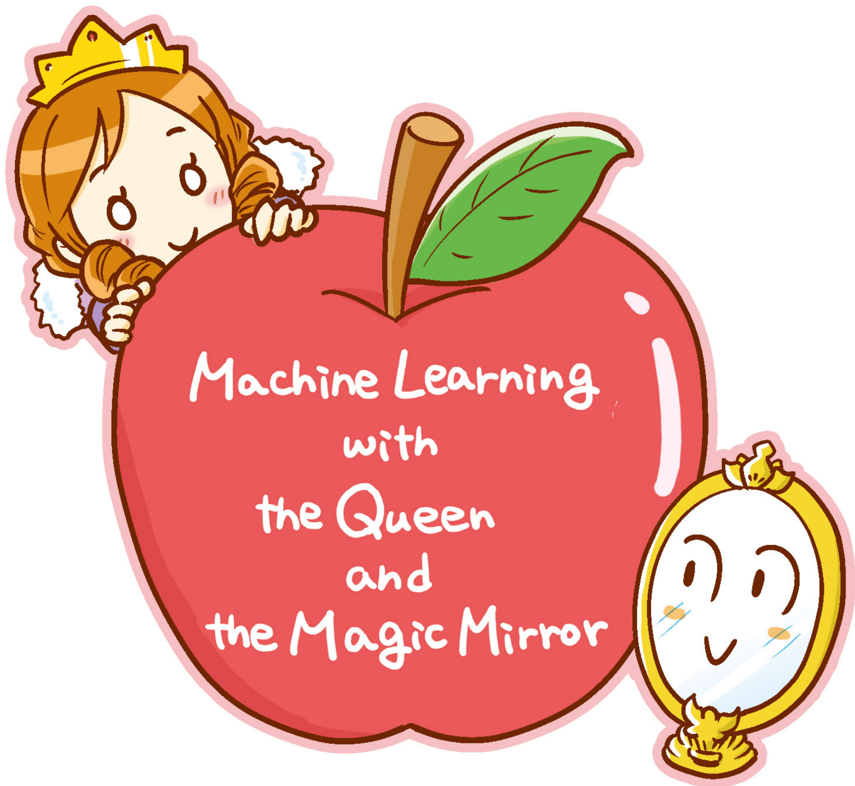


マルチメディア情報ハイディング・エンリッチメント研究会（EMM）

知的情報処理の最前線 -スパースモデリング vs 深層学習-

東北大学大学院情報科学研究科応用情報科学専攻

大関　真之

TOHOKU UNIVERSITY

CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

TOHOKU UNIVERSITY

世の中で何が起きているのか？

世の中で何が起きているのか？

人工知能？

世の中で何が起きているのか？

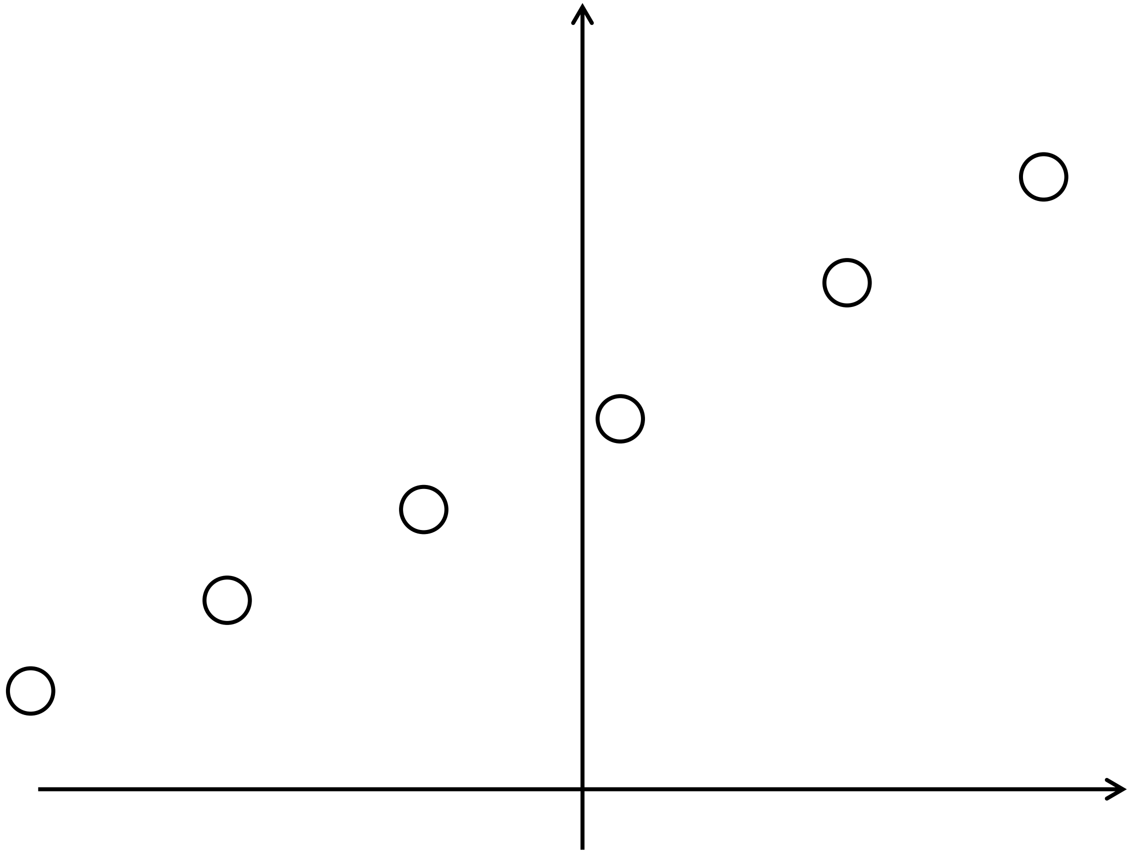
逆問題の解決

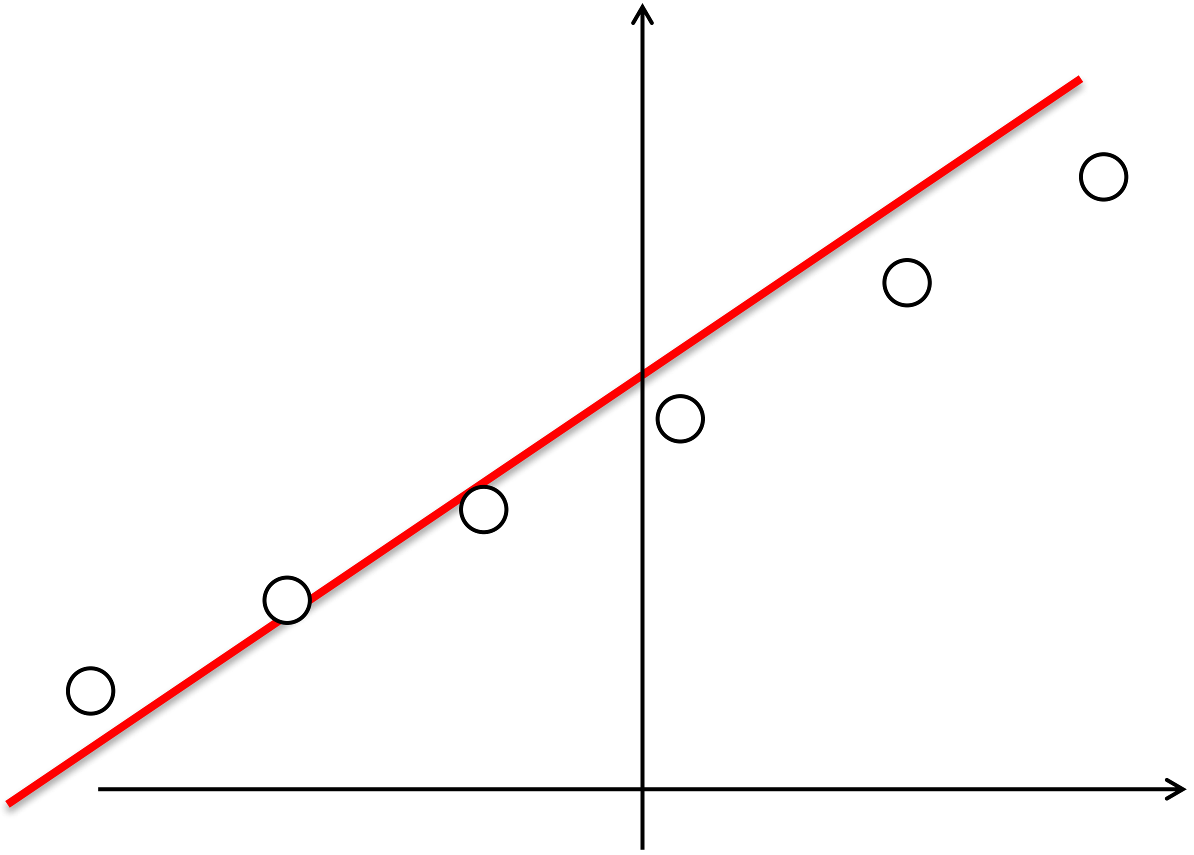
$$y = f(\mathbf{x})$$

