u}) + (\nabla g(\mathbf{u}))^{\mathrm{T}} (\mathbf{u} - \mathbf{u}[t]) + \frac{L}{2} \|\mathbf{u} - \mathbf{u}[t]\|_2^2 \right\}$$

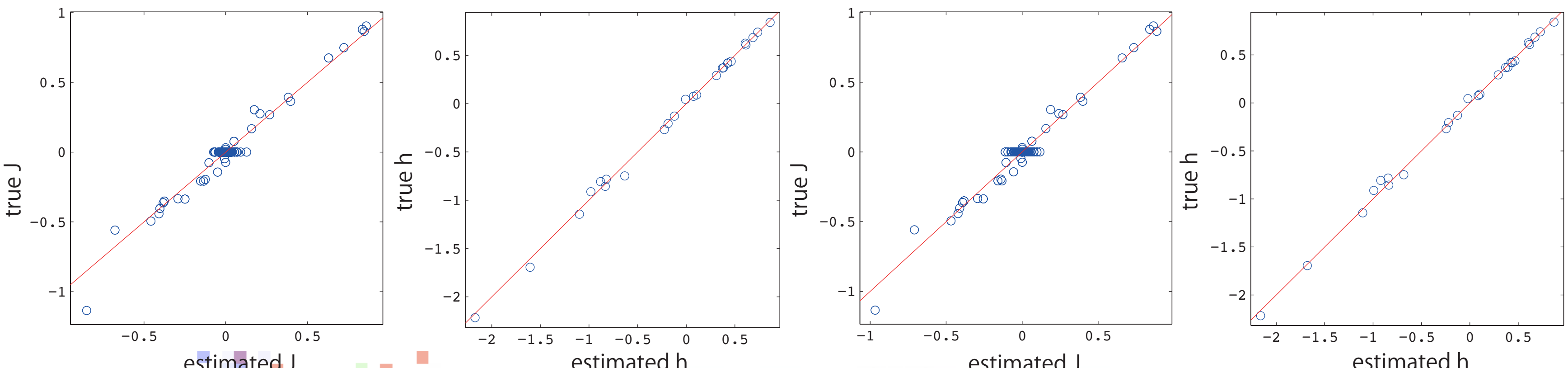
L1ノルムによる変数選択

M. Ohzeki: J. Phys. Soc. Jpn. 84, 054801 (2015)

- スパース相関推定：実験結果
  - N=25、D=5000
    - 相互作用に1割程度の非零成分（ガウスランダム）



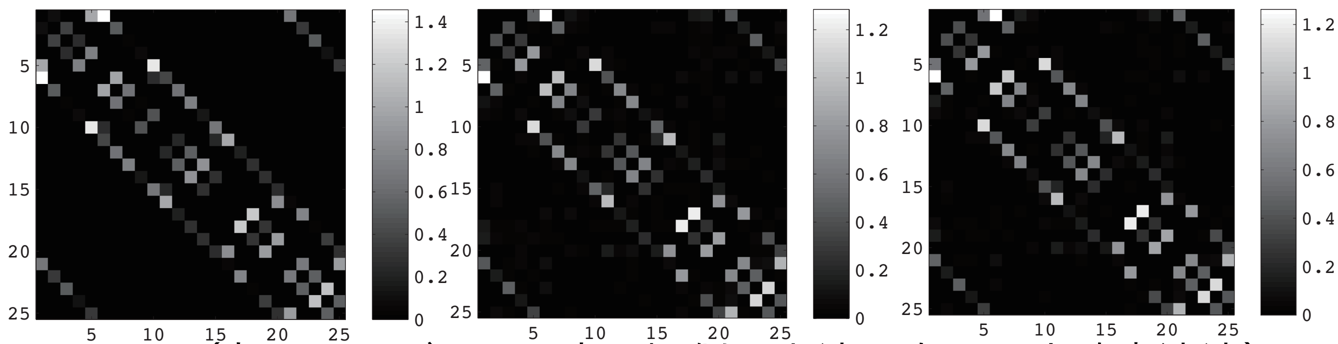
（左：オリジナル、中：疑似最尤法、右：最小確率流法）



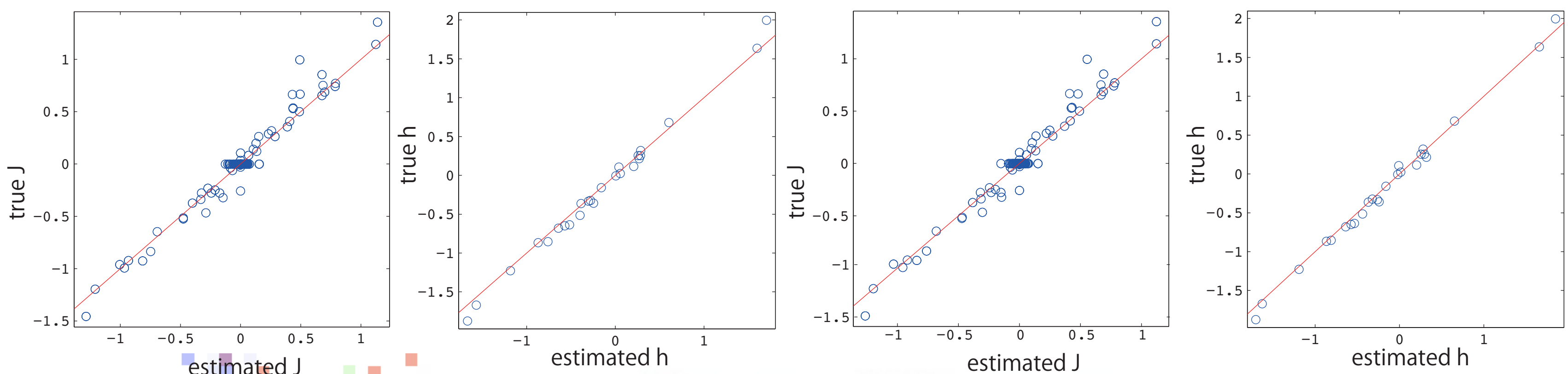
L1ノルムによる変数選択

M. Ohzeki: J. Phys. Soc. Jpn. 84, 054801 (2015)

- スパース相関推定：実験結果
  - N=25、D=5000
    - 相互作用を2次元最近接に限定、1割程度の非零成分（ガウスランダム）



（左：オリジナル、中：疑似最尤法、右：最小確率流法）



多くの答案から生徒の能力を明らかにする

項目応答理論

項目応答理論

F. Baker: Item response theory : parameter estimation techniques[2004]

- 試験結果により得られるデータからパラメータ推定
  - 各種導入実績
    - テストの品質管理
      - 情報処理技術者試験
      - TOEFL、TOEIC
      - 医療系資格試験、etc
    - アンケート結果からの顧客の指向調査

項目応答理論

F. Baker: Item response theory : parameterestimation techniques[2004]

- 1-パラメータロジスティクモデル
  - テストの品質管理に用いられる確率論的模型

$$P(\mathbf{x}|\boldsymbol{\theta},\mathbf{d}) = \frac{1}{Z(\boldsymbol{\theta},\mathbf{d})}\prod_{j=1}^{J}\exp\left\{\sum_{i=1}^{I}(\theta_i - d_j)x_{ij}\right\},$$

- $x_{ij} = \pm 1$: i 番目の被験者が j 番目の項目に正答したかどうか

  正答の場合 $x_{ij} = 1$ 誤答の場合 $x_{ij} = -1$

- θ_i: i 番目の被験者の能力
- d_j: j 番目の項目の難易度

ボルツマン機械学習との対応

- カンニングを想定した拡張
  - 被験者間の相関を考慮する

$$P(\mathbf{x}|\boldsymbol{\theta},\mathbf{d},\mathbf{w}) = $$

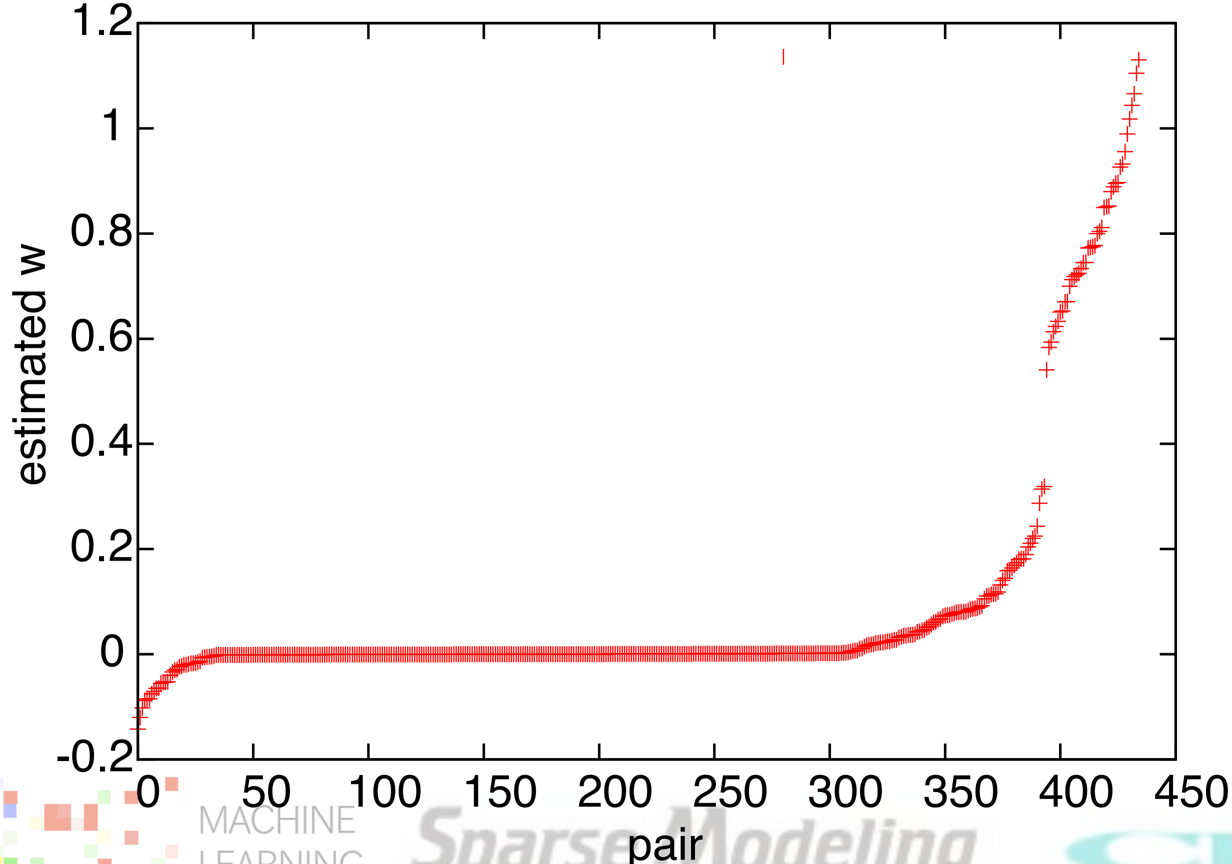
$$\prod_{j=1}^{J}\frac{1}{Z_j}\prod_{i=1}^{I}\exp\left\{(\theta_i - d_j)x_{ij} + \sum_{k\in\partial(i)} w_{ik}x_{ij}x_{kj}\right\}.$$