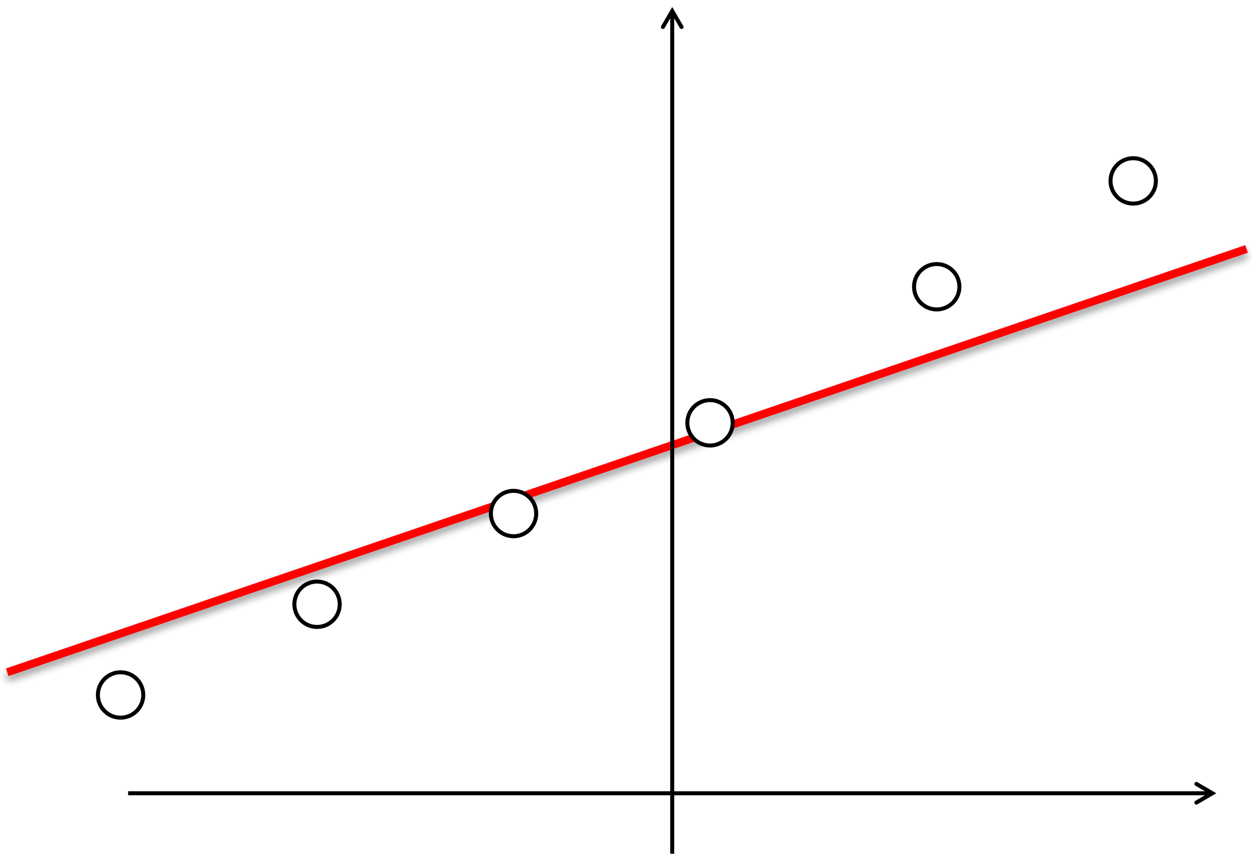
世の中で何が起きているのか？

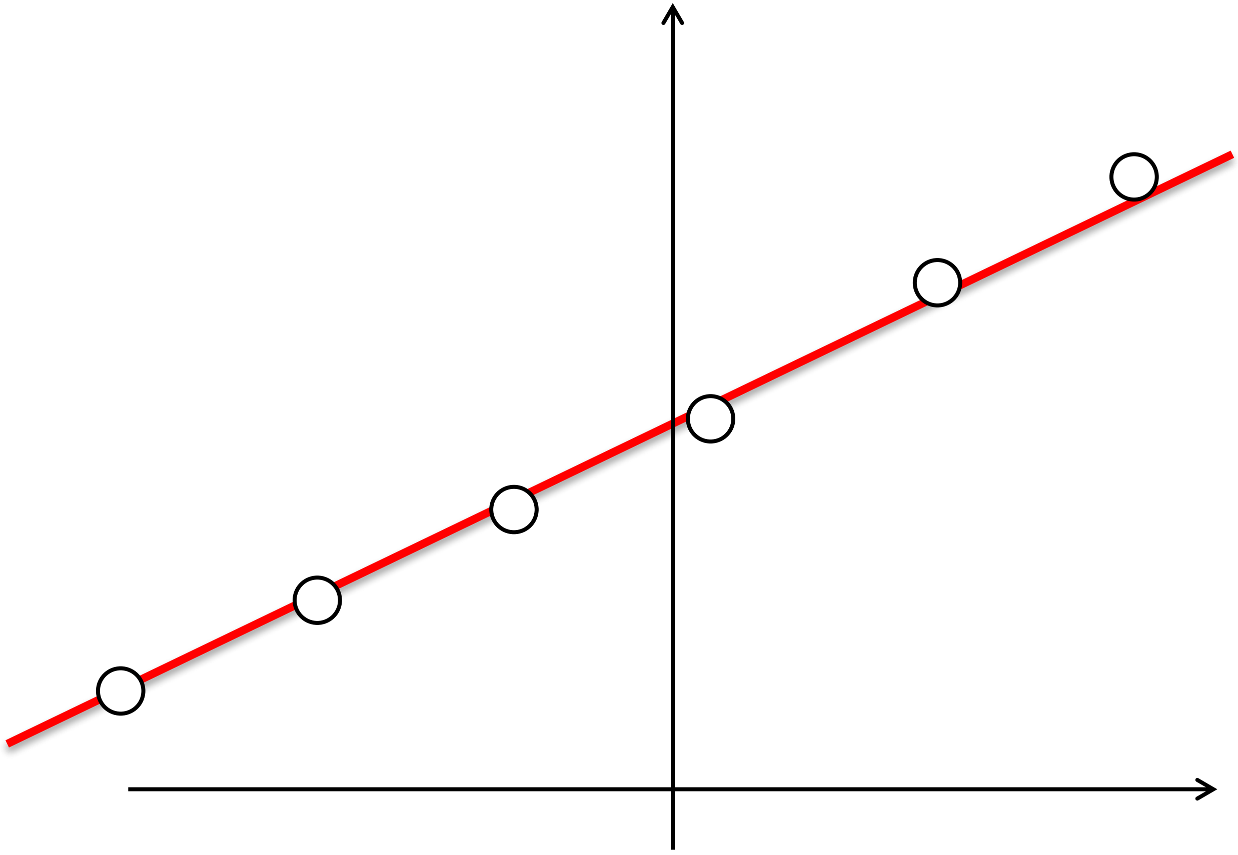
逆問題の解決

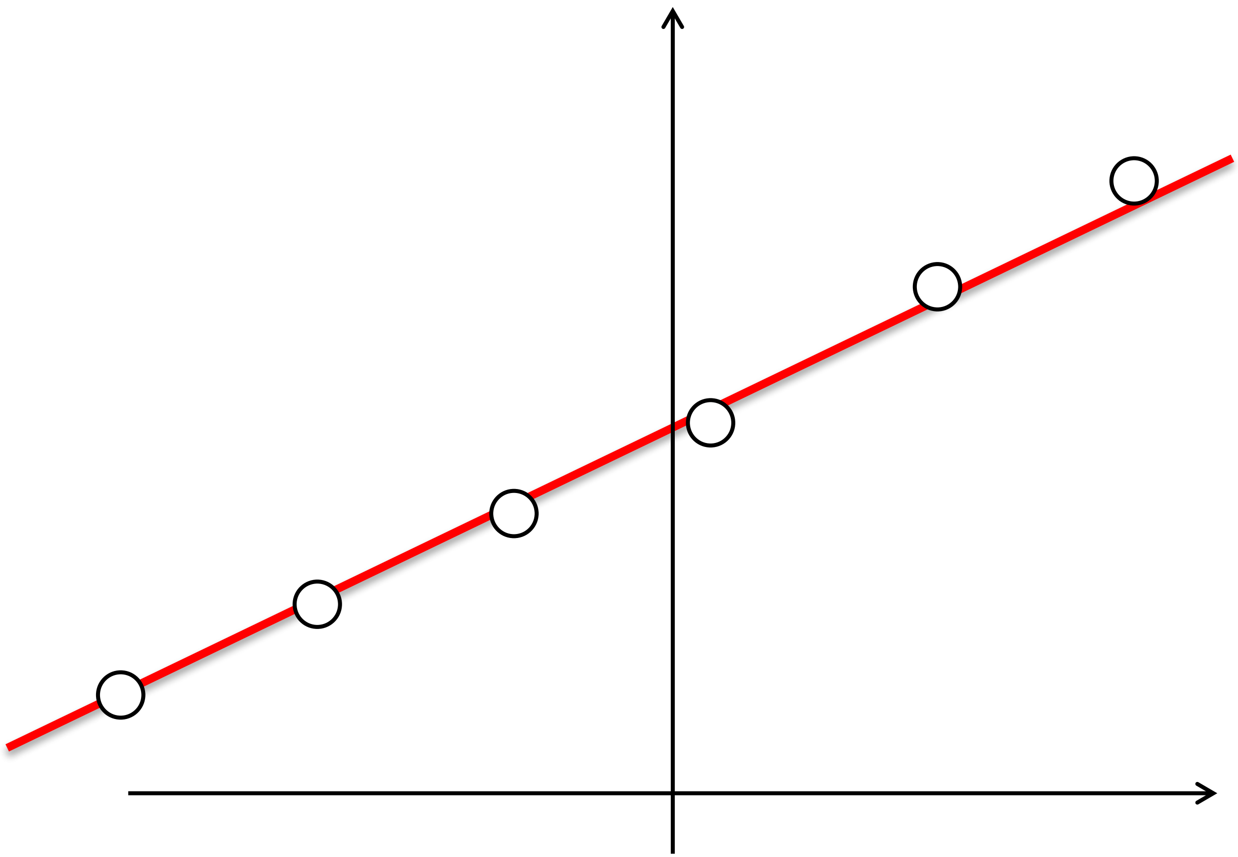
$$y = f(\mathbf{x})$$





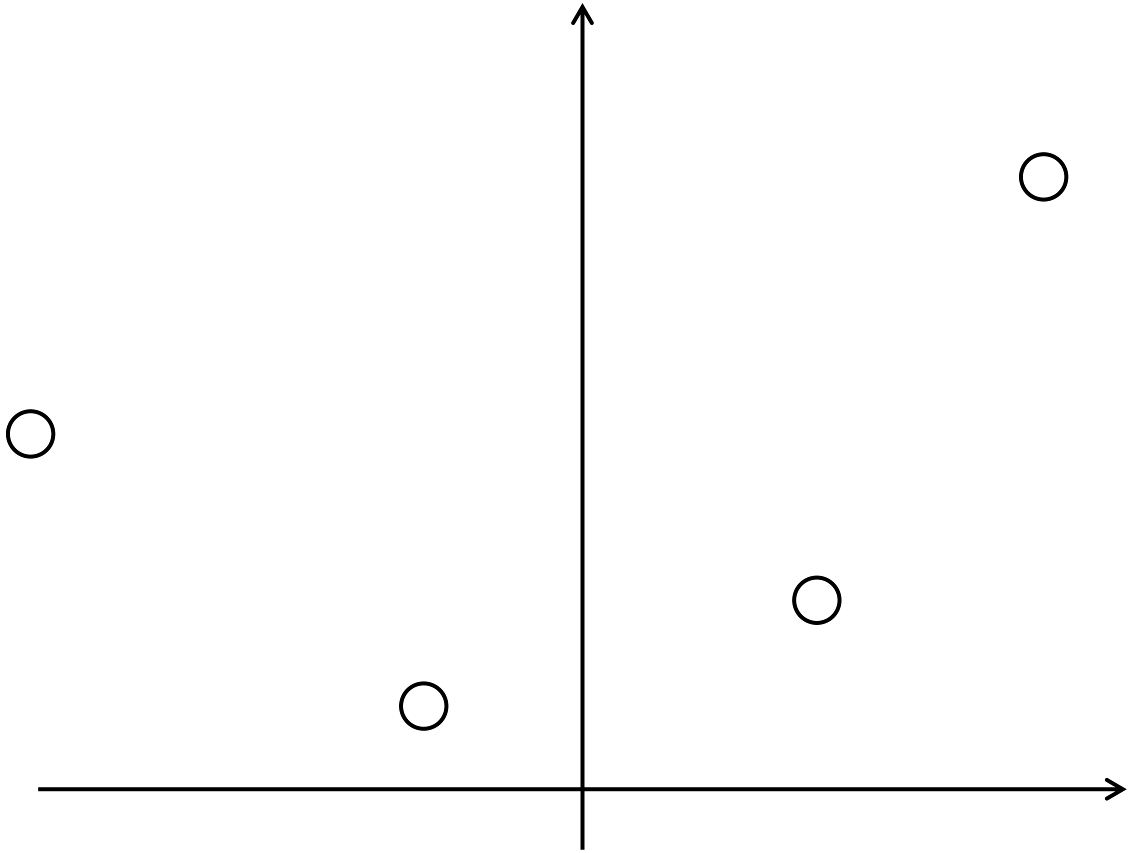


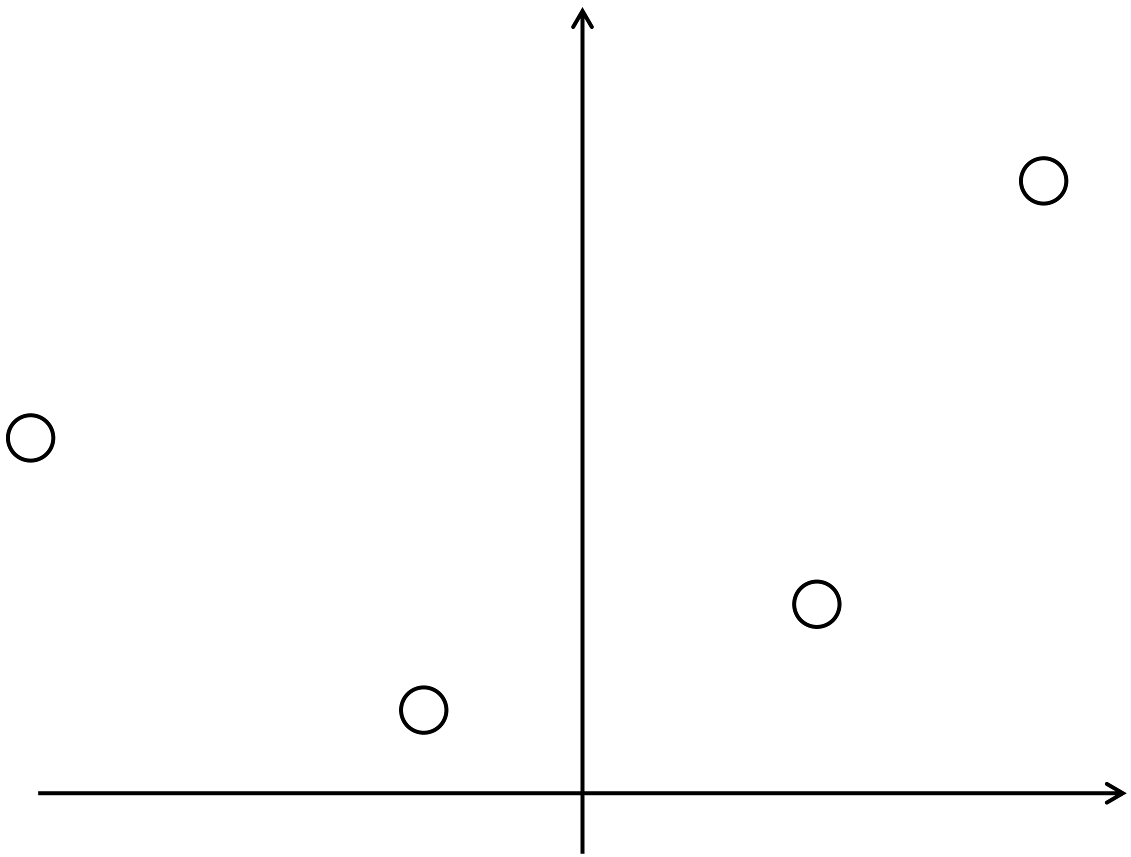




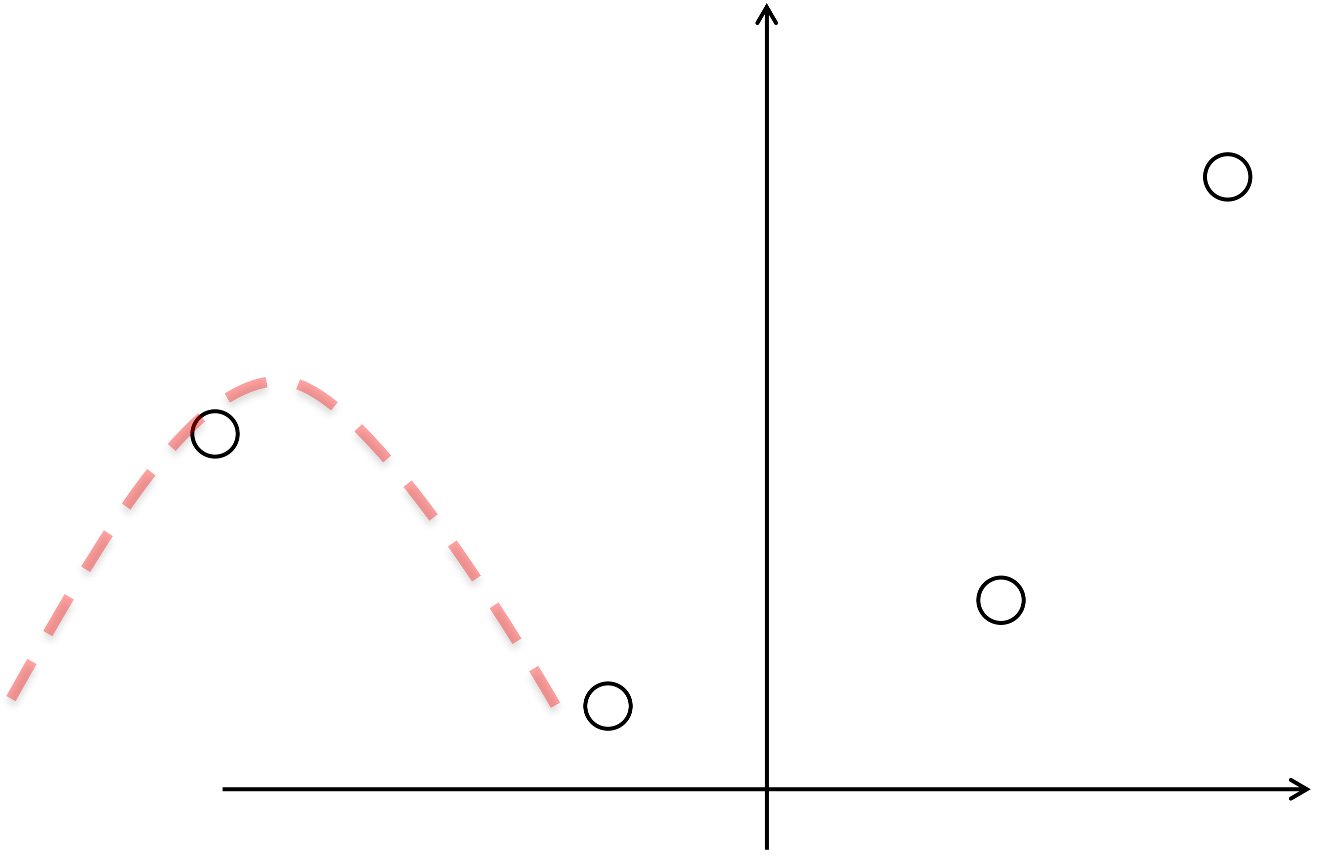
$$\min_{\mathbf{a}} \left\{ \sum_{d=1}^{D} \left(y^{(d)} - a x^{(d)} - b \right)^2 \right\}$$

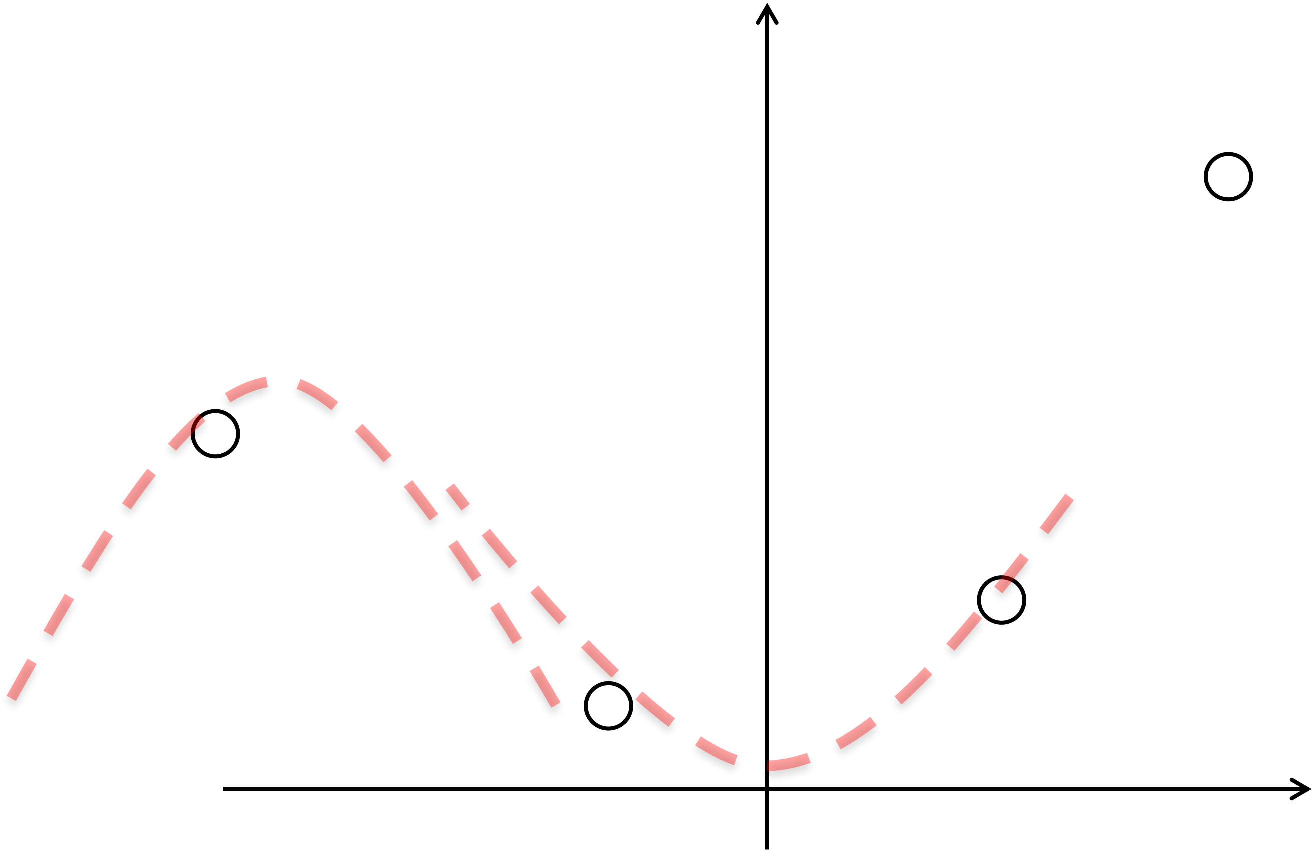
複雑になったら？？

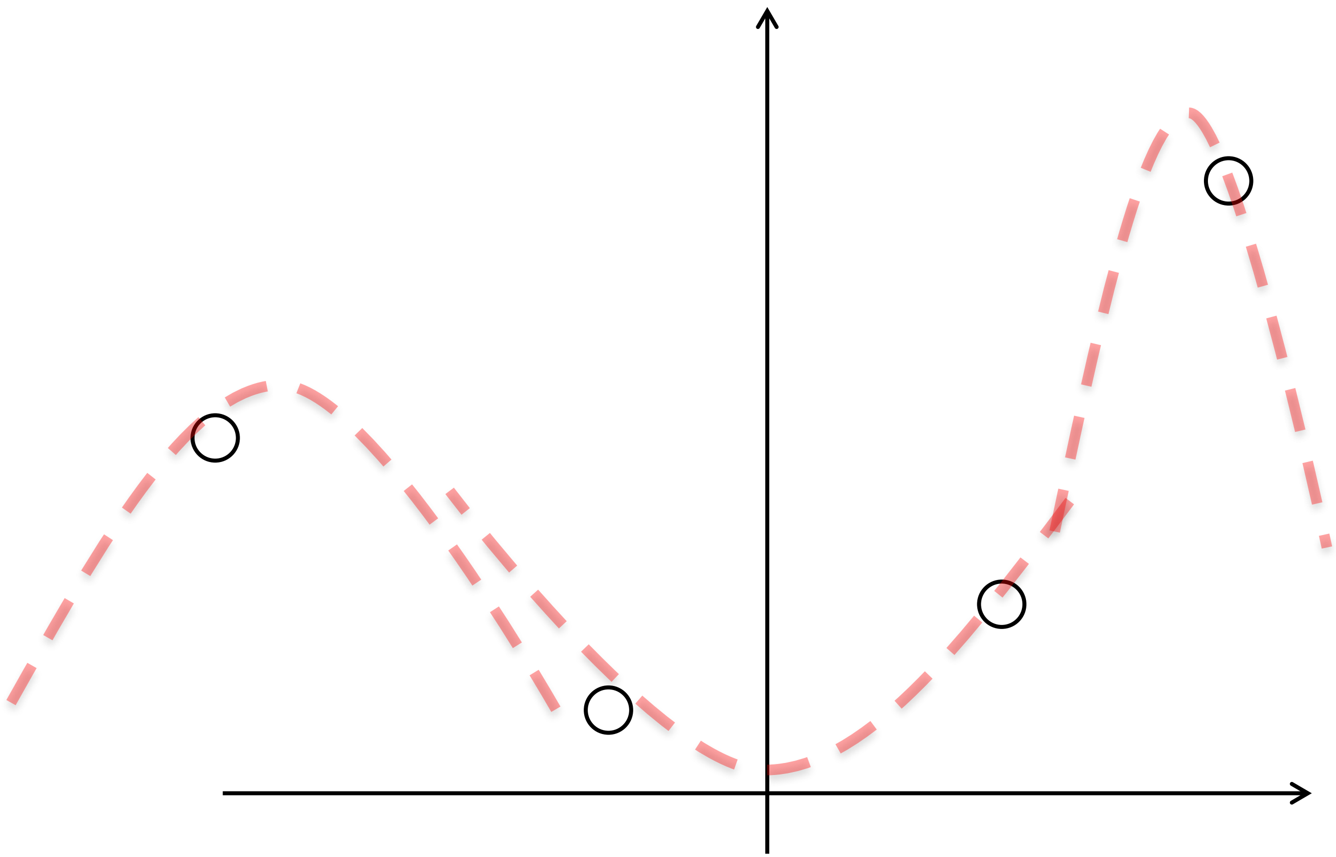


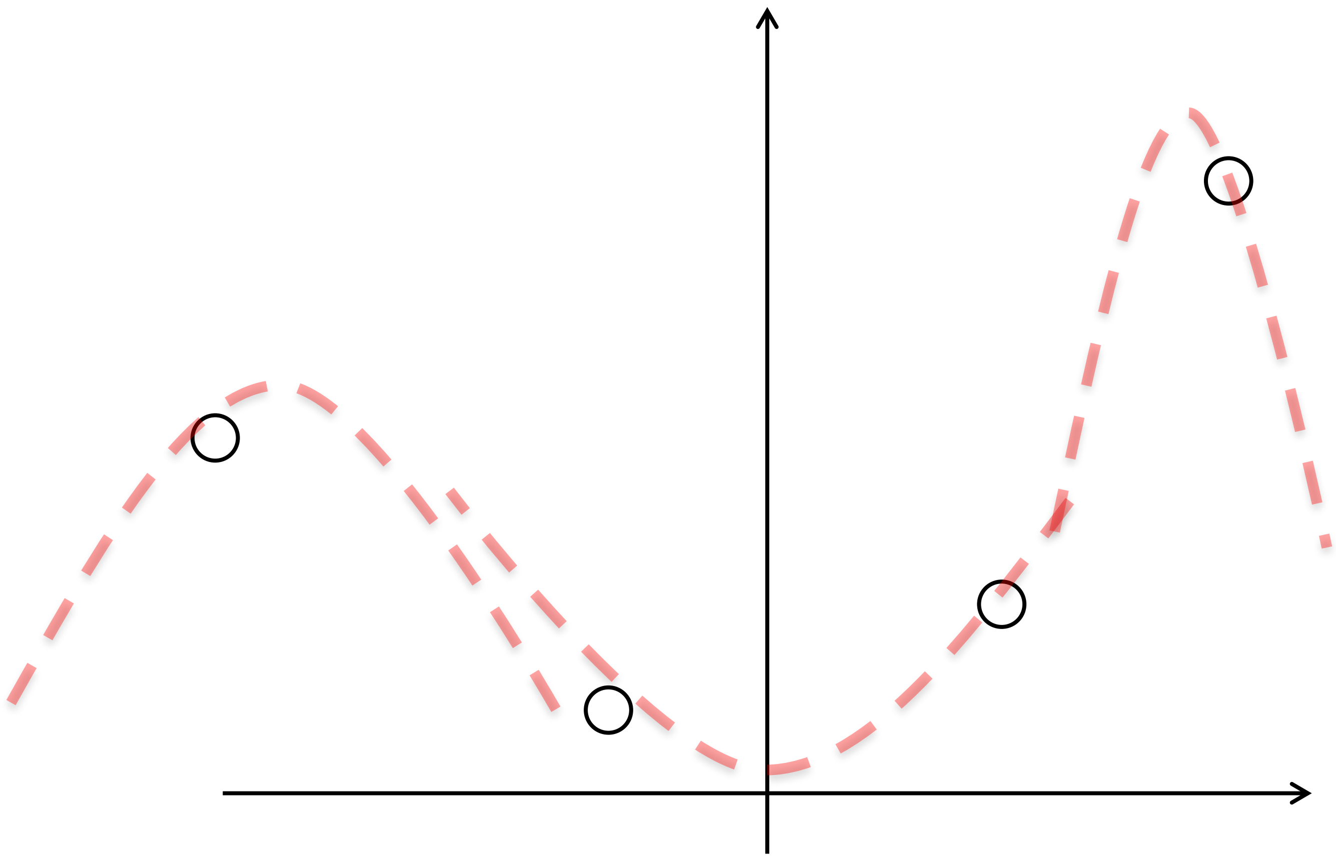


直線？曲線？何次関数？？

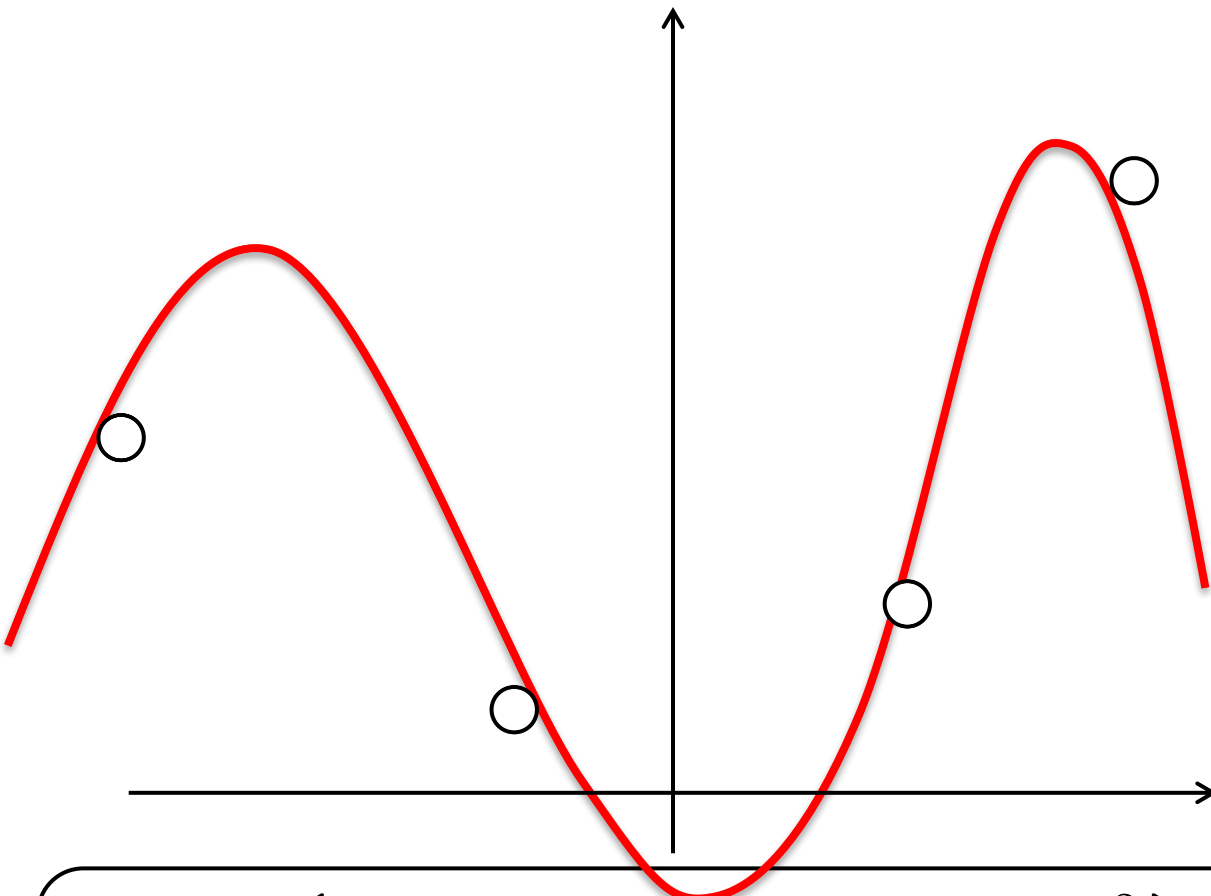








$$\min_{\mathbf{a}} \left\{ \sum_{d=1}^{D} \left(y^{(d)} - \sum_{k=0}^{N} a_k f_k(x^{(d)}) \right)^2 \right\}$$



$$\min_{\mathbf{a}} \left\{ \sum_{d=1}^{D} \left(y^{(d)} - \sum_{k=0}^{N} a_k f_k(x^{(d)}) \right)^2 \right\}$$