  - カンニング係数の導入
    - w_{ik}: 被験者 i と被験者 k の協力関係
  - 今回はレポート問題（全結合）で教え合う（対称的）状況を仮定

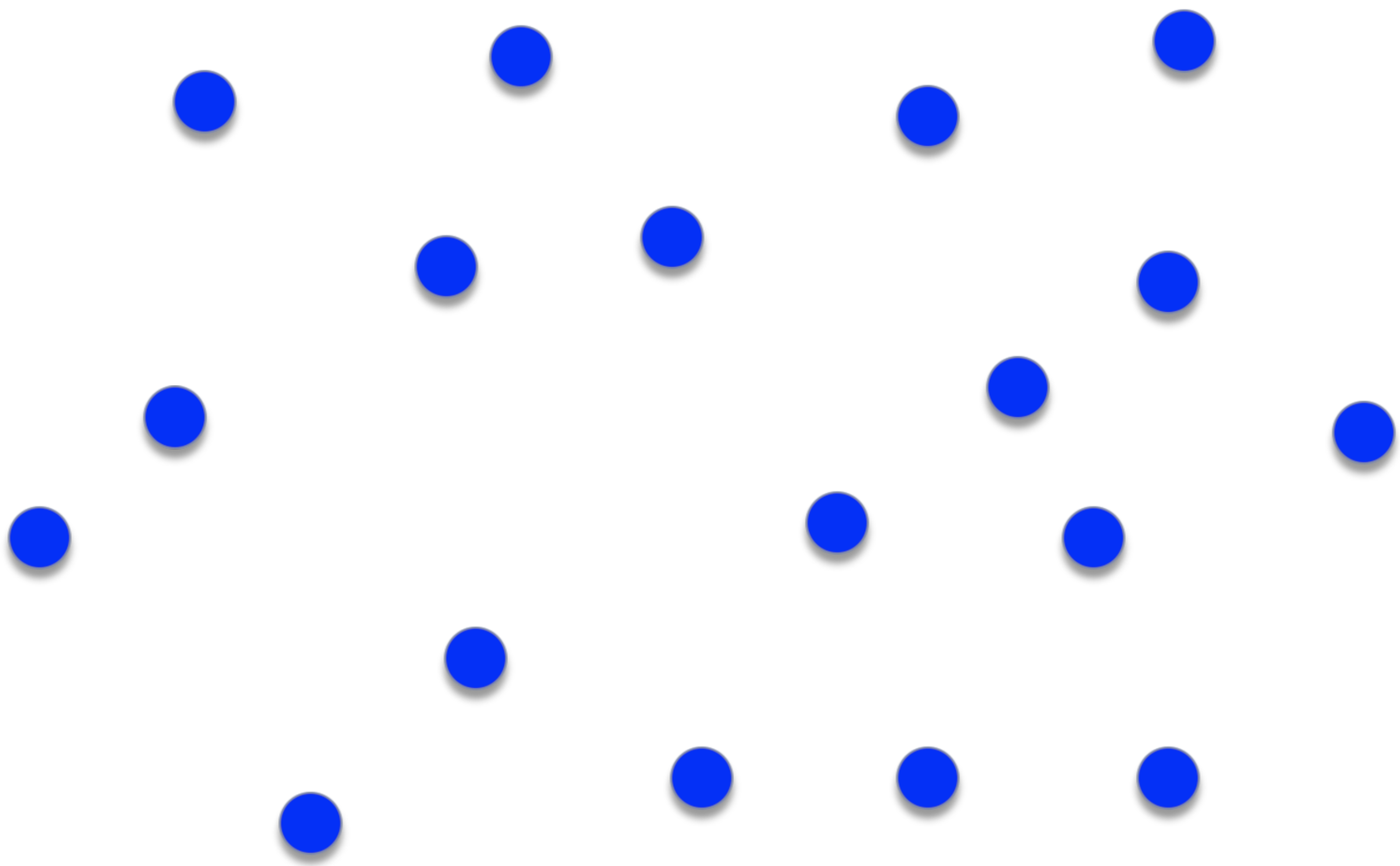
あらゆる関係性を考えればよいか？

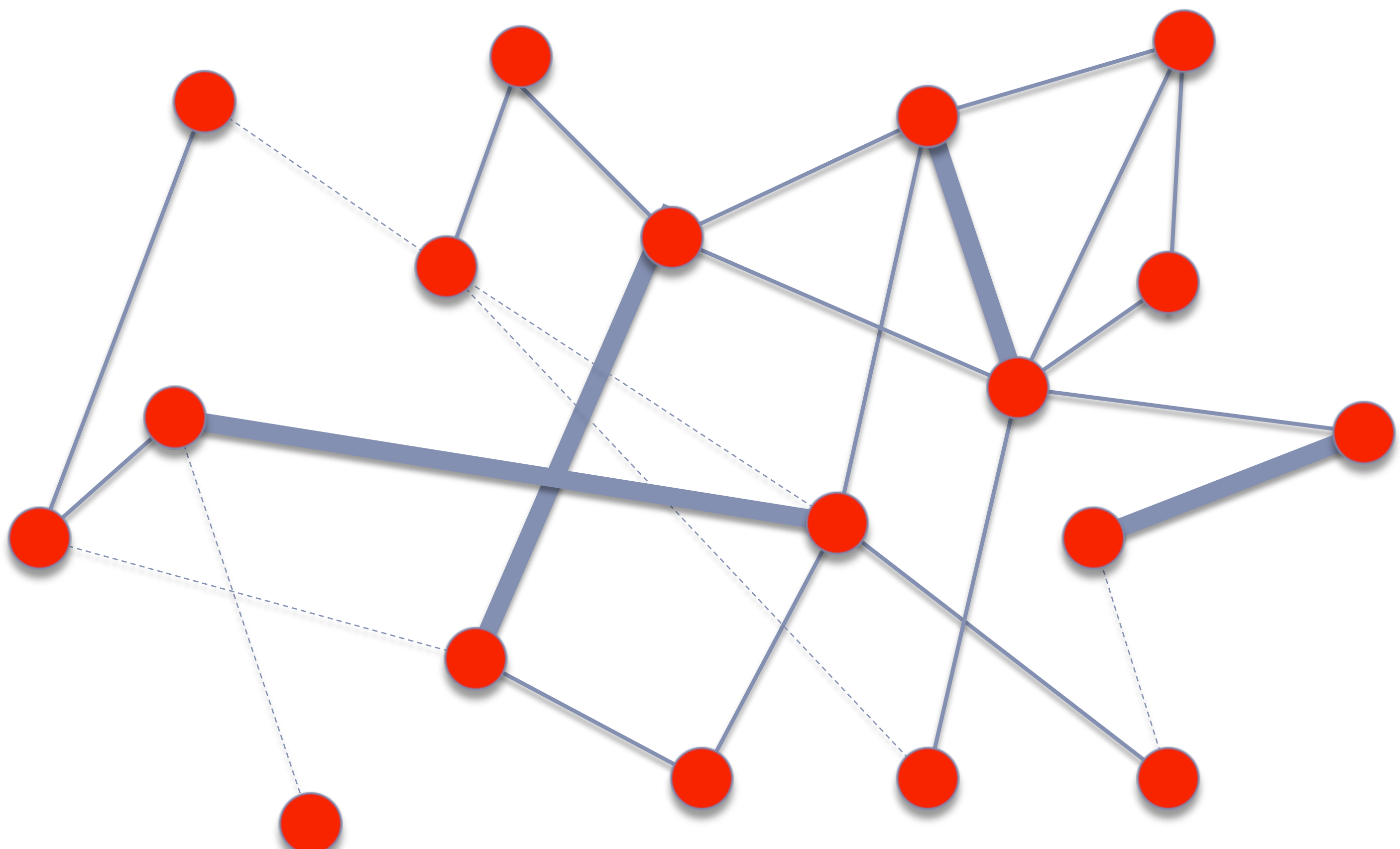
L1ノルムによる推定結果

- L1ノルムによるスパース解推定
  - 問題数J=1000のとき

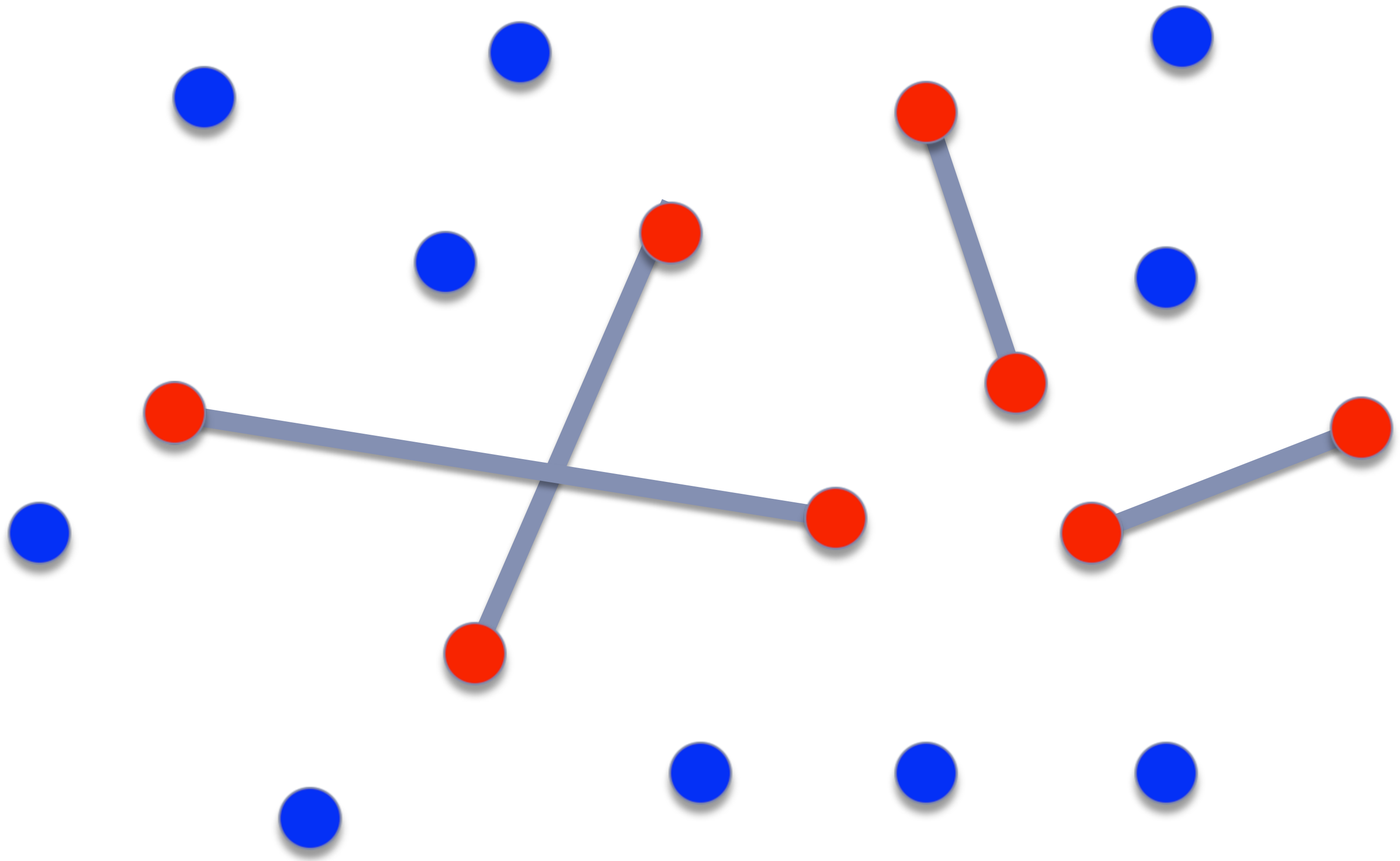


QUANTUM ANNEALING MACHINE LEARNING SparseModeling CREST





全員カンニング？！



理想的なカンニング

あるかないかはっきり言ってほしい！

あるかないかはっきり言ってほしい！
L0ノルムによる正則化

あるかないかはっきり言ってほしい！

L0ノルムによる正則化

組み合わせ最適化問題？？

あるかないかはっきり言ってほしい！

L0ノルムによる正則化

~~組み合わせ最適化問題？？~~

近似解法（貪欲法）

はっきりと零・非零を分けた推定：貪欲法の利用

A. Decelle and F. Ricci-Tersenghi: Phys. Rev. Lett. 112 (2014) 070603.

- デシメーションアルゴリズムの導入
  - 基本は尤度関数（うまく説明している度合い）の最大化

$$\max_{\mathbf{w},\theta,\mathbf{d}}\left\{\mathcal{L}(\mathbf{w},\theta,\mathbf{d})\right\}$$

QUANTUM ANNEALING MACHINE LEARNING Sparse Modeling CREST

はっきりと零・非零を分けた推定：貪欲法の利用

A. Decelle and F. Ricci-Tersenghi: Phys. Rev. Lett. 112 (2014) 070603.

- デシメーションアルゴリズムの導入
  - 基本は尤度関数（うまく説明している度合い）の最大化

$$\max_{\mathbf{w},\theta,\mathbf{d}}\left\{\mathcal{L}(\mathbf{w},\theta,\mathbf{d})\right\}$$

1. まず最尤推定を行う.

$$\mathbf{w} = (w_{12}, w_{23}, w_{24}, \cdots, w_{I-1,I})$$

はっきりと零・非零を分けた推定：貪欲法の利用

A. Decelle and F. Ricci-Tersenghi: Phys. Rev. Lett. 112 (2014) 070603.