機械学習の躍進

- **データに潜む関係性を暴く**

$$y = f(\mathbf{x})$$

- x: 入力、y:出力、これらをつなぐ**関数fを知りたい**

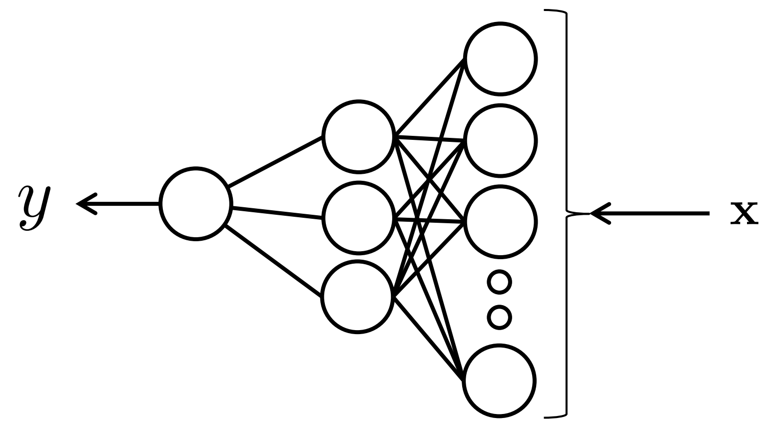
機械学習の躍進

- データに潜む関係性を暴く

$$y = f(\mathbf{x})$$

  - x: 入力、y:出力、これらをつなぐ関数fを知りたい
  - 単純だけど繰り返しによる複雑さの極致=深層学習

How to 深層学習？



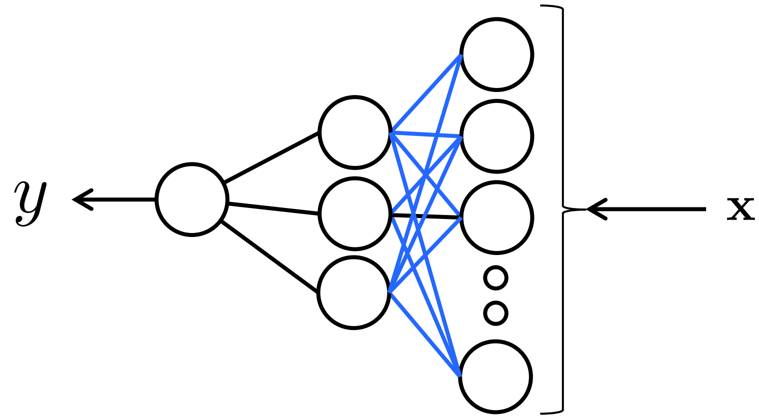
- 関数の**モデリングの放棄** = とことん楽をする
  - 線形変換

$$\boxed{\mathbf{v} = W\mathbf{x}}$$

  - 非線形変換

$$\boxed{\mathbf{x}' = \phi(\mathbf{v})}$$

How to 深層学習？



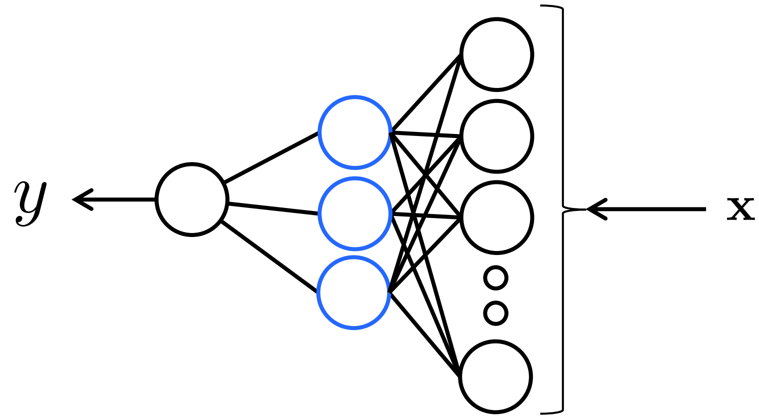
- 関数の**モデリングの放棄** = とことん楽をする
  - 線形変換

$$\boxed{\mathbf{v} = W\mathbf{x}}$$

  - 非線形変換

$$\boxed{\mathbf{x}' = \phi(\mathbf{v})}$$

How to 深層学習？



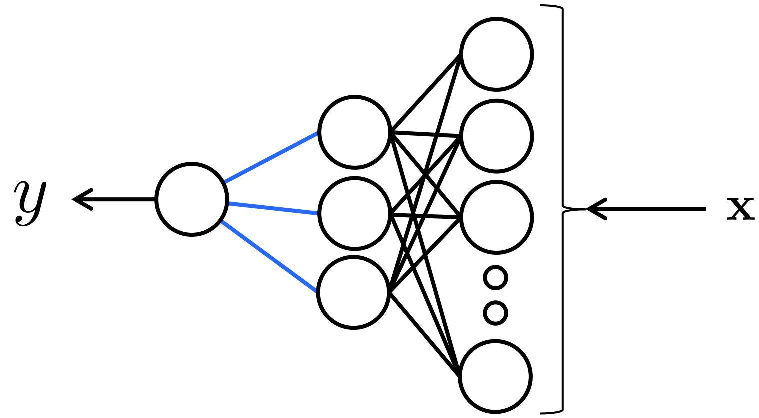
- 関数の**モデリングの放棄** = とことん楽をする
  - 線形変換

$$\boxed{\mathbf{v} = W\mathbf{x}}$$

  - 非線形変換

$$\boxed{\mathbf{x}' = \phi(\mathbf{v})}$$

How to 深層学習？



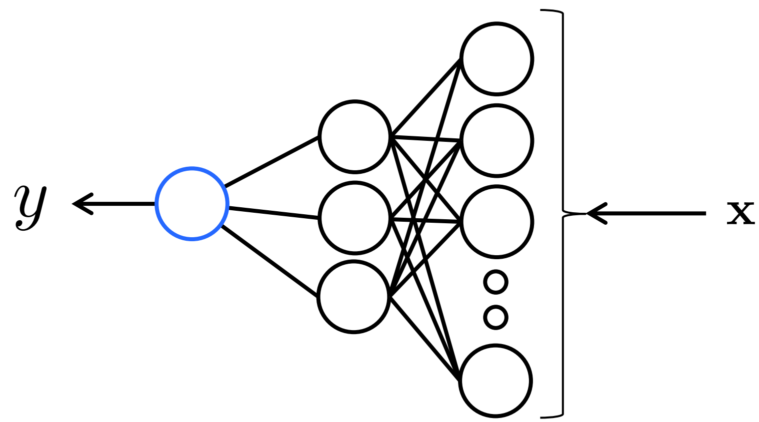
- 関数の**モデリングの放棄** = とことん楽をする
  - 線形変換

$$\boxed{\mathbf{v} = W\mathbf{x}}$$

  - 非線形変換

$$\boxed{\mathbf{x}' = \phi(\mathbf{v})}$$

How to 深層学習？



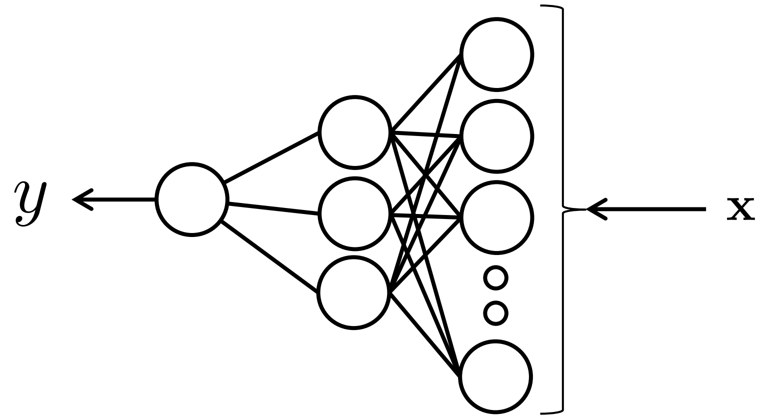
- 関数の**モデリングの放棄** = とことん楽をする
  - 線形変換

$$\boxed{\mathbf{v} = W\mathbf{x}}$$

  - 非線形変換

$$\boxed{\mathbf{x}' = \phi(\mathbf{v})}$$

How to 深層学習？

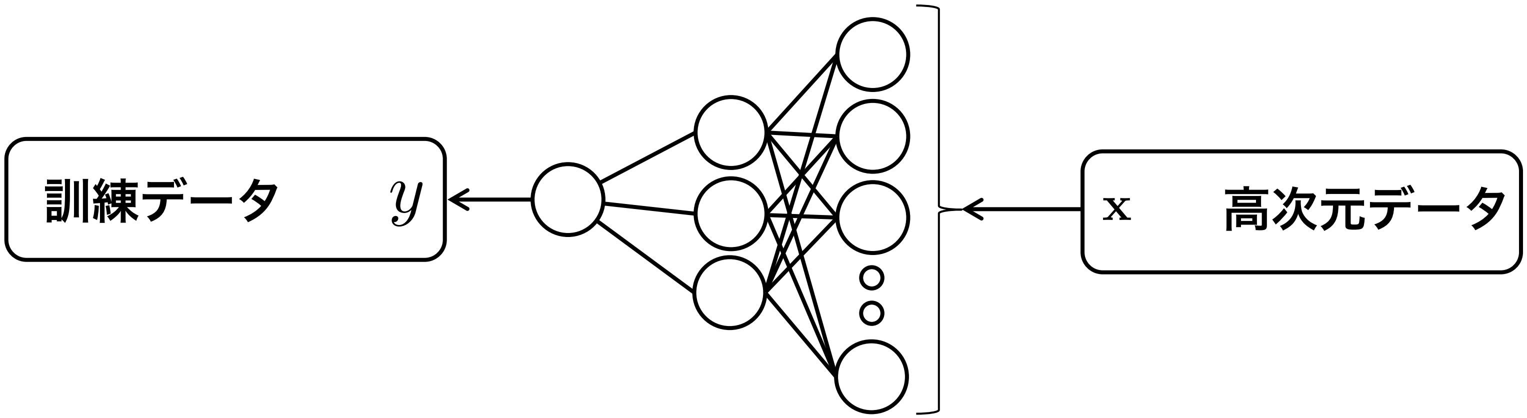


- 関数のモデリングの放棄 = とことん楽をする
  - 繰り返し構造の関数を用意 $f_k[\mathbf{x}] = \phi_k\left(W_k\mathbf{x}\right)$

$$y = f(\mathbf{x}) = f_n[f_{n-1}[\cdots f_1[\mathbf{x}]]]$$

  - 多層ニューラルネットワーク

How to 深層学習？



- **正解データに合わせる最適化を実行する**
  - 損失関数の例:連続値を出力とする場合

$$E_f(\mathbf{y}^{(1:D)},\mathbf{x}^{(1:D)}) = \frac{1}{2}\sum_{d=1}^{D}\left\|\mathbf{y}^{(d)} - f\left(\mathbf{x}^{(d)}\right)\right\|^2$$

CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

世の中で何が起きているのか？

逆問題の解決

$$y = f(\mathbf{x})$$

世の中で何が起きているのか？

逆問題の解決

$$y = f(\mathbf{x})$$

関数は分かっているとして…

- データの本質を暴く

$$y = f(\mathbf{x})$$

- x: 入力、y:出力、出力が出てきた**原因・要因・本質xを知りたい**

関数は分かっているとして…

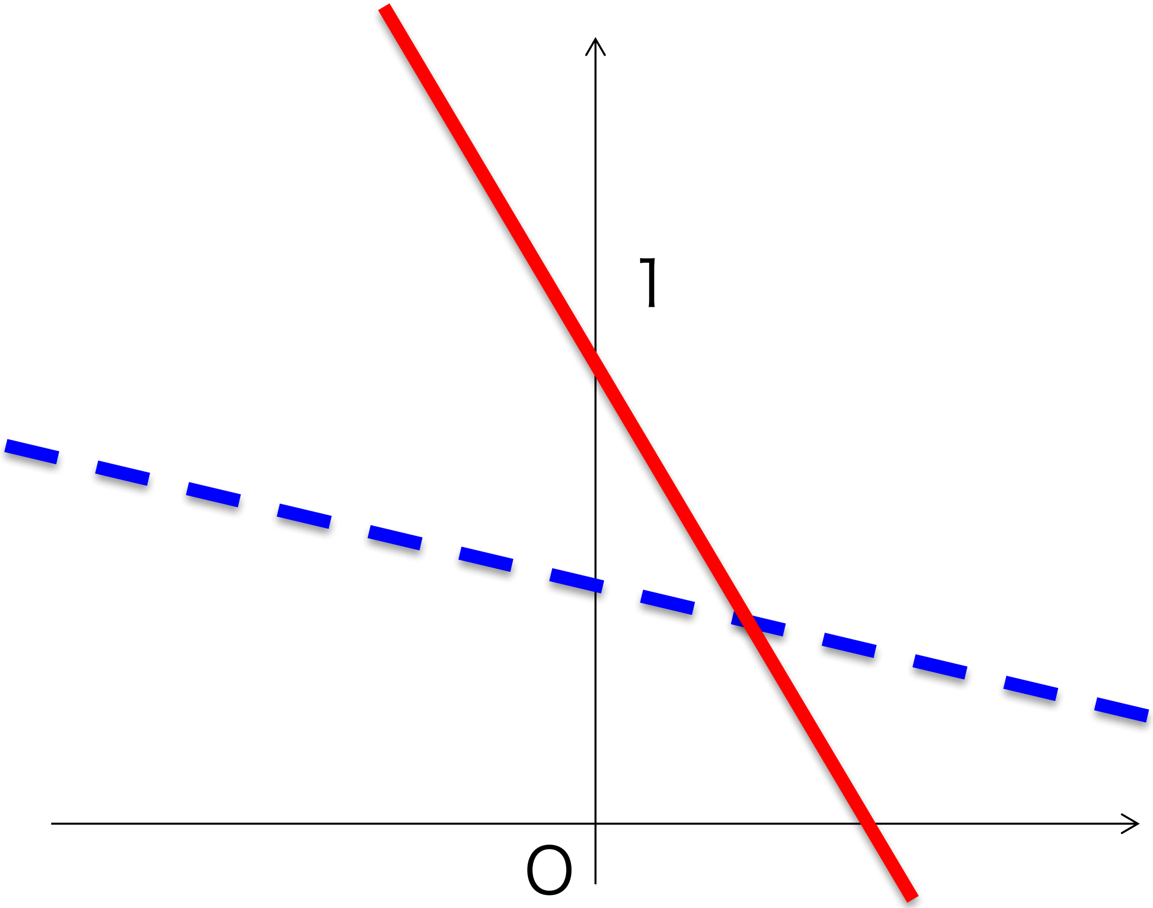
- データの本質を暴く

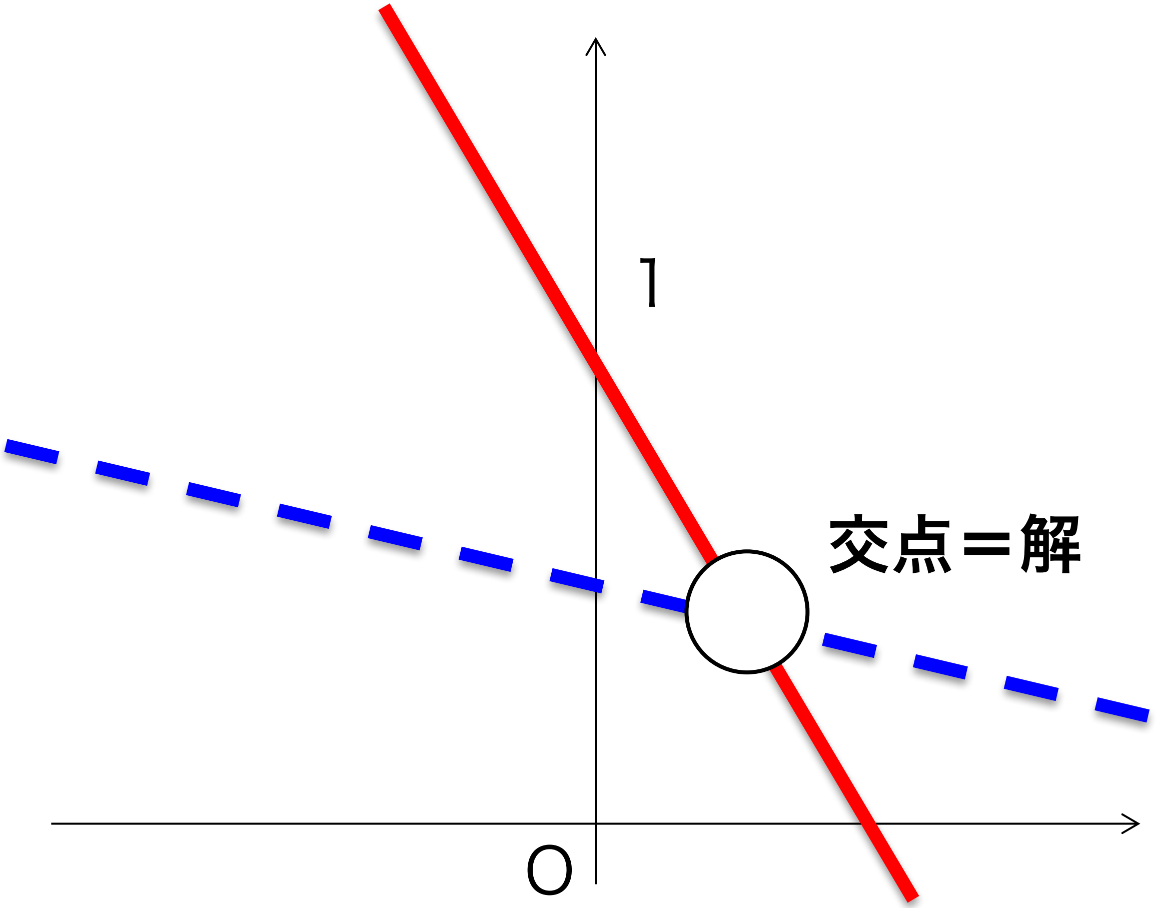
$$y = f(\mathbf{x})$$

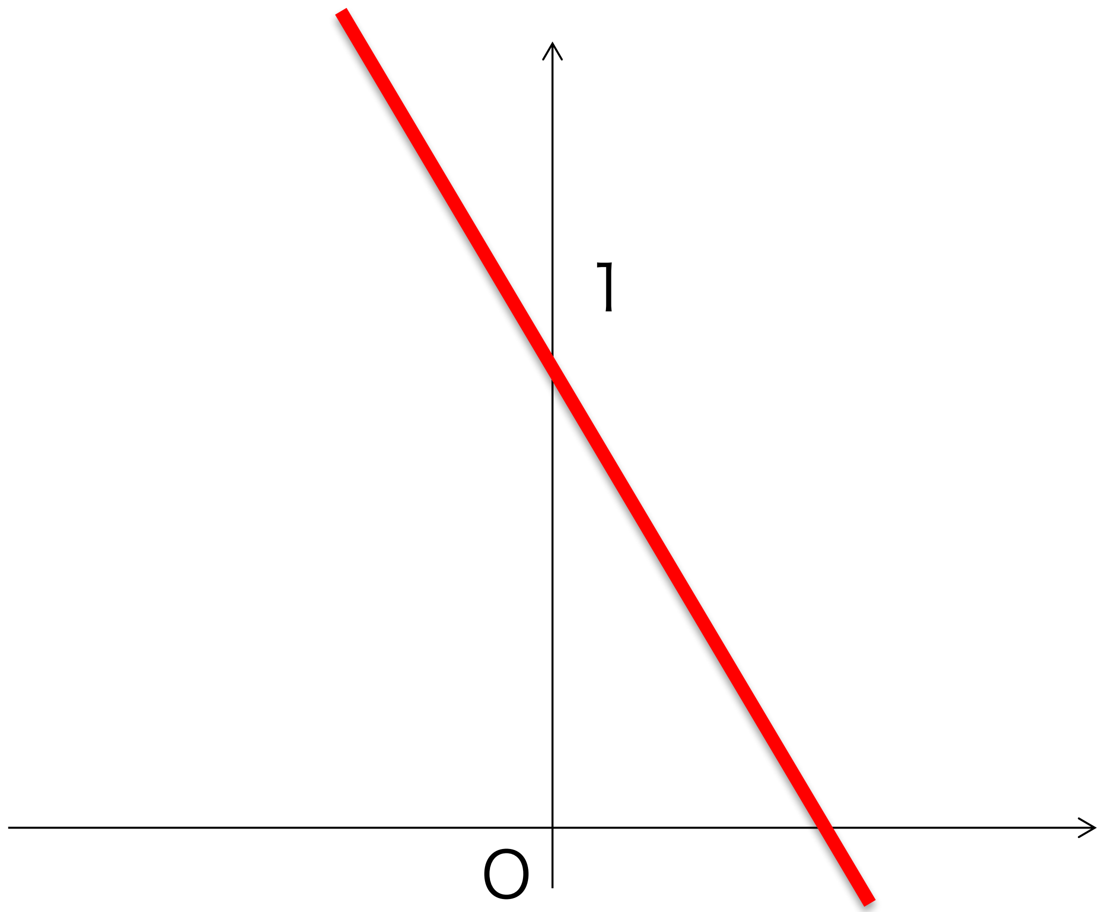
- x: 入力、y:出力、出力が出てきた原因・要因・本質xを知りたい
- 本質を探るために必要な方法論=スパースモデリング

要するに
Xを求める話

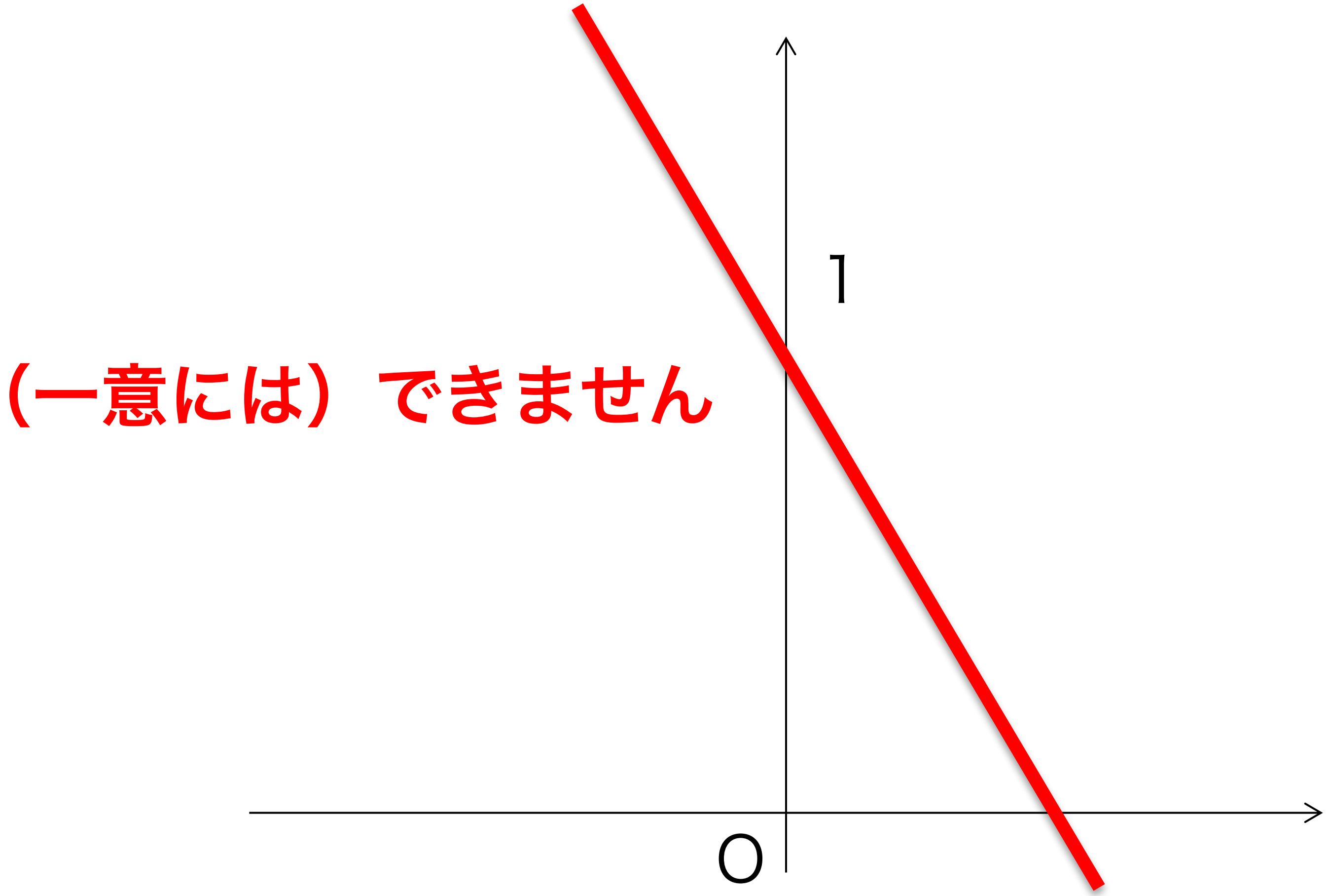
要するに
Xを求める話
連立方程式


1
O


1
交点＝解
O



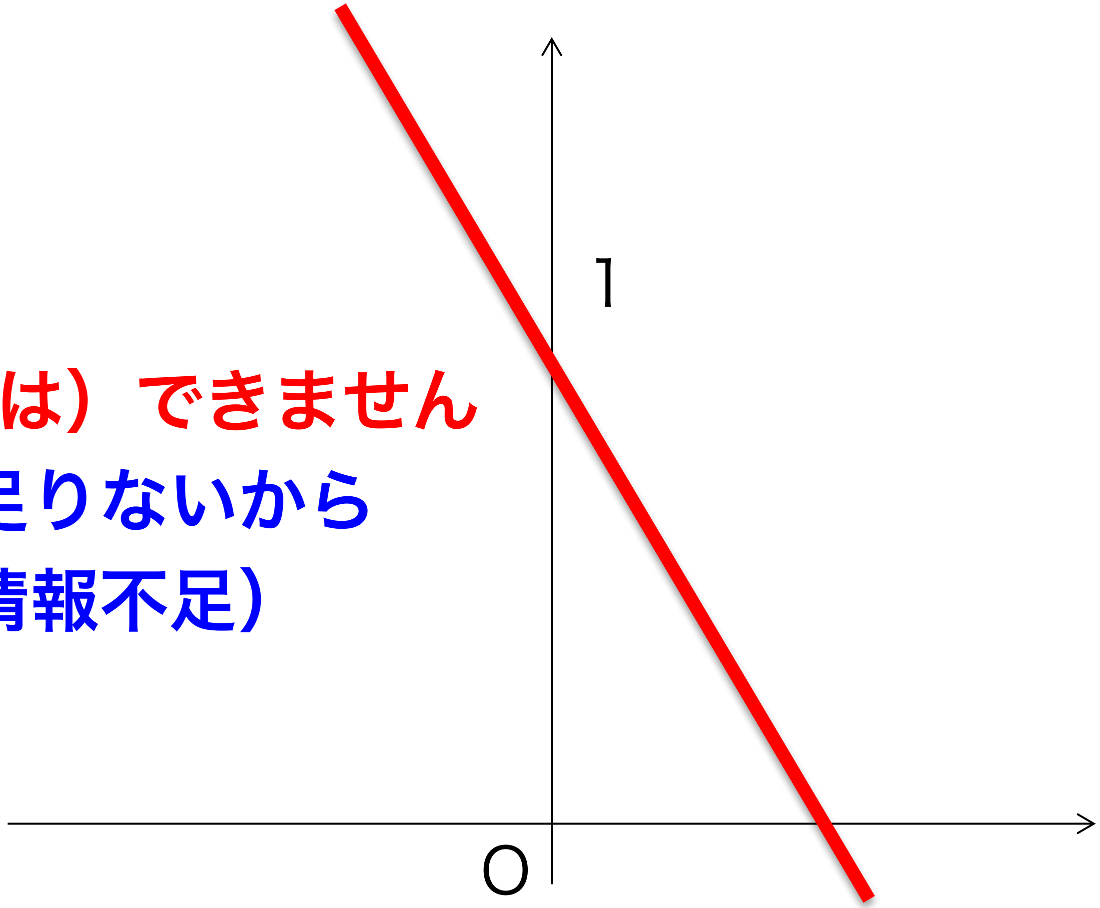
1
O


1
（一意には）できません
O

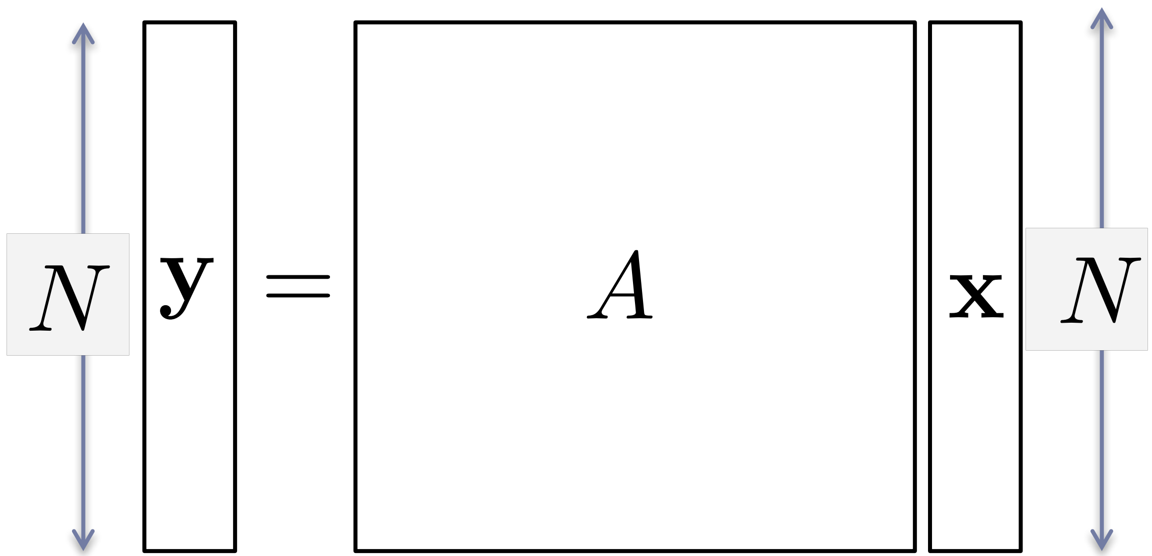
（一意には）できません

式が足りないから

（情報不足）

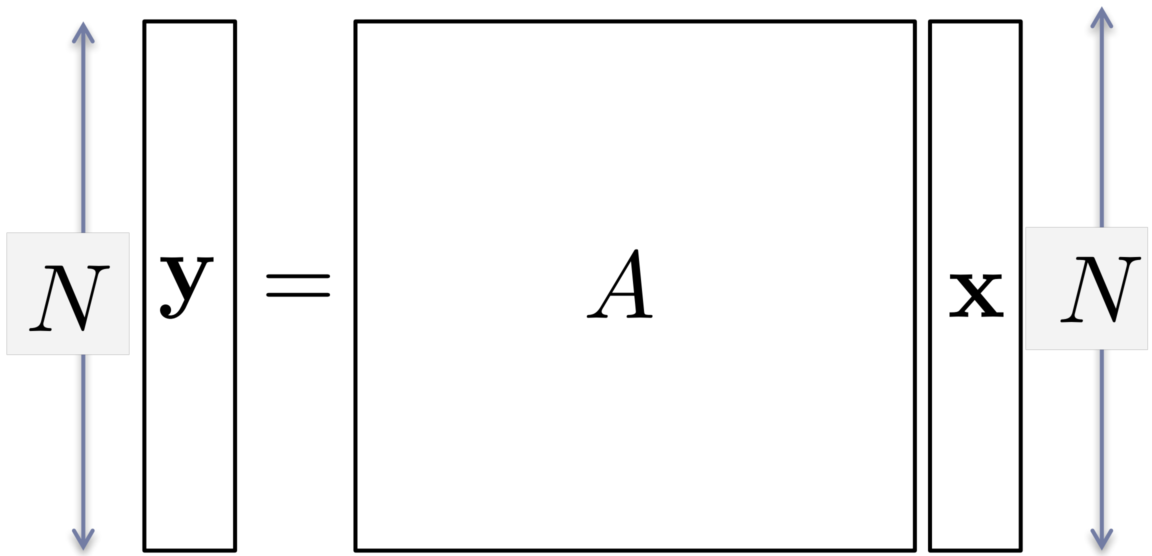


本質をつかむためには情報が必要
これが常識



$$\mathbf{y} = A\mathbf{x}$$

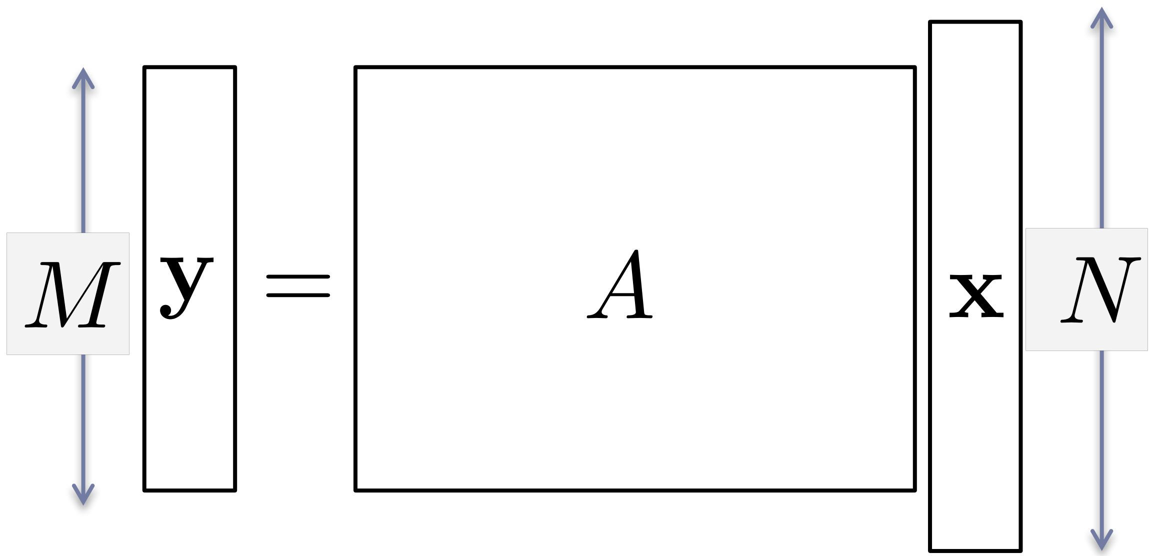
を満たす解を求めよ



$$\mathbf{y} = A\mathbf{x}$$

を満たす解を求めよ

逆行列があるからOK



$$\mathbf{y} = A\mathbf{x}$$

を満たす解を求めよ

劣決定系方程式でもXを当てる技術

圧縮センシング

そんなことがなぜ可能か？

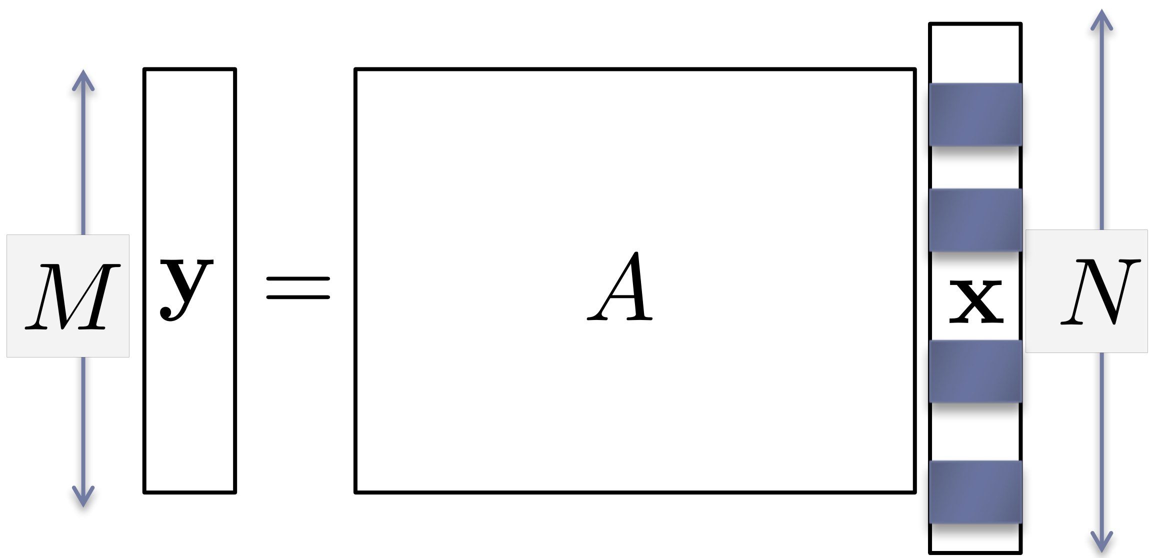
そんなことがなぜ可能か？

スパース性

そんなことがなぜ可能か？

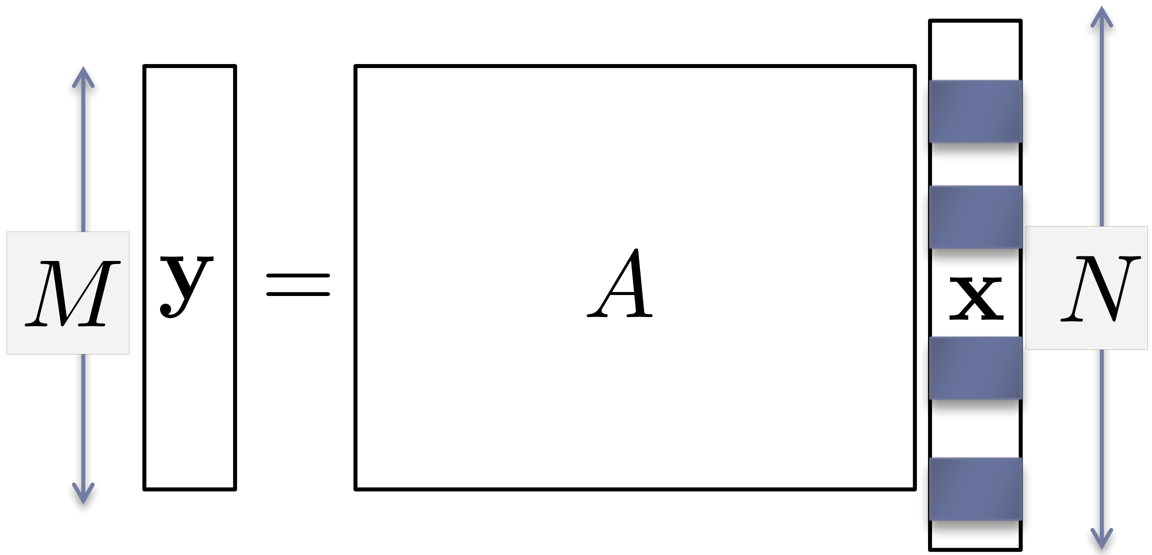
スパース性

ほとんどの要素は不要だった



$$\mathbf{y} = A\mathbf{x}$$