- デシメーションアルゴリズムの導入
  - 基本は尤度関数（うまく説明している度合い）の最大化

$$\max_{\mathbf{w},\theta,\mathbf{d}}\left\{\mathcal{L}(\mathbf{w},\theta,\mathbf{d})\right\}$$

1. まず最尤推定を行う.

$$\mathbf{w} = (w_{12}, w_{23}, w_{24}, \cdots, w_{I-1,I})$$

2. 推定結果において、絶対値が小さいものを消す.

$$\mathbf{w} = (\boxed{0.002}, 0.78, \boxed{0.0001}, \cdots, 0.3)$$

はっきりと零・非零を分けた推定：貪欲法の利用

A. Decelle and F. Ricci-Tersenghi: Phys. Rev. Lett. 112 (2014) 070603.

- デシメーションアルゴリズムの導入
  - 基本は尤度関数（うまく説明している度合い）の最大化

$$\max_{\mathbf{w},\theta,\mathbf{d}}\left\{\mathcal{L}(\mathbf{w},\theta,\mathbf{d})\right\}$$

1. まず最尤推定を行う.

$$\mathbf{w} = (w_{12}, w_{23}, w_{24}, \cdots, w_{I-1,I})$$

2. 推定結果において、絶対値が小さいものを消す.

$$\mathbf{w} = (\boxed{0.002}, 0.78, \boxed{0.0001}, \cdots, 0.3)$$

3. 残った非零サポートのみで最尤推定を行う.

$$\mathbf{w} = (0, w_{23}, 0, \cdots, w_{I,I-1})$$

A. Decelle and F. Ricci-Tersenghi: Phys. Rev. Lett. 112 (2014) 070603.

- デシメーションアルゴリズムの導入
  - 基本は尤度関数（うまく説明している度合い）の最大化

$$\max_{\mathbf{w},\theta,\mathbf{d}}\left\{\mathcal{L}(\mathbf{w},\theta,\mathbf{d})\right\}$$

1. まず最尤推定を行う.

$$\mathbf{w} = (w_{12}, w_{23}, w_{24}, \cdots, w_{I-1,I})$$

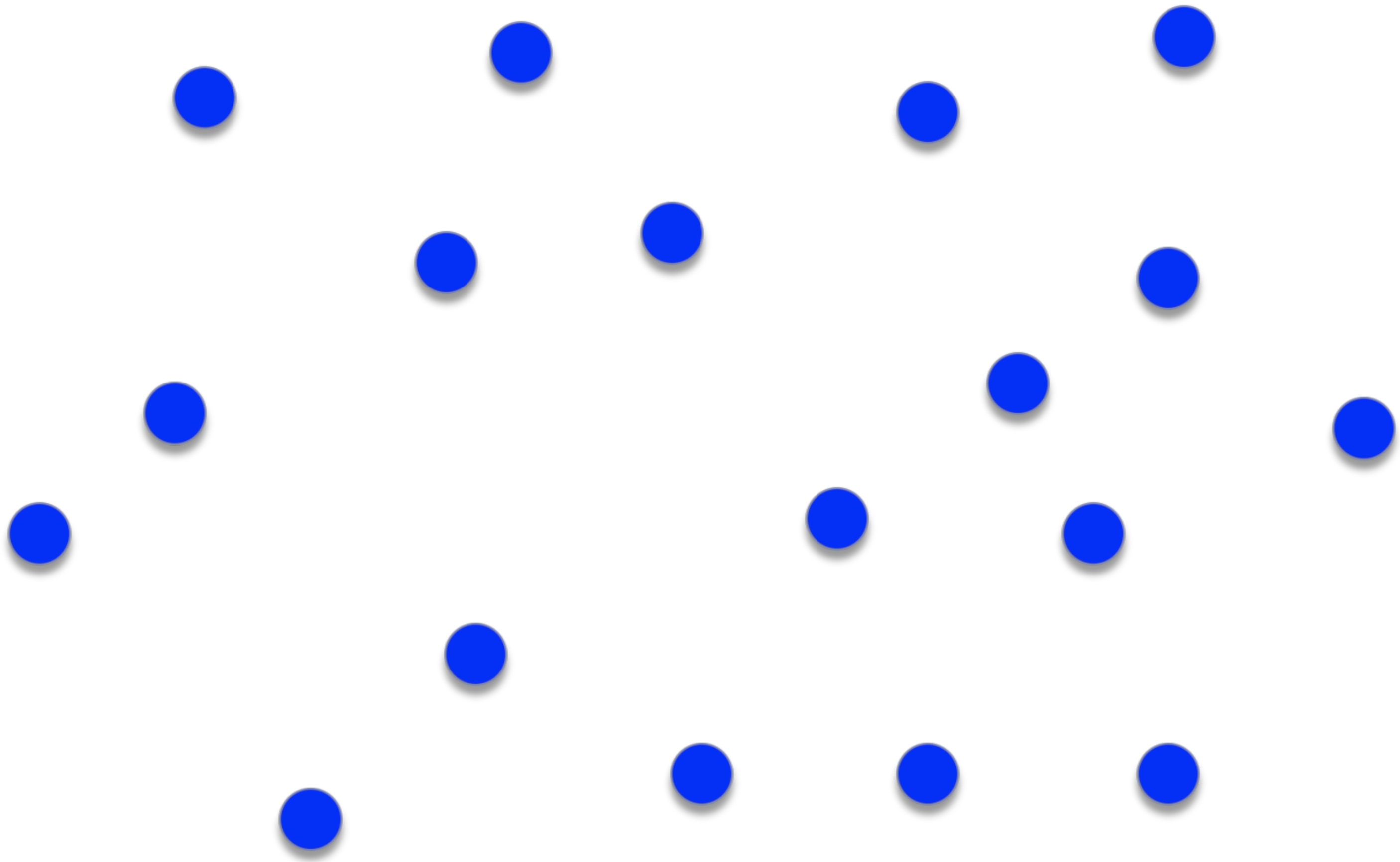
2. 推定結果において、絶対値が小さいものを消す.