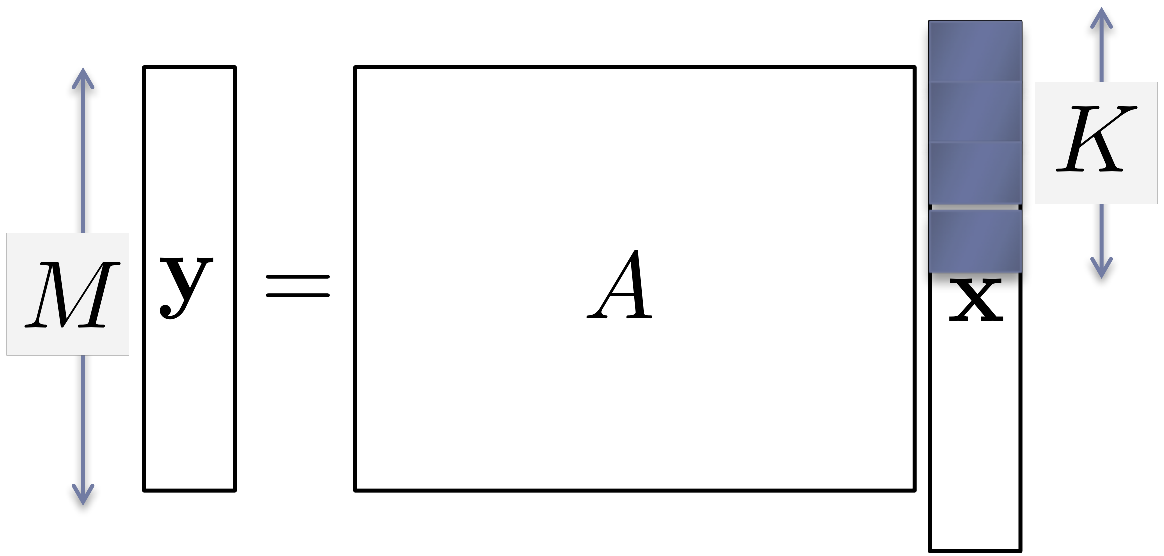
を満たす解を求めよ



$$\mathbf{y} = A\mathbf{x}$$

を満たす解を求めよ

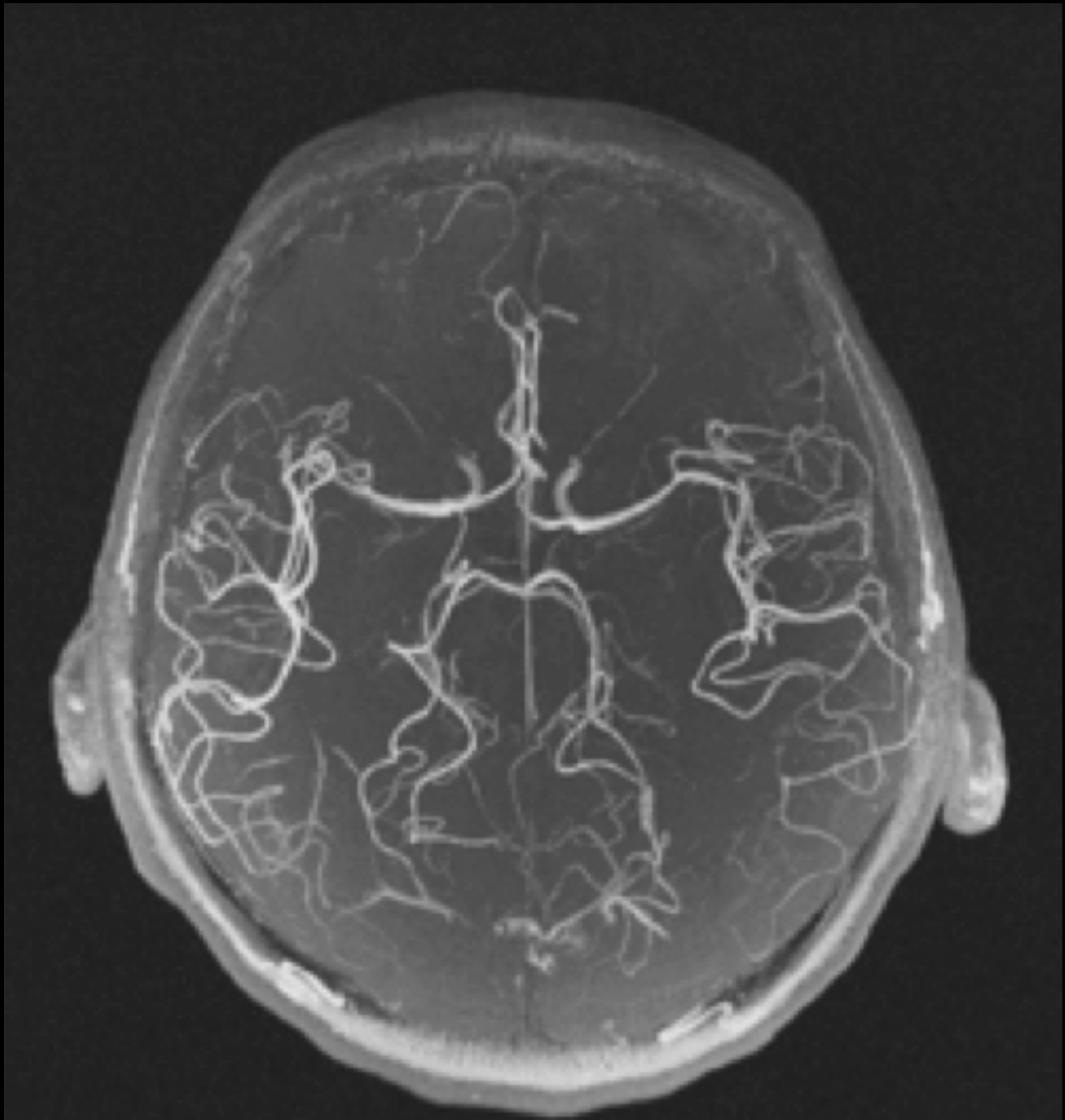
できる！



$$\mathbf{y} = A\mathbf{x}$$

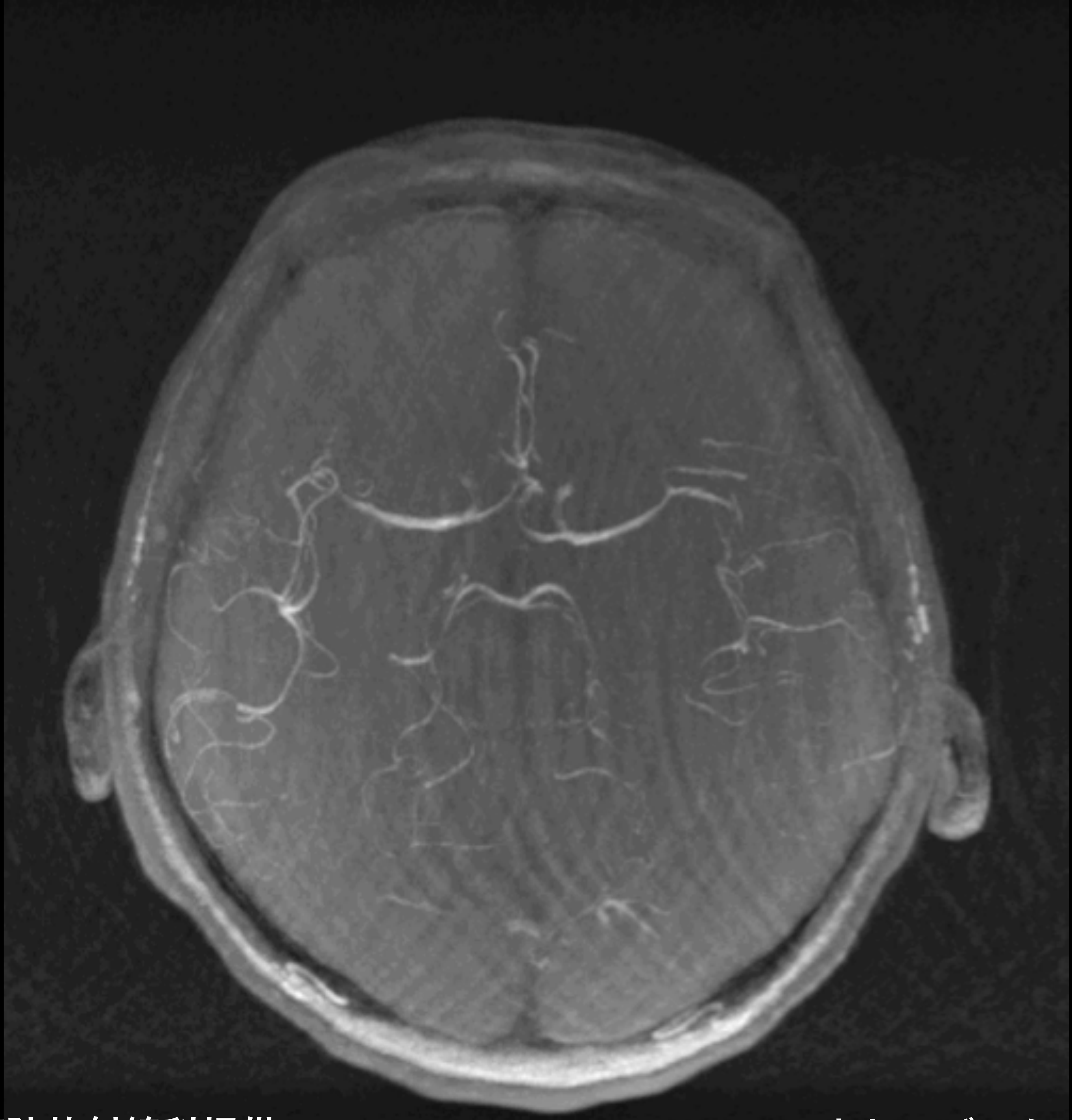
を満たす解を求めよ

できる！



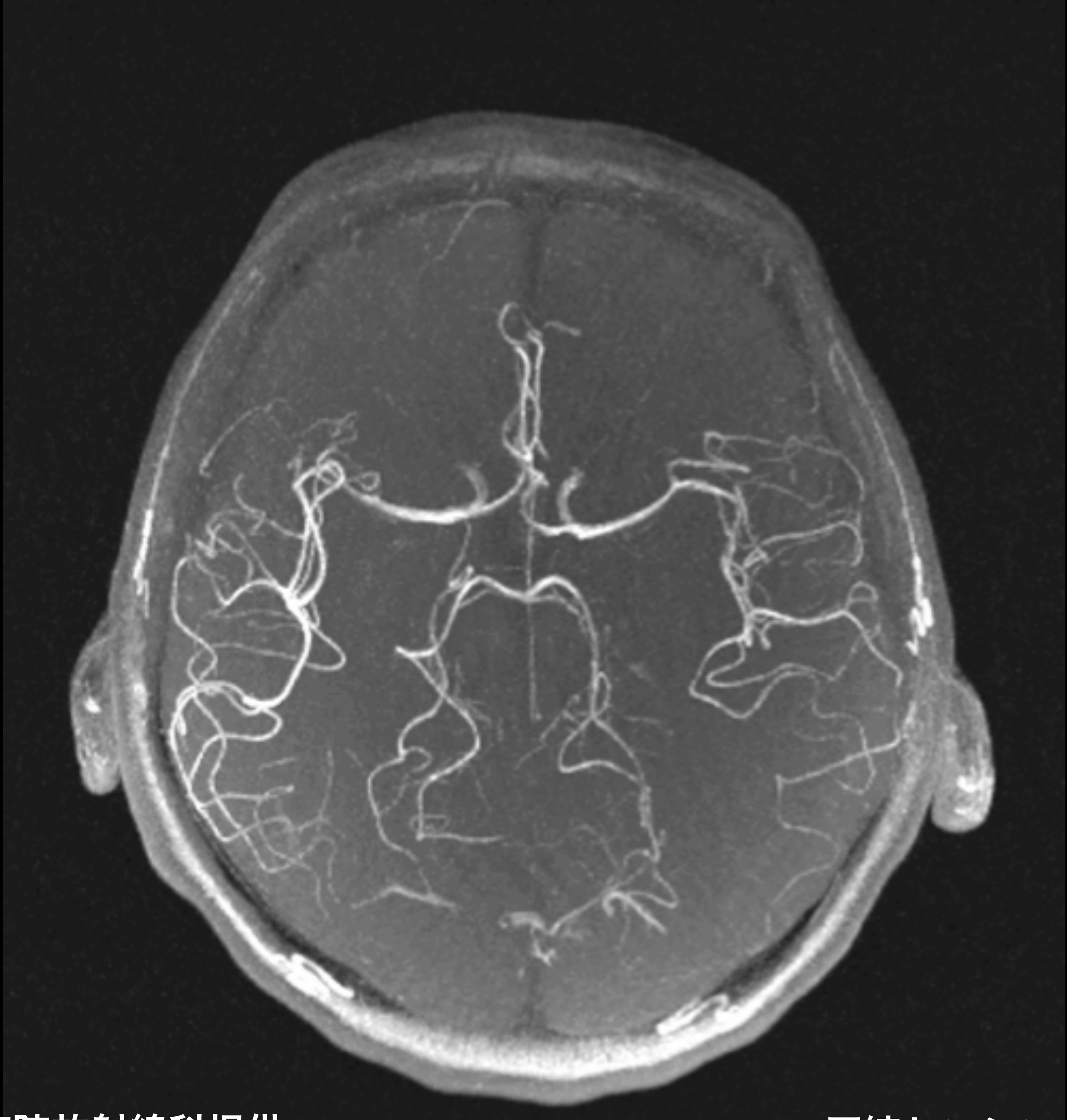
京都大学病院放射線科提供

MRI画像の例



京都大学病院放射線科提供

少ないデータ



京都大学病院放射線科提供

圧縮センシング利用

天文学＋圧縮センシング

K. Akiyama, et al. Astrophys. J. (2017)

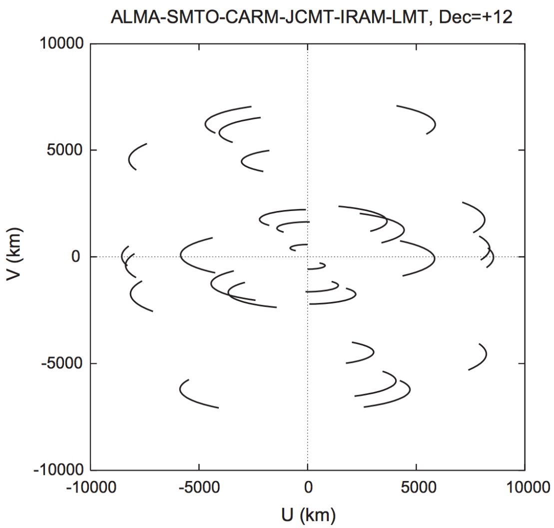
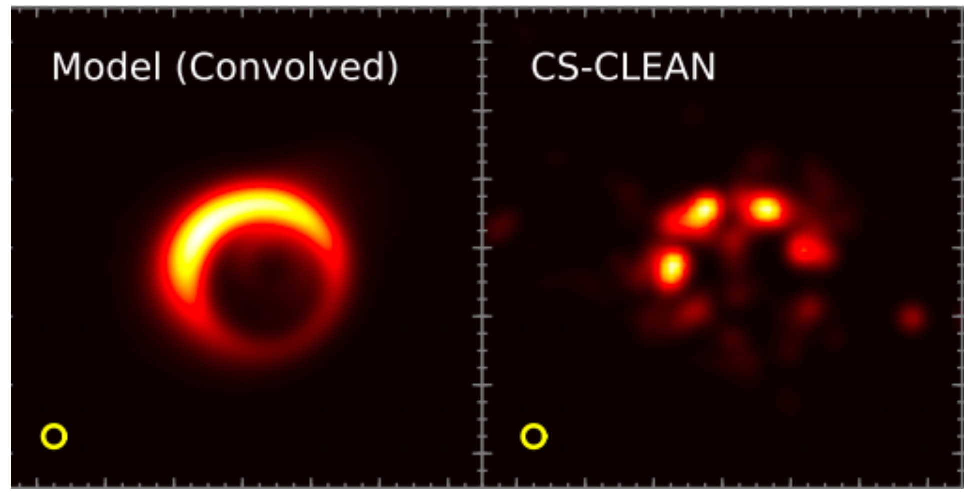
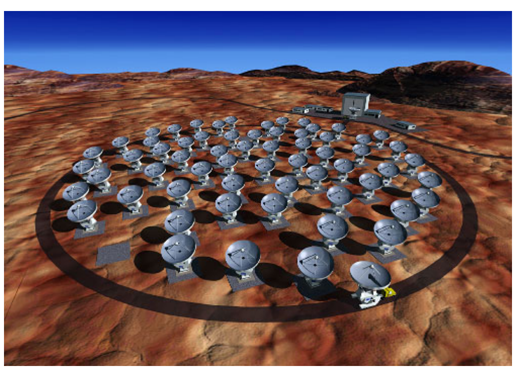


Fig. 5. Simulated UV coverage of M87 with six-station sub-mm VLBI array of EHT. Here it is assumed that observations are conducted at an elevation larger than 20° at each station.





CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

天文学＋圧縮センシング

K. Akiyama, et al. Astrophys. J. (2017)

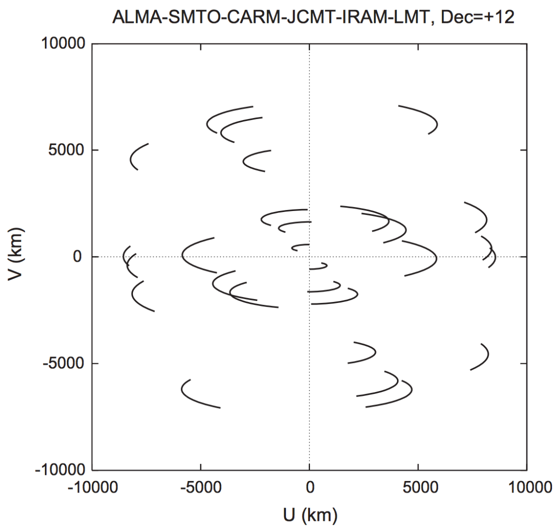
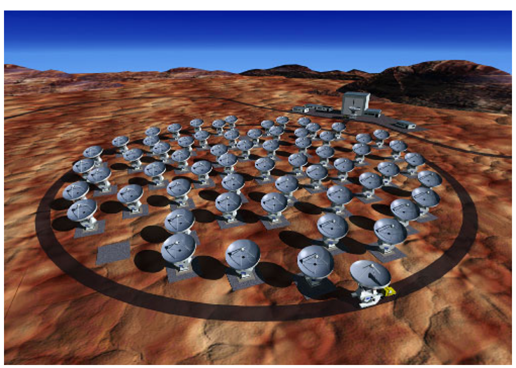
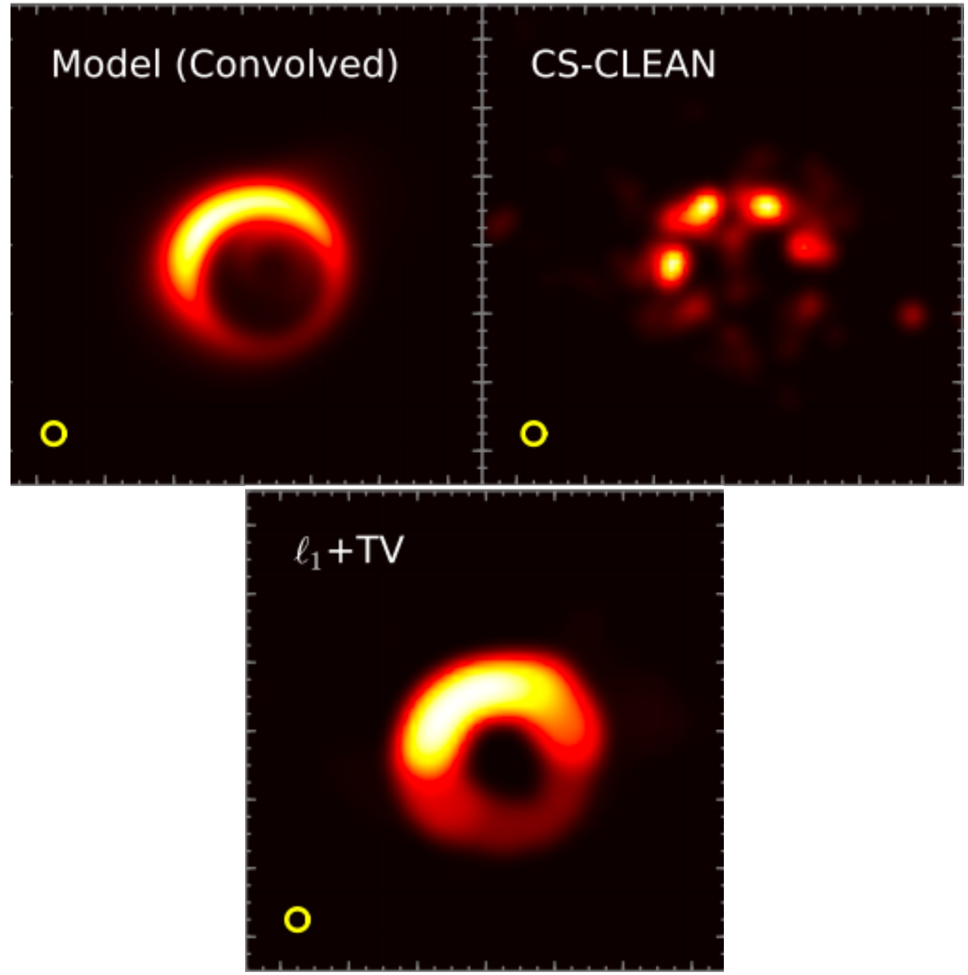


Fig. 5. Simulated UV coverage of M87 with six-station sub-mm VLBI array of EHT. Here it is assumed that observations are conducted at an elevation larger than 20° at each station.





不足した観測情報＋推定方法

圧縮センシング

スパース解推定

- L0ノルム最小化によるスパース解推定
  - 以下の最小化問題でスパース解を探索しよう！

$$\boxed{\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}}$$

スパース解推定

- L0ノルム最小化によるスパース解推定
  - 以下の最小化問題でスパース解を探索しよう！

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

L0ノルム=非零の個数

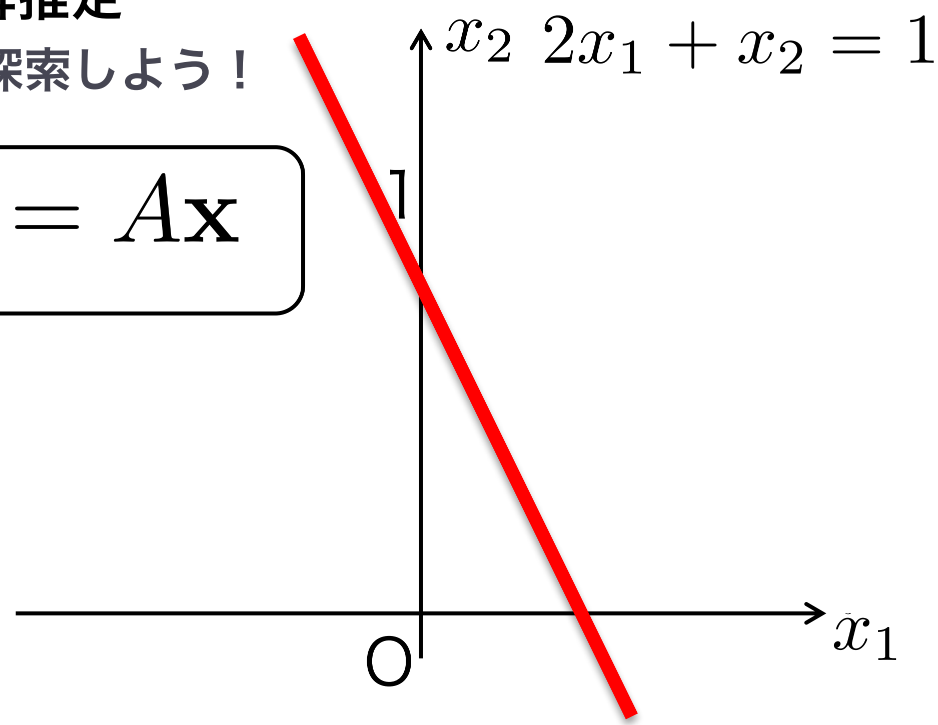
CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

スパース解推定

- **L0ノルム最小化によるスパース解推定**
  - 以下の最小化問題でスパース解を探索しよう！

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

L0ノルム=非零の個数

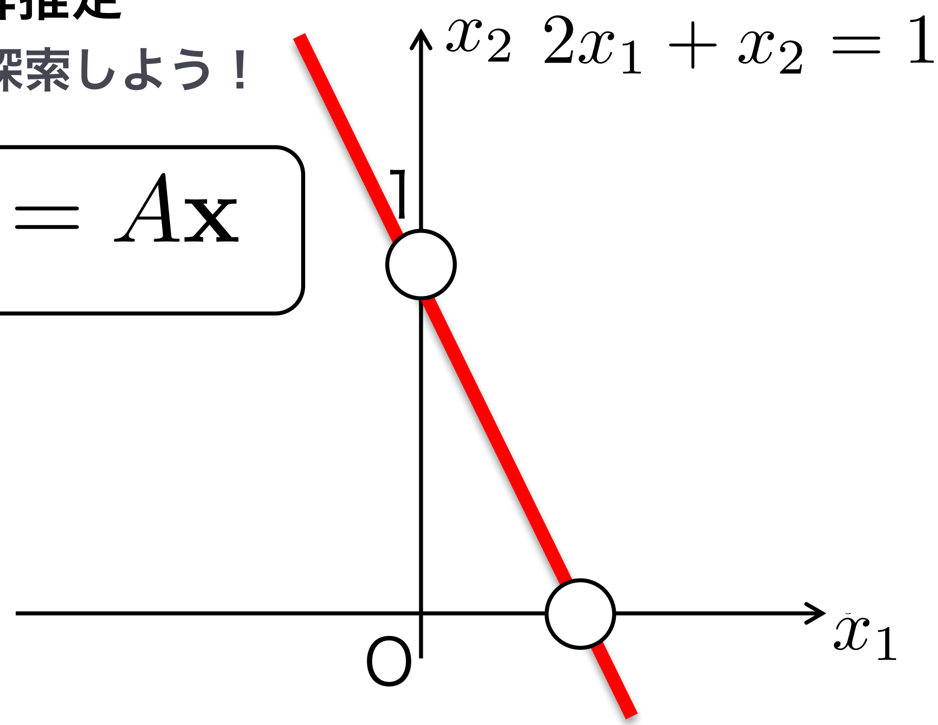


スパース解推定

- **L0ノルム最小化によるスパース解推定**
  - 以下の最小化問題でスパース解を探索しよう！

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

L0ノルム=非零の個数



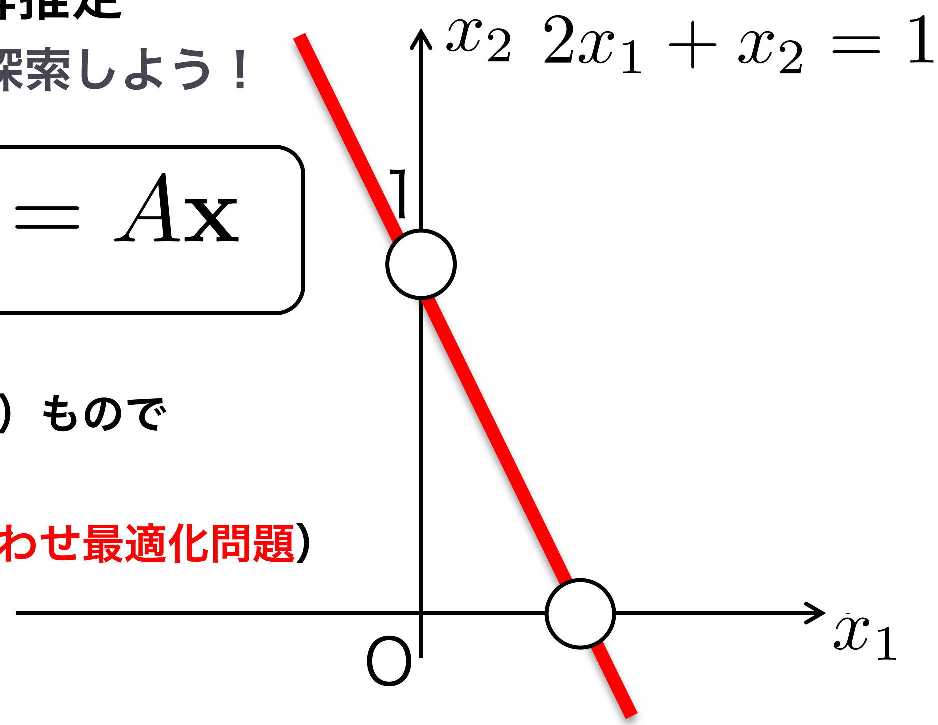
スパース解推定

- L0ノルム最小化によるスパース解推定
  - 以下の最小化問題でスパース解を探索しよう！

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

L0ノルム=**非零の個数**

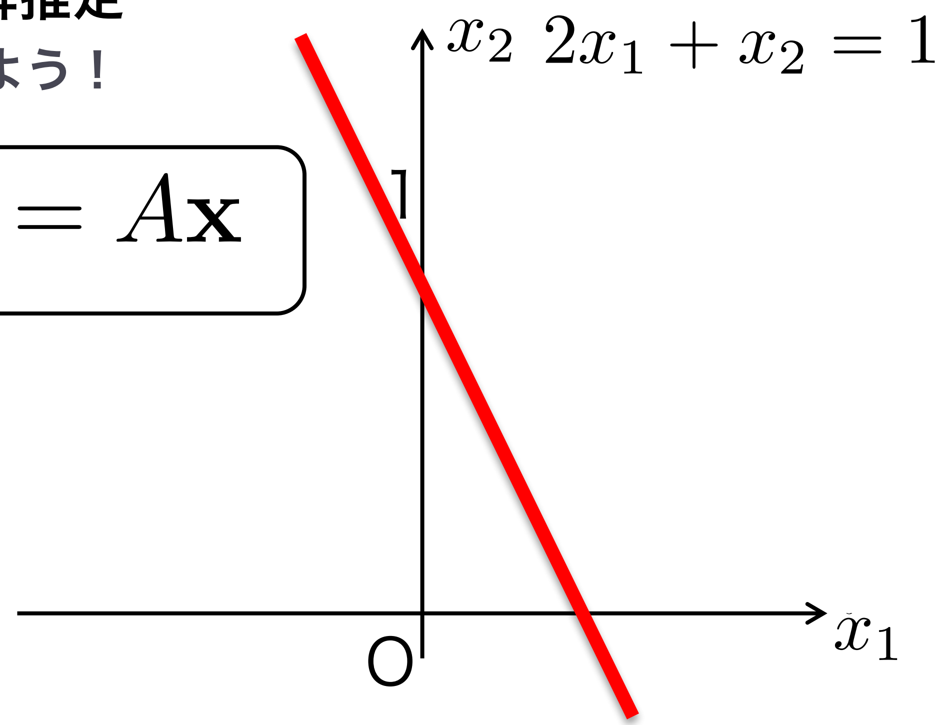
- 非零成分の個数が小さい（スパース）もので方程式を満たすものを探す
- 残念ながら計算量的に困難（**組み合わせ最適化問題**）