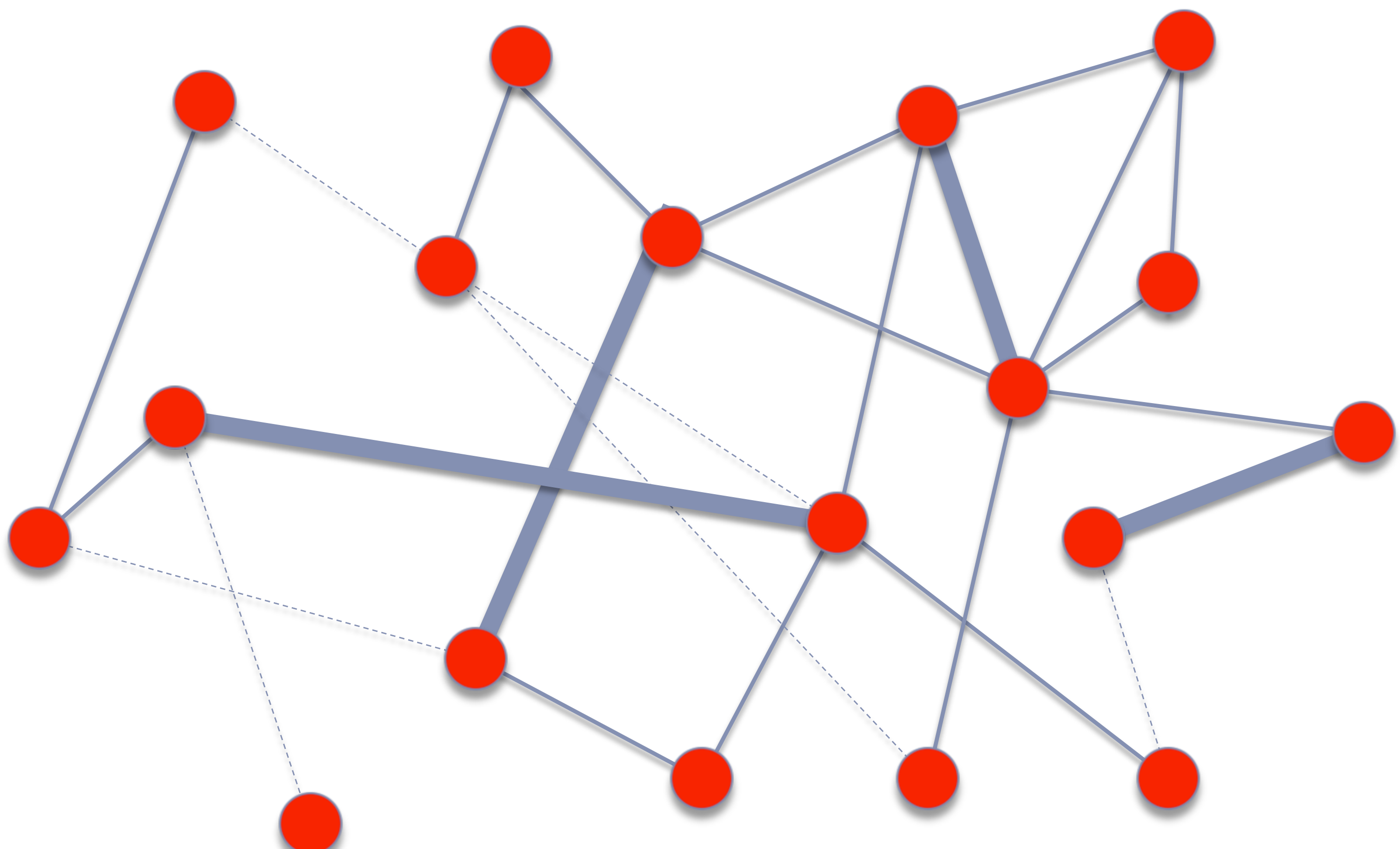
$$\mathbf{w} = (\boxed{0.002}, 0.78, \boxed{0.0001}, \cdots, 0.3)$$

3. 残った非零サポートのみで最尤推定を行う.

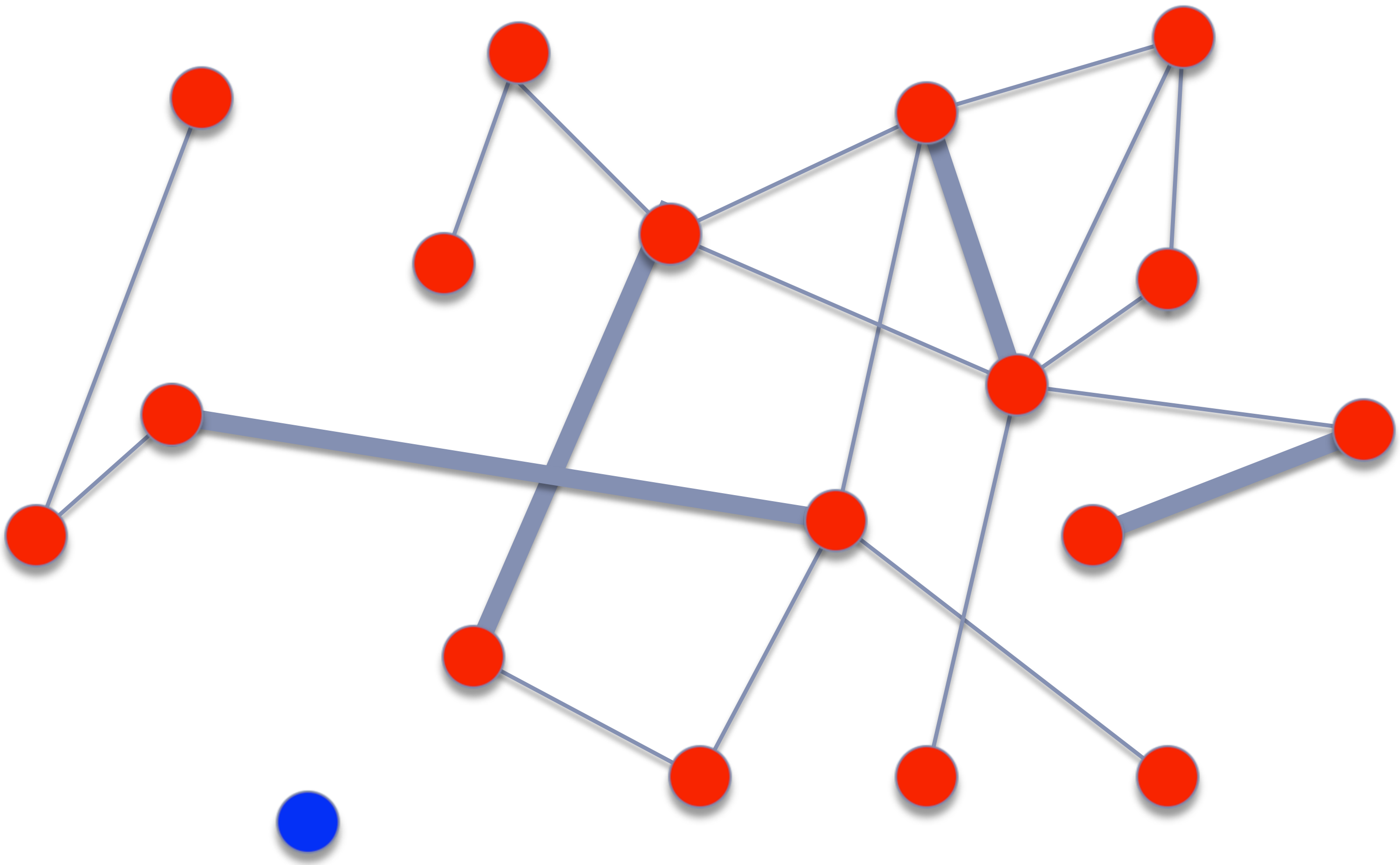
$$\mathbf{w} = (0, w_{23}, 0, \cdots, w_{I,I-1})$$

- 利点
  - 零となるところをはっきりと零と指定する

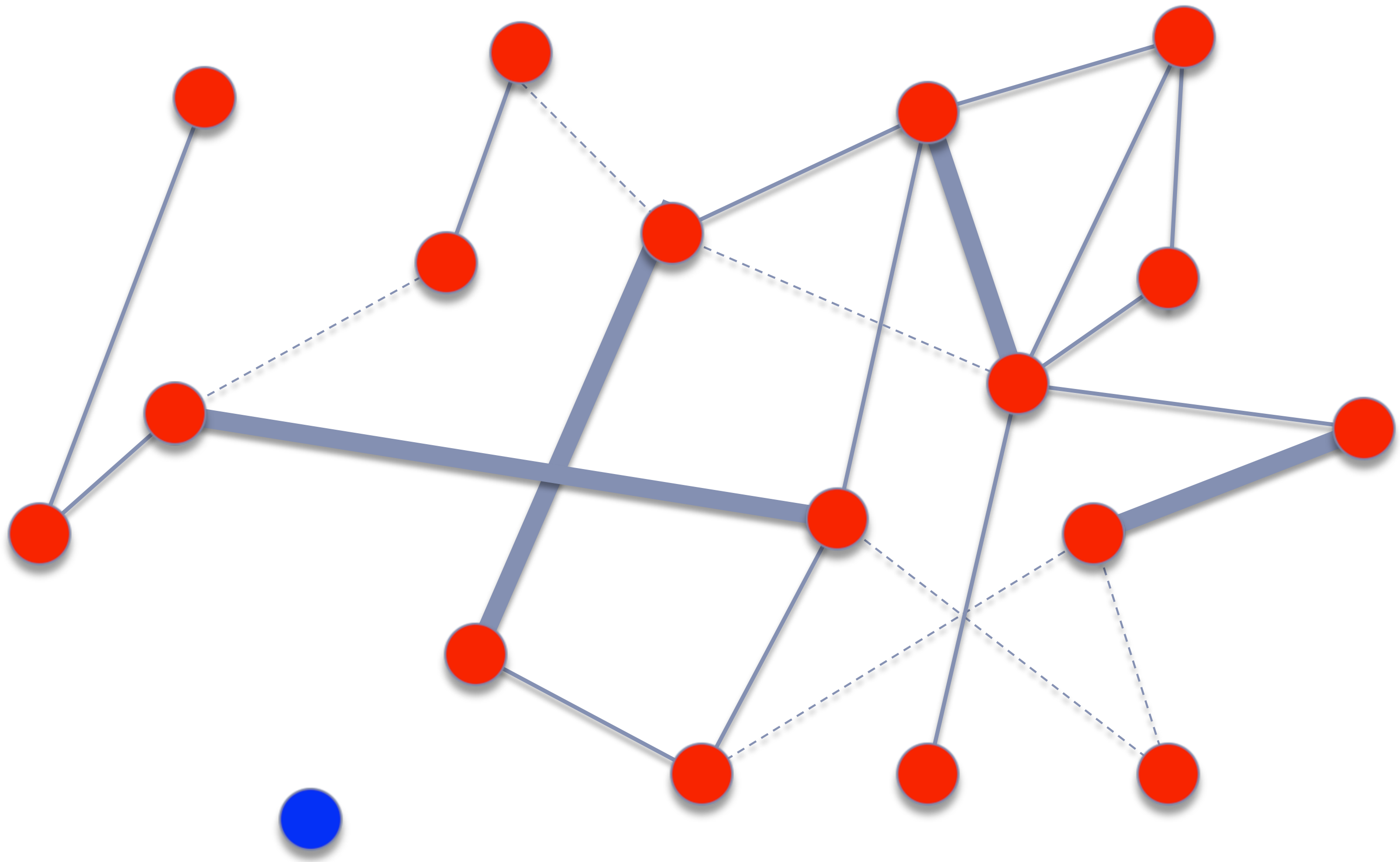




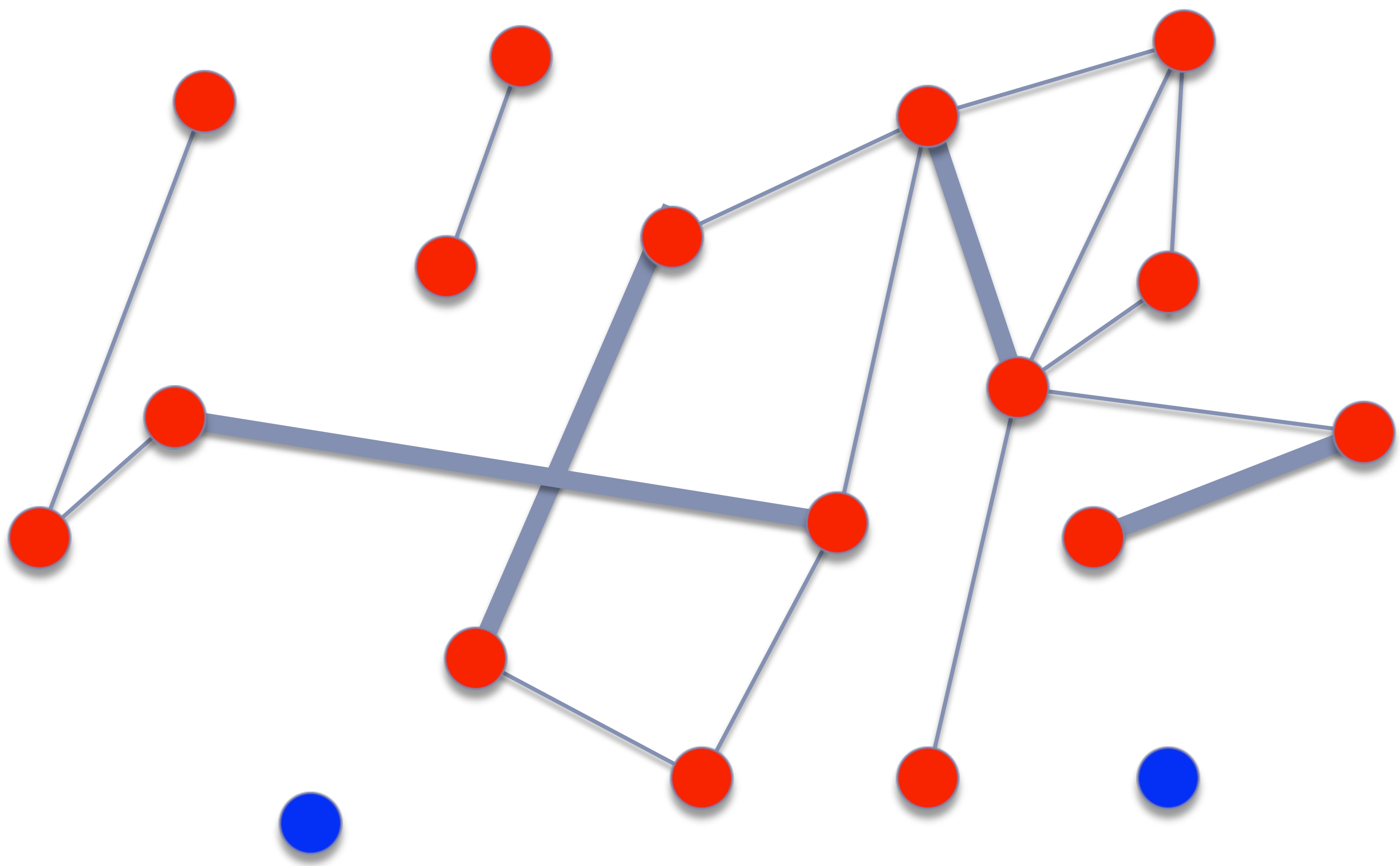
従来の推定法（性悪説？！）



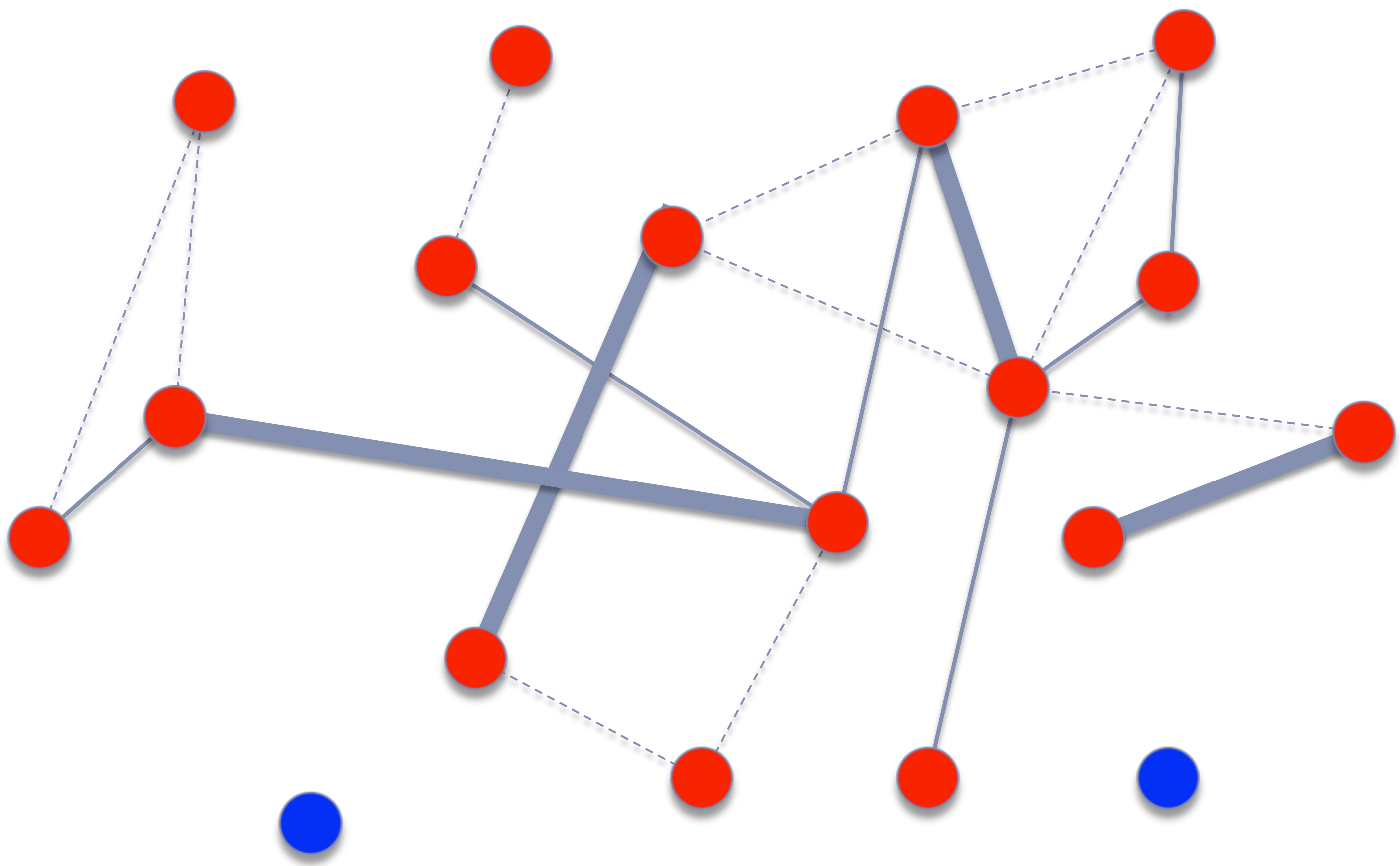
デシメーション（性善説）



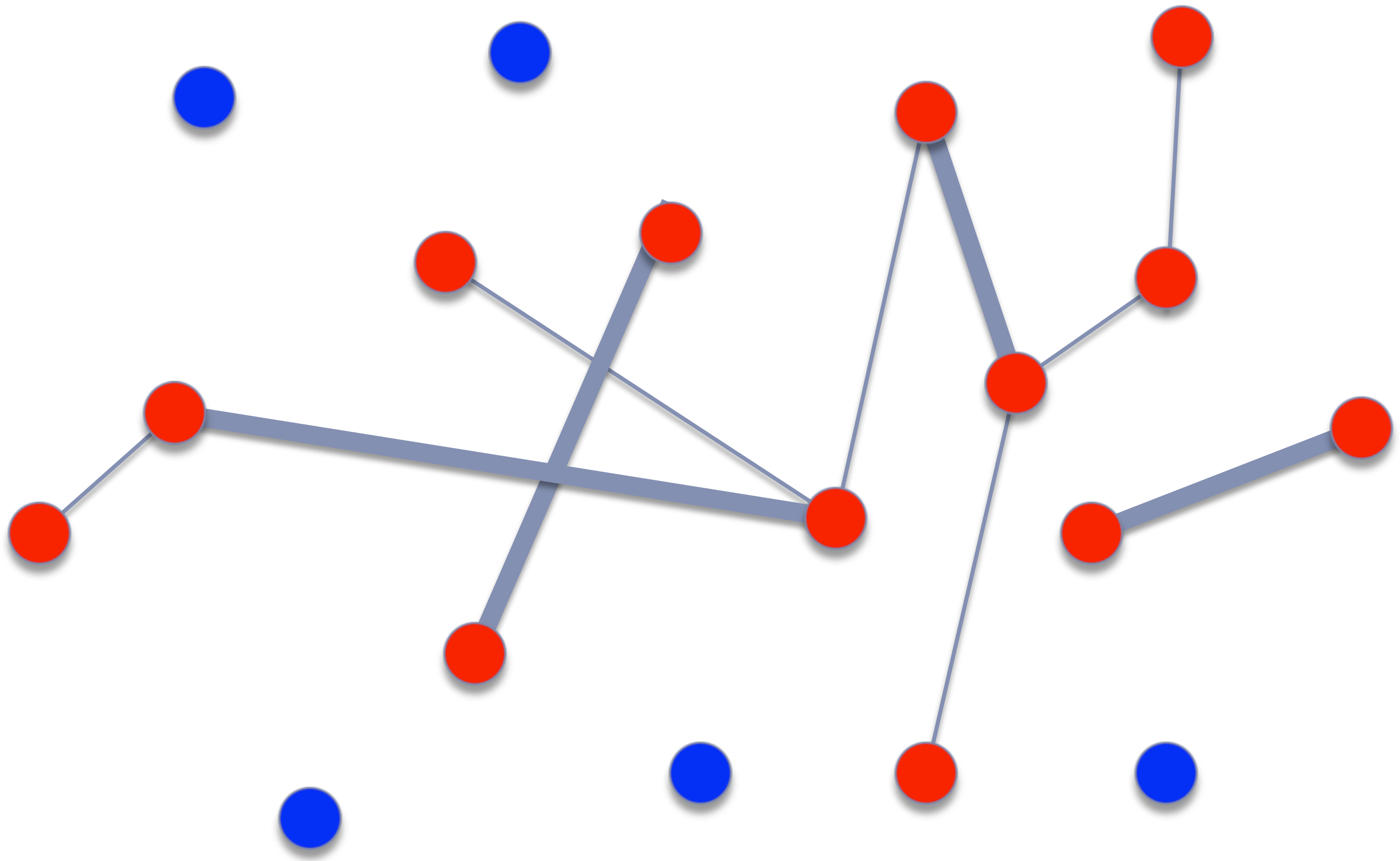
従来の推定法（性悪説？！）



デシメーション（性善説）



従来の推定法（性悪説？！）



デシメーション（性善説）

はっきりと零・非零を分けた推定：貪欲法の利用

A. Decelle and F. Ricci-Tersenghi: Phys. Rev. Lett. 112 (2014) 070603.

- 直観的には…
  - 性善説にもとづく推定
    - カンニング係数の絶対値が小さいのだから、疑いが小さい
    - 疑いが小さい被験者をいつまでも疑わない
    - だってカンニングは基本的にはないものだから！！

はっきりと零・非零を分けた推定：貪欲法の利用

A. Decelle and F. Ricci-Tersenghi: Phys. Rev. Lett. 112 (2014) 070603.

- 直観的には…
  - 性善説にもとづく推定
    - カンニング係数の絶対値が小さいのだから、疑いが小さい
    - 疑いが小さい被験者をいつまでも疑わない
    - だってカンニングは基本的にはないものだから！！

性善説で精度が向上

「ボルツマン機械学習は，もともと１つ１つの要素の相互作用を見つけるのが得意なプログラムです。それぞれの問題での正解，不正解のデータから，生徒

QUANTUM ANNEALING MACHINE LEARNING Sparse Modeling CREST

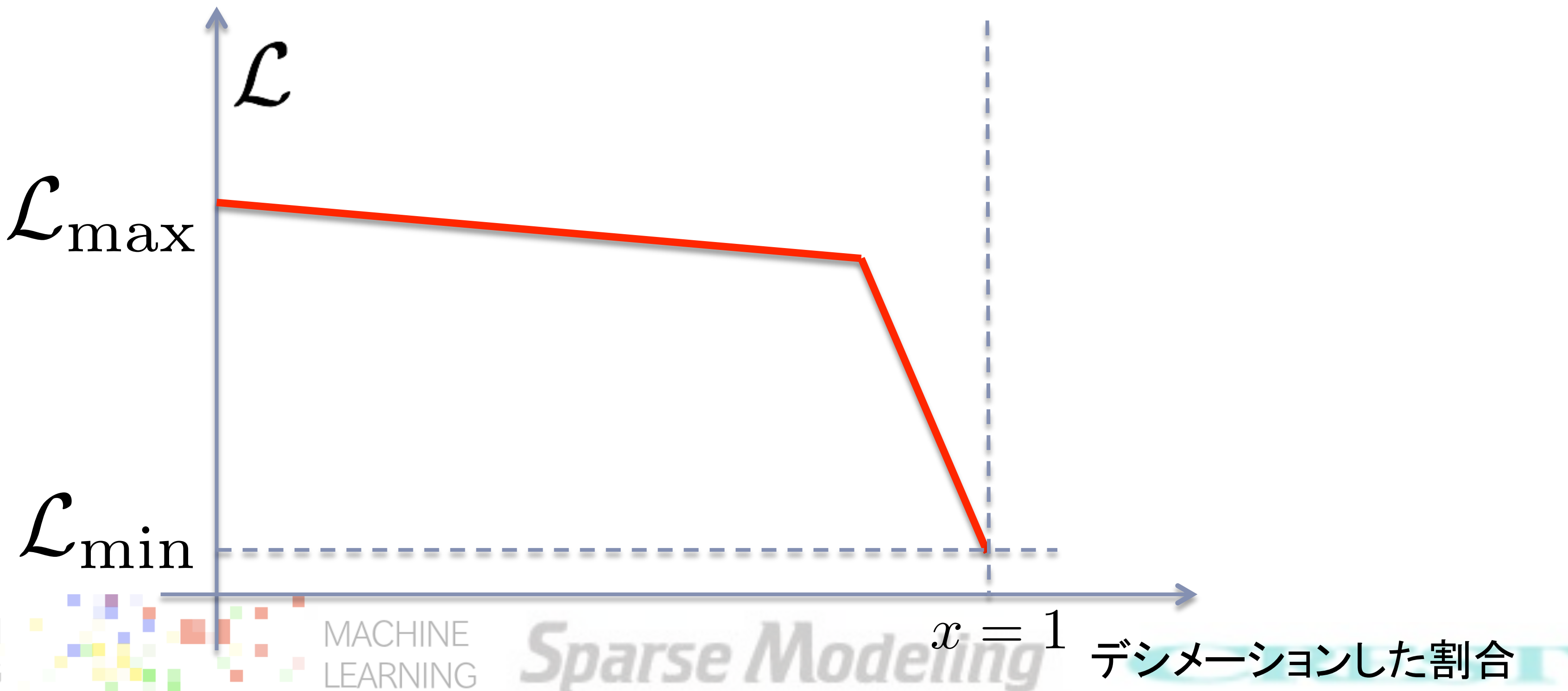
はっきりと零・非零を分けた推定：貪欲法の利用

A. Decelle and F. Ricci-Tersenghi: Phys. Rev. Lett. 112 (2014) 070603.

- 直観的には…
  - 性善説にもとづく推定
    - カンニング係数の絶対値が小さいのだから、疑いが小さい
    - 疑いが小さい被験者をいつまでも疑わない
    - だってカンニングは基本的にはないものだから！！

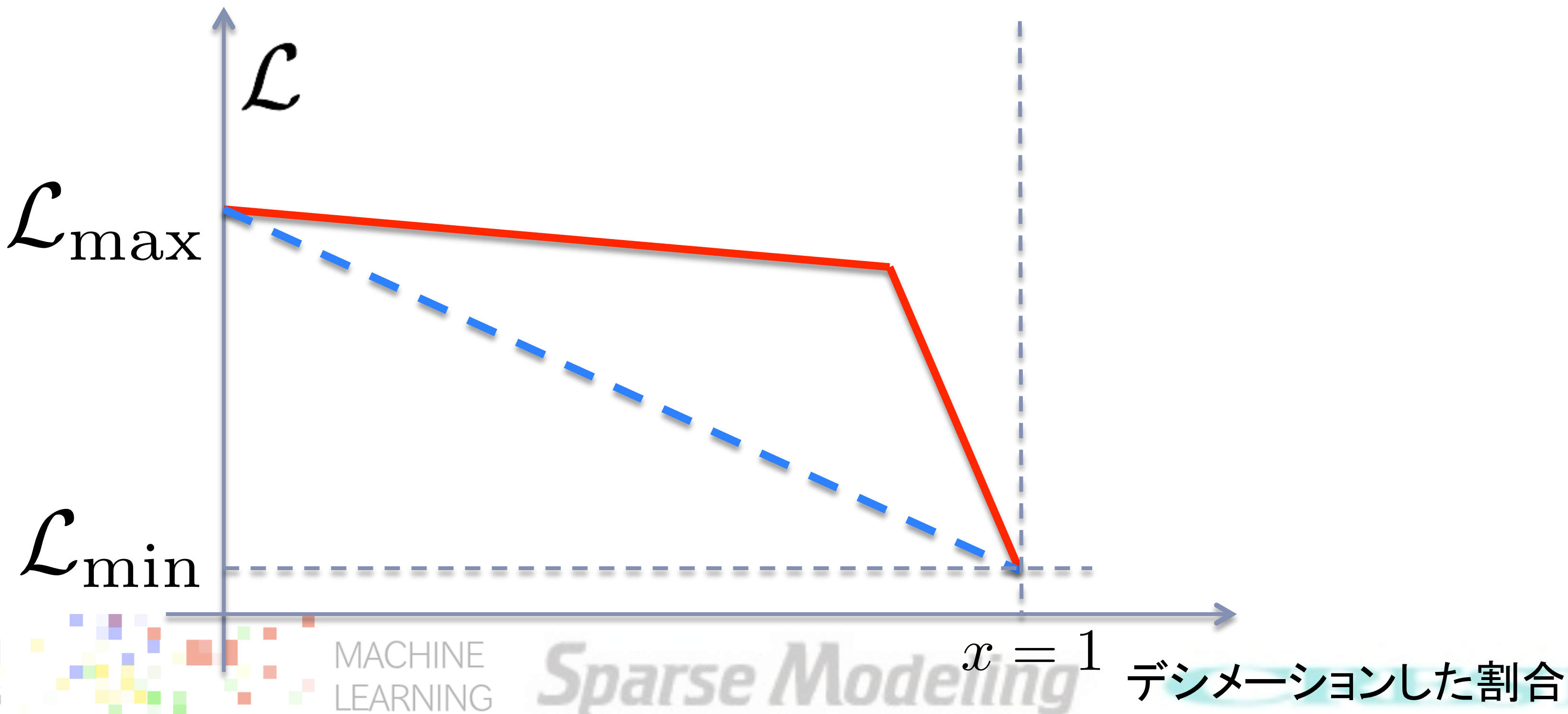
デシメーションアルゴリズム

- どこまでデシメーションすればよいのか？
  - スパース性の仮定のもとでは、ほとんどのカンニング係数は0である
  - 非常に小さいカンニング係数を0としても尤度関数はほぼ変わらない



デシメーションアルゴリズム

- どこまでデシメーションすればよいのか？
  - スパース性の仮定のもとでは、ほとんどのカンニング係数は0である
  - 非常に小さいカンニング係数を0としても尤度関数はほぼ変わらない

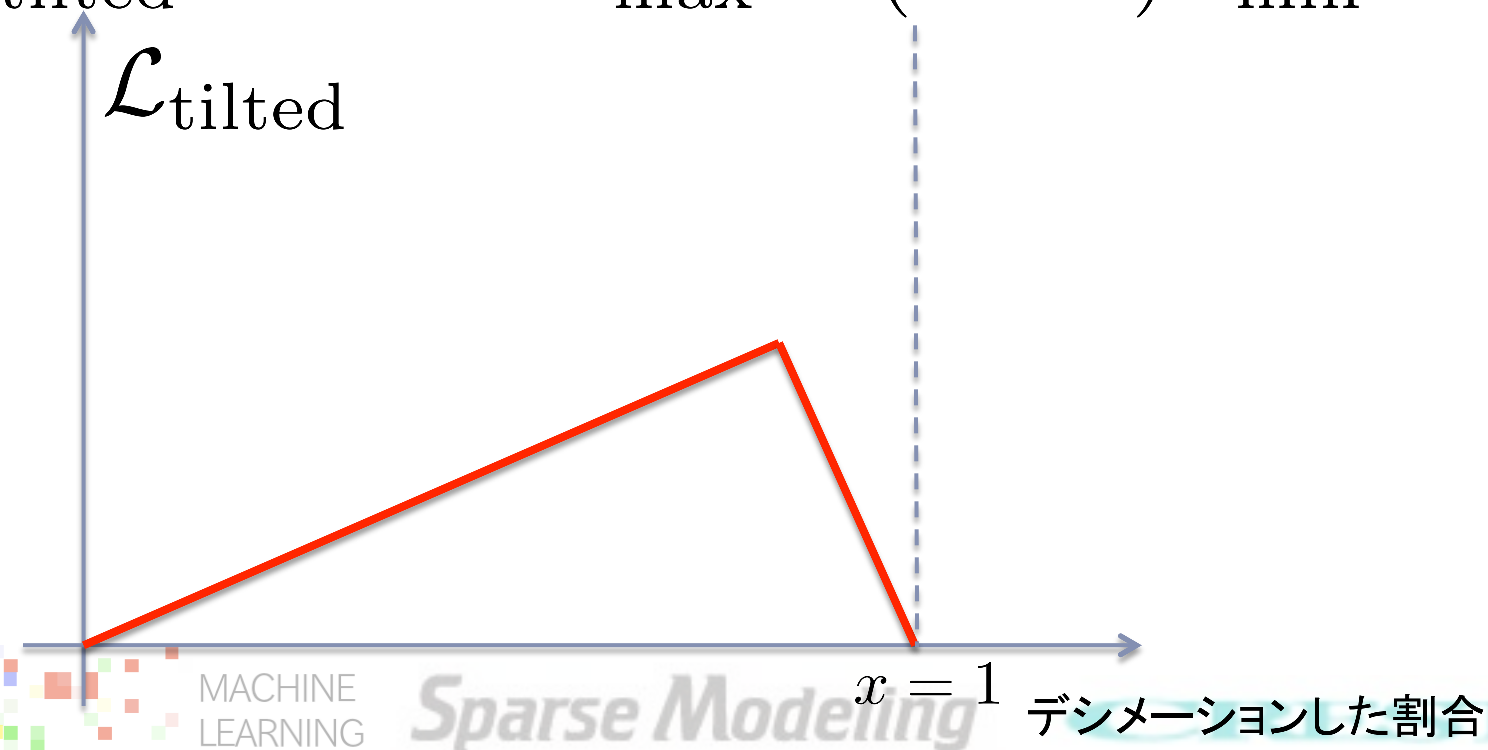


QUANTUM ANNEALING
MACHINE LEARNING
Sparse Modeling

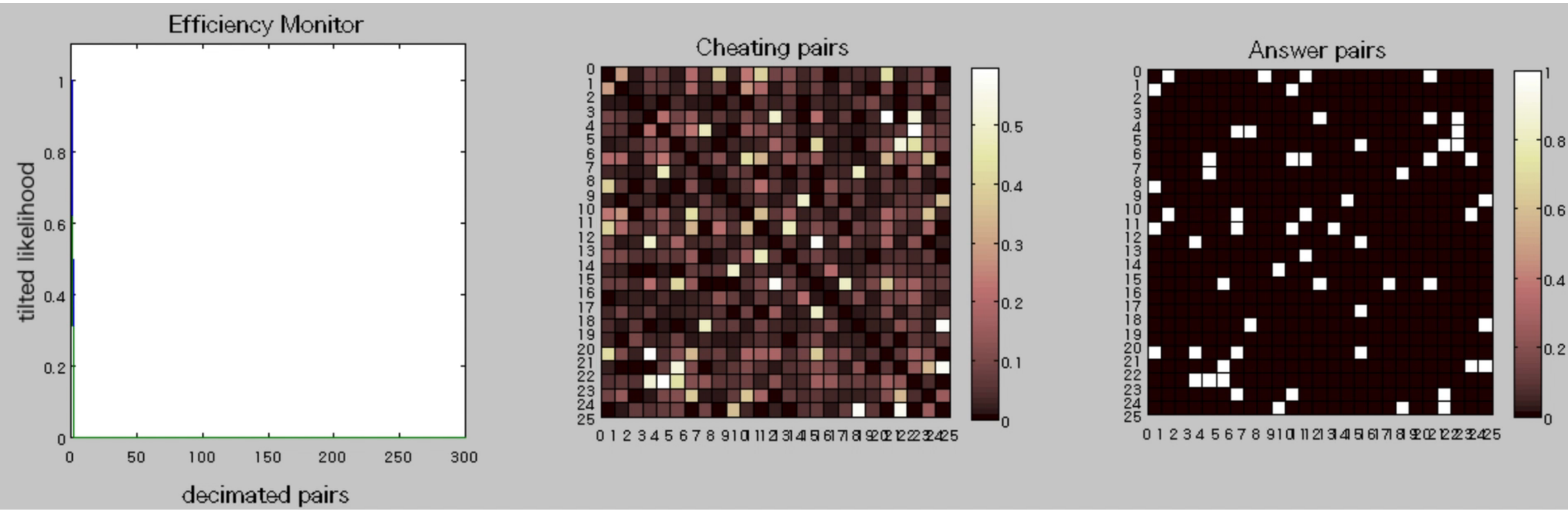
デシメーションアルゴリズム

- どこまでデシメーションすればよいのか？
  - スパース性の仮定のもとでは、ほとんどのカンニング係数は0である
  - 非常に小さいカンニング係数を0としても尤度関数はほぼ変わらない

$$\mathcal{L}_{\mathrm{tilted}} = \mathcal{L} - x\mathcal{L}_{\max} - (1-x)\mathcal{L}_{\min}$$



デシメーションアルゴリズムの様子

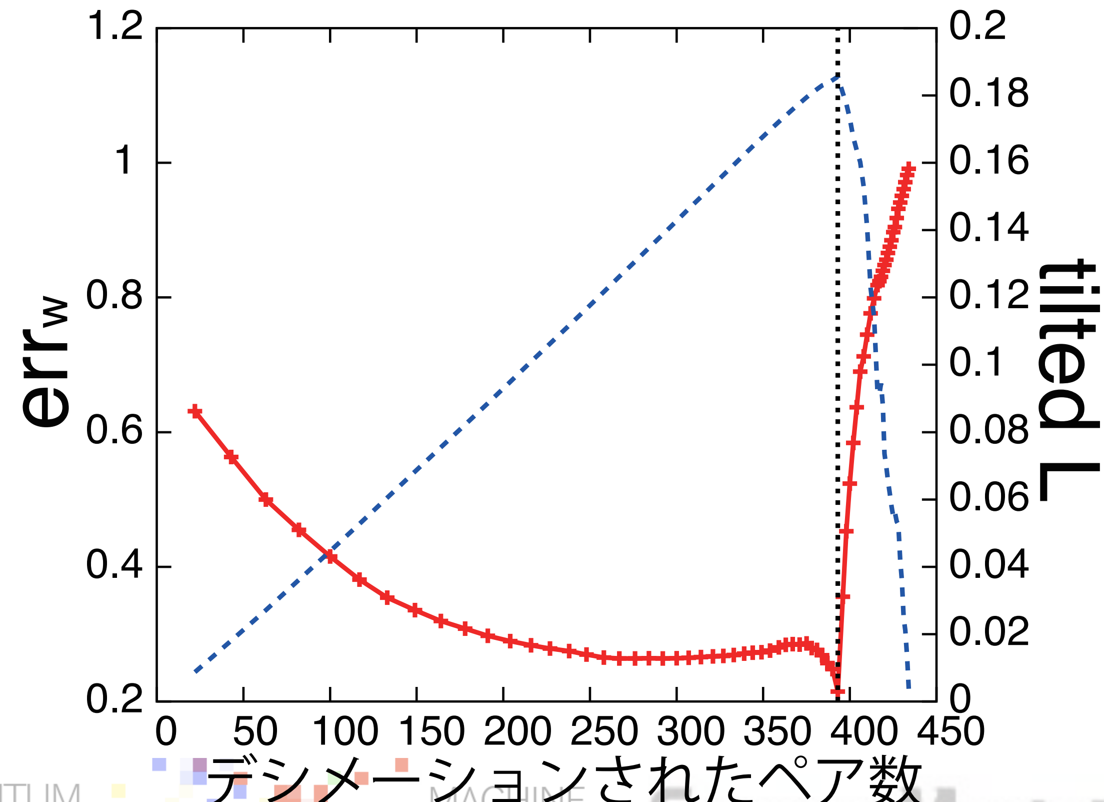


QUANTUM ANNEALING MACHINE LEARNING Sparse Modeling CREST

数値計算結果

S. Yamanaka, M. Ohzeki, A. Decelle: JPSJ 84, 024801 (2015)

- カンニング係数の推定結果
  - $\mathcal{L}_{\mathrm{tilted}}$ のピーク位置と推定誤差の最小値をとる位置が一致

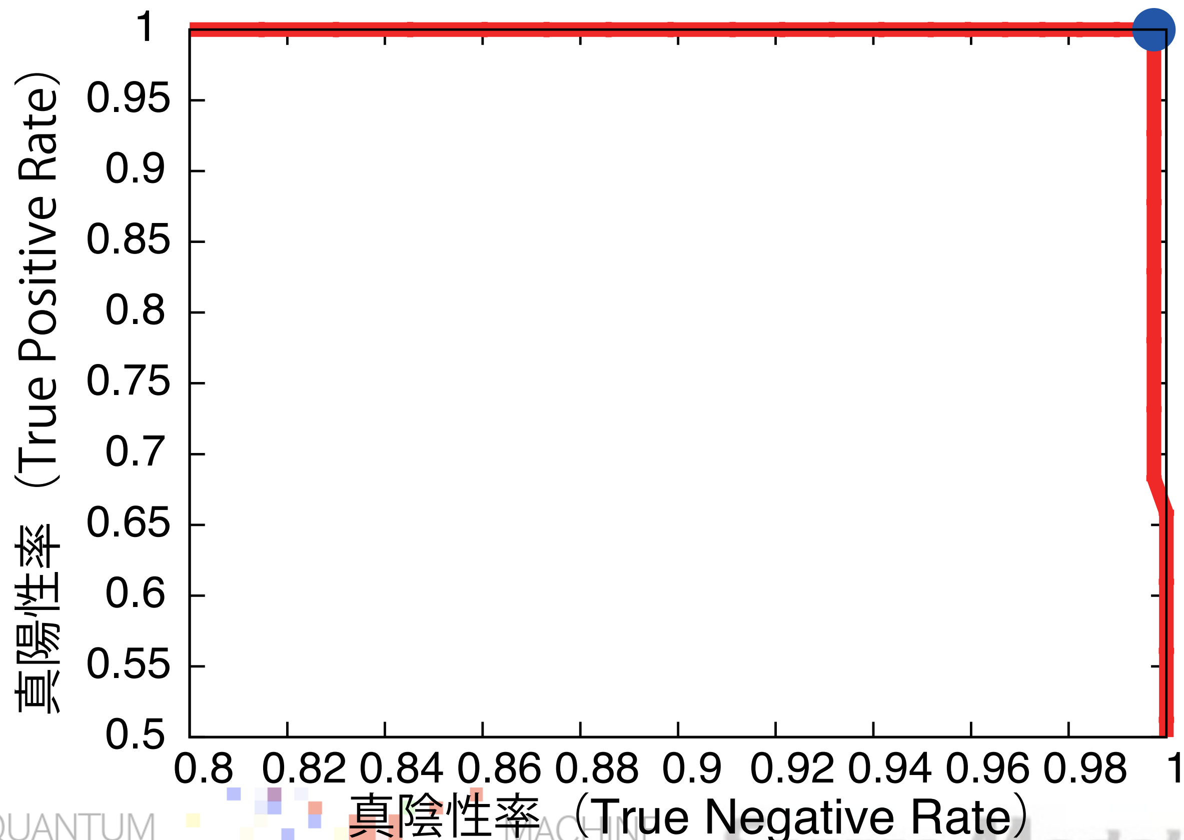


被験者数 : I = 30
ペア数 : I(I-1)/2 = 435
カンニングをしている割合
0.1
項目数 : J = 1000

数値計算結果

S. Yamanaka, M. Ohzeki, A. Decelle: JPSJ 84, 024801 (2015)

- カンニング係数の推定結果
  - $\mathcal{L}_{\mathrm{tilted}}$ のピーク位置でROC曲線の最良点にある

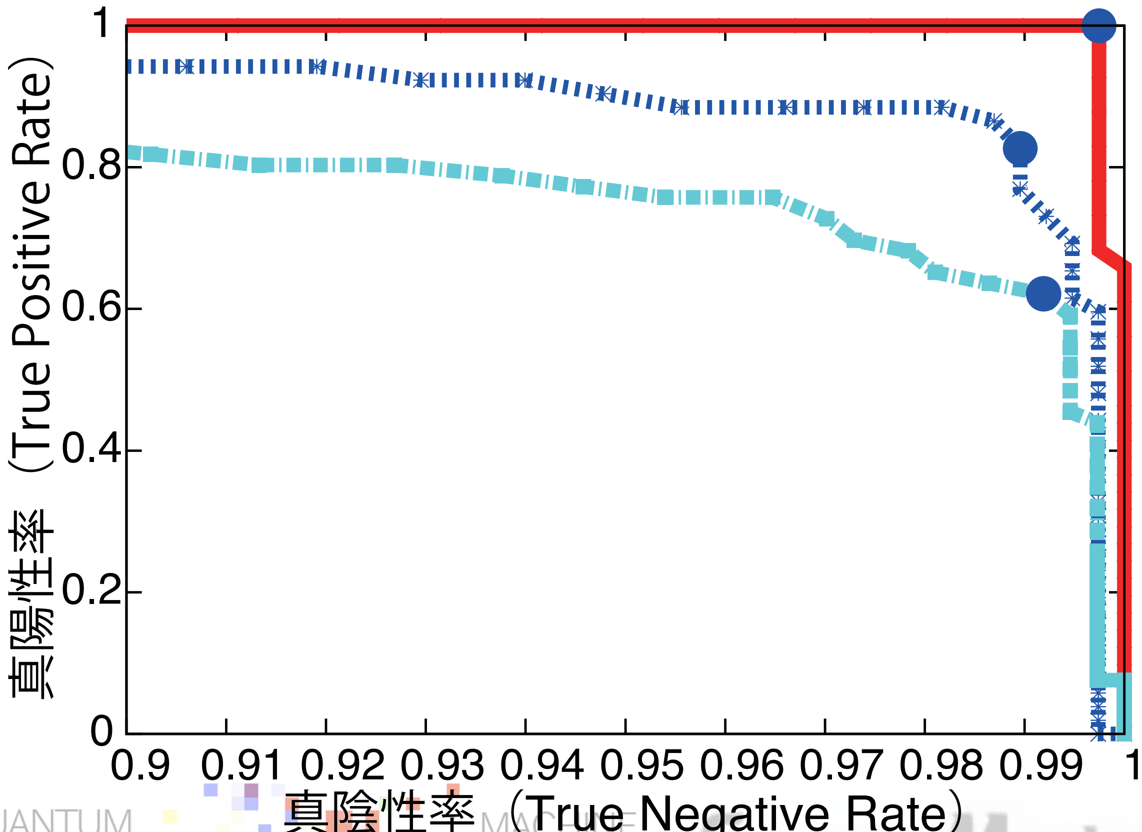


被験者数 : I = 30
ペア数 : I(I-1)/2 = 435
カンニングをしている割合 0.1
項目数 : J = 1000
真陽性率 :
カンニングをしていると推定されたペア/実際にカンニングをしているペア
真陰性率 :
カンニングをしていないと推定されたペア/実際にカンニングをしていないペア

スパース性との関係性

S. Yamanaka, M. Ohzeki, A. Decelle: JPSJ 84, 024801 (2015)

- カンニング係数の推定結果
  - ROC曲線が左下に後退、真陽性率が減少



被験者数：I = 30
ペア数：I(I-1)/2 = 435
カンニングをしている割合
0.1->0.125->0.15
項目数：J = 1000
真陽性率：
カンニングをしていると推定されたペア
/実際にカンニングをしているペア
真陰性率：
カンニングをしていないと推定されたペア
/実際にカンニングをしていないペア

まとめ

- ボルツマン機械学習=データ駆動型の統計力学
  - 期待値計算を如何に正確に実行するか（マルコフ連鎖モンテカルロ法）
  - 期待値計算を如何に高速に実行するか（平均場近似等）
  - 大規模なデータ利用を如何にさぼるか（確率勾配法）
  - 複雑度を増しても計算は容易なものが登場
    - 深層学習における隠れ変数ありボルツマンマシン
    - リザバーコンピューティング
- 本質部分の抽出＝ビッグデータに対する対処策
  - 複雑になればなるほど、本質が曖昧に
  - スパースモデリングの導入によりクリアな理解へ

QUANTUM ANNEALING MACHINE LEARNING Sparse Modeling CREST