スパース解推定

- L1ノルム最小化によるスパース解推定
  - 代わりに次の最小化問題で探索しよう！

$$\boxed{\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}}$$

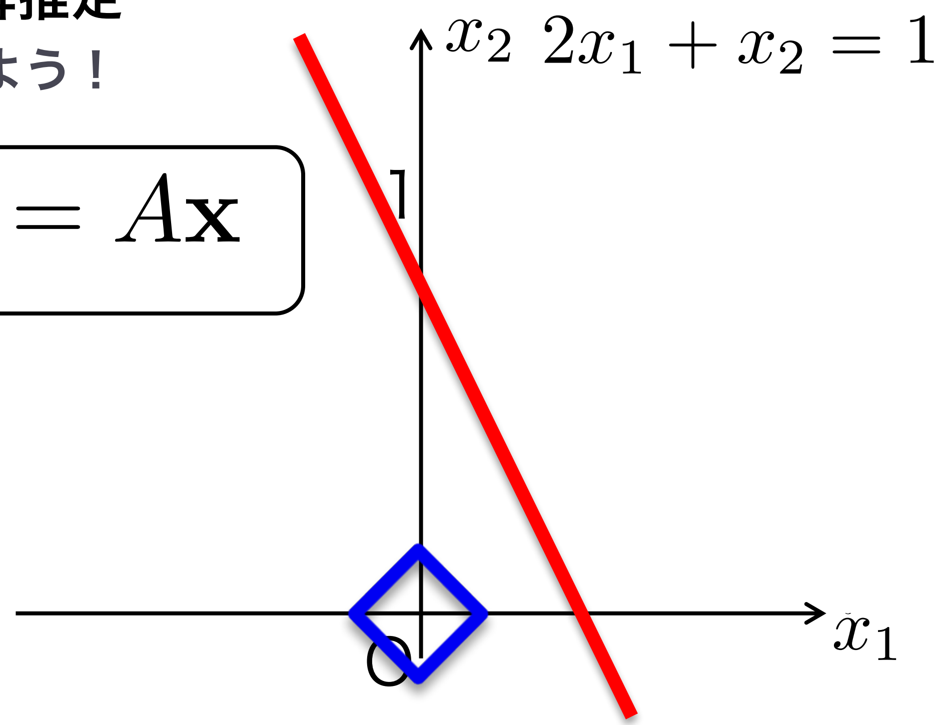


$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_N|$$

スパース解推定

- L1ノルム最小化によるスパース解推定
  - 代わりに次の最小化問題で探索しよう！

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

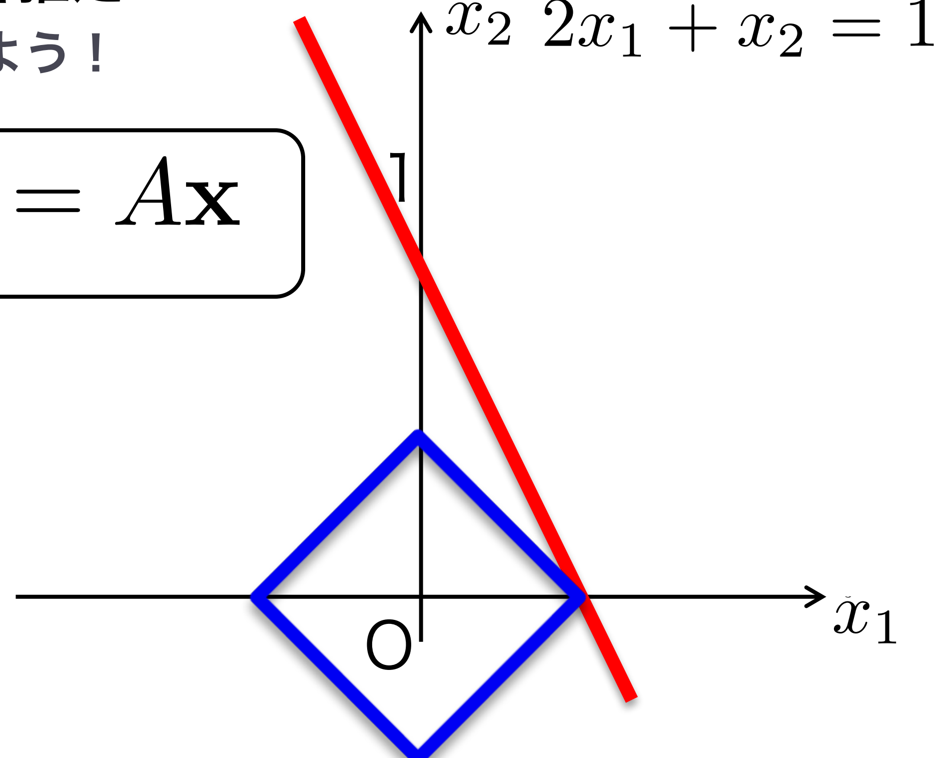


$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_N|$$

スパース解推定

- L1ノルム最小化によるスパース解推定
  - 代わりに次の最小化問題で探索しよう！

$$\boxed{\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}}$$



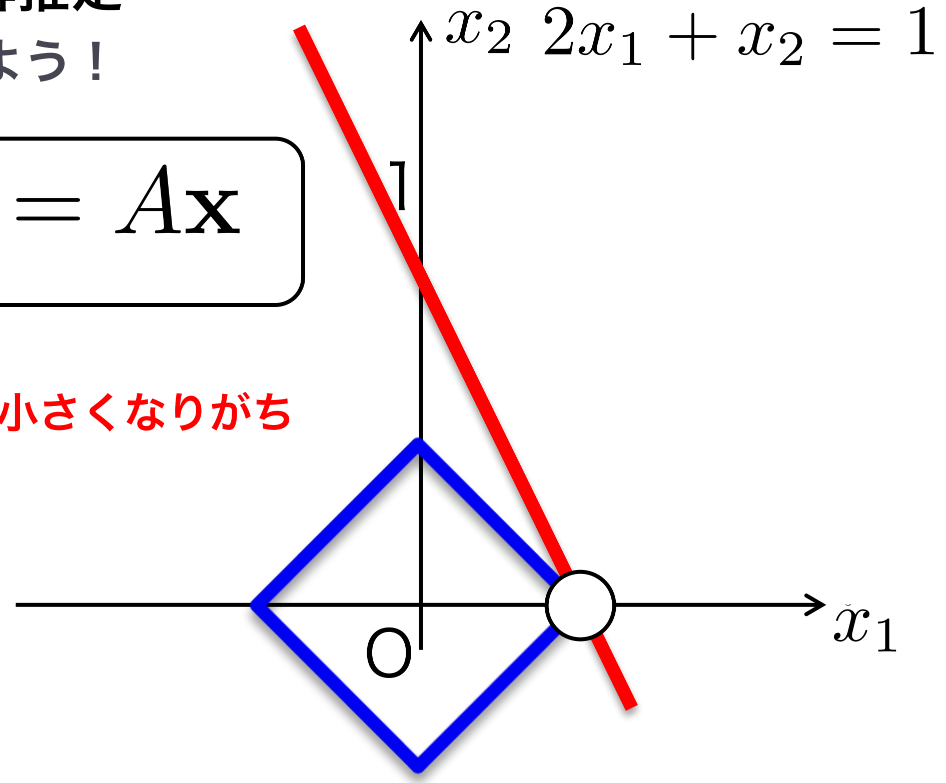
$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_N|$$

スパース解推定

- L1ノルム最小化によるスパース解推定
  - 代わりに次の最小化問題で探索しよう！

$$\boxed{\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}}$$

- 零が多ければ小さくなる＋大きさも小さくなりがち
- 計算量は非常に軽い（Nの3乗程度）



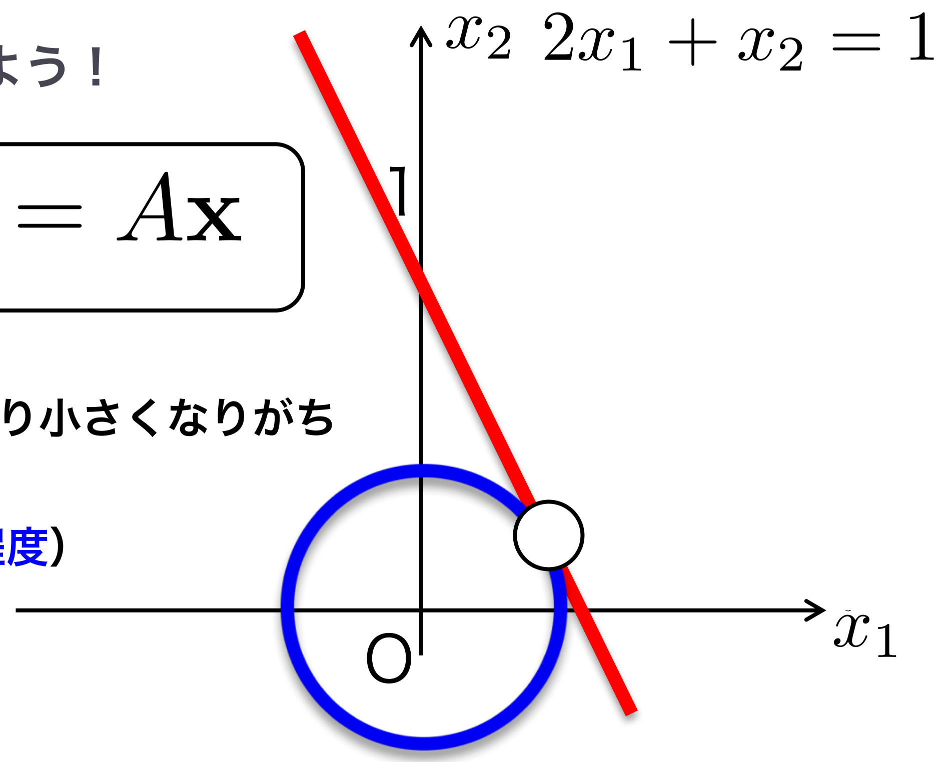
$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_N|$$

QUANTUM ANNEALING
MACHINE LEARNING

スパース解推定

- L2ノルムでは？？
  - 代わりに次の最小化問題で探索しよう！

$$\boxed{\min_{\mathbf{x}} \|\mathbf{x}\|_2 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}}$$



- 零が多ければ小さい＋大きさがかなり小さくなりがち
  +小さいかすが残る
- 計算量は非常に軽い（Nの2〜3乗程度）

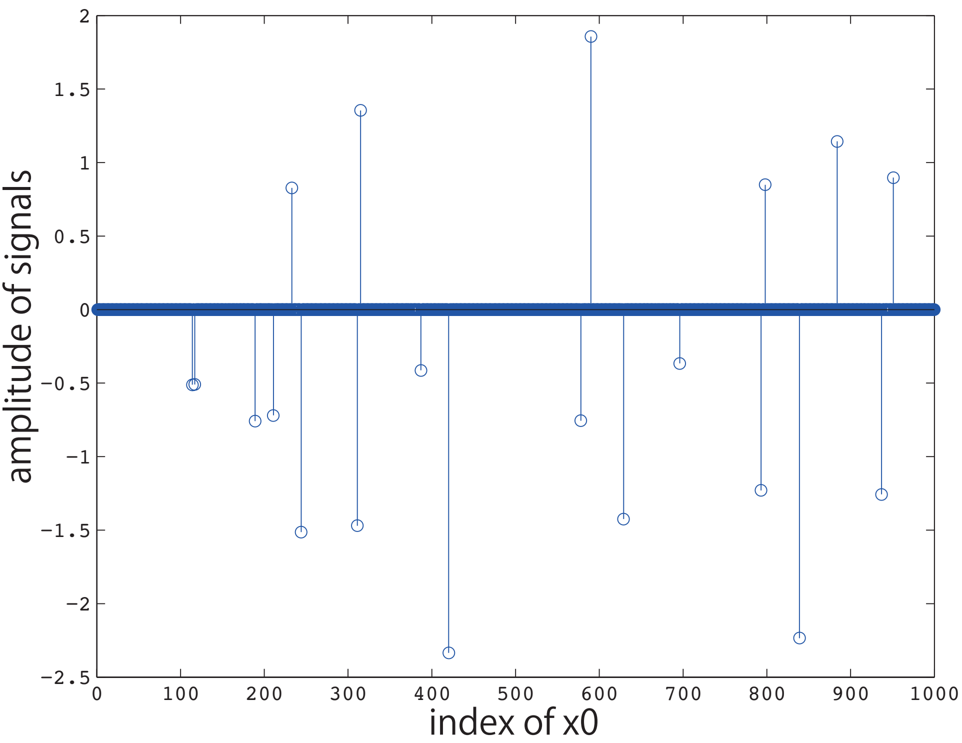
$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_N^2}$$

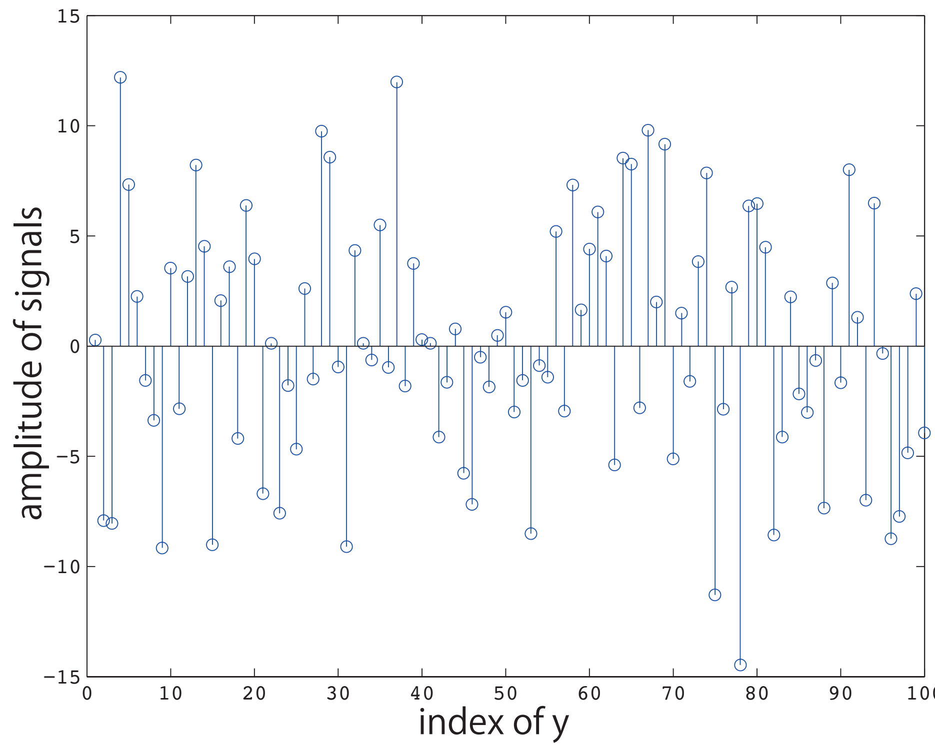
L1ノルムはスパースな解の選択法

L1ノルムはスパースな解の選択法
当たるかどうか？

実践例

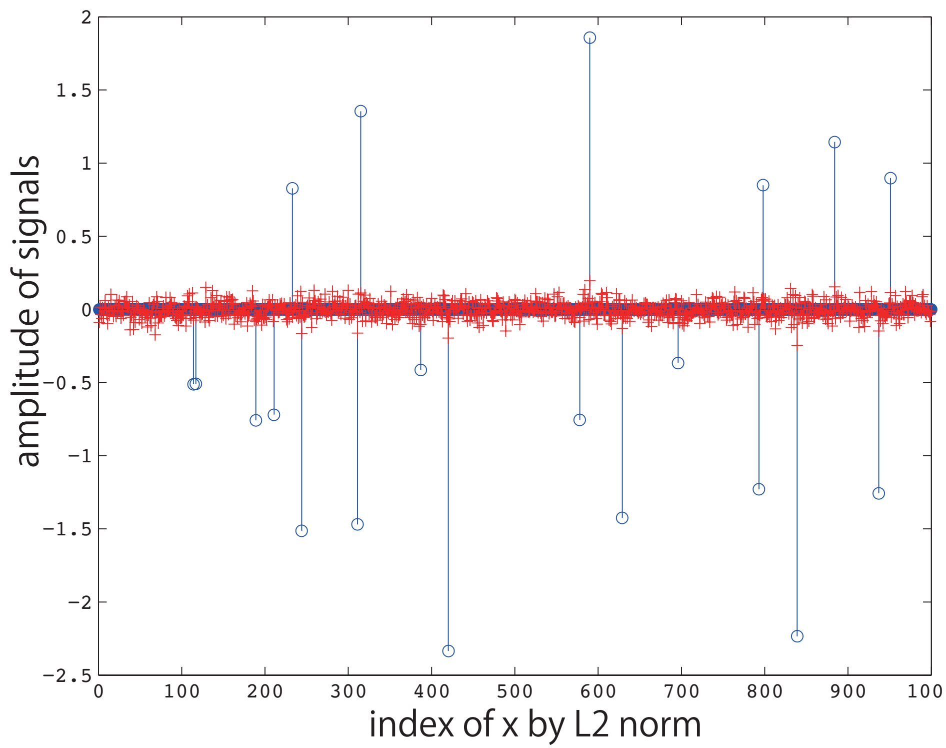
- M=100,N=1000,K=20,ガウスランダム観測行列の場合





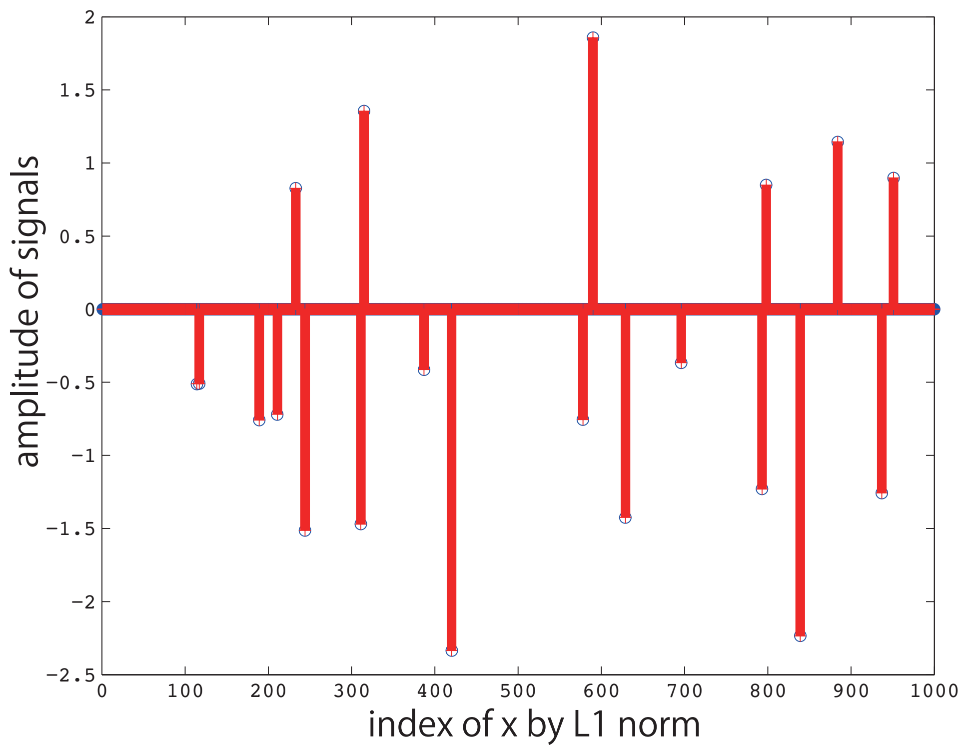
実践例

- L2ノルムを利用した場合



実践例

- L1ノルムを利用した場合



L1ノルム最小化の性能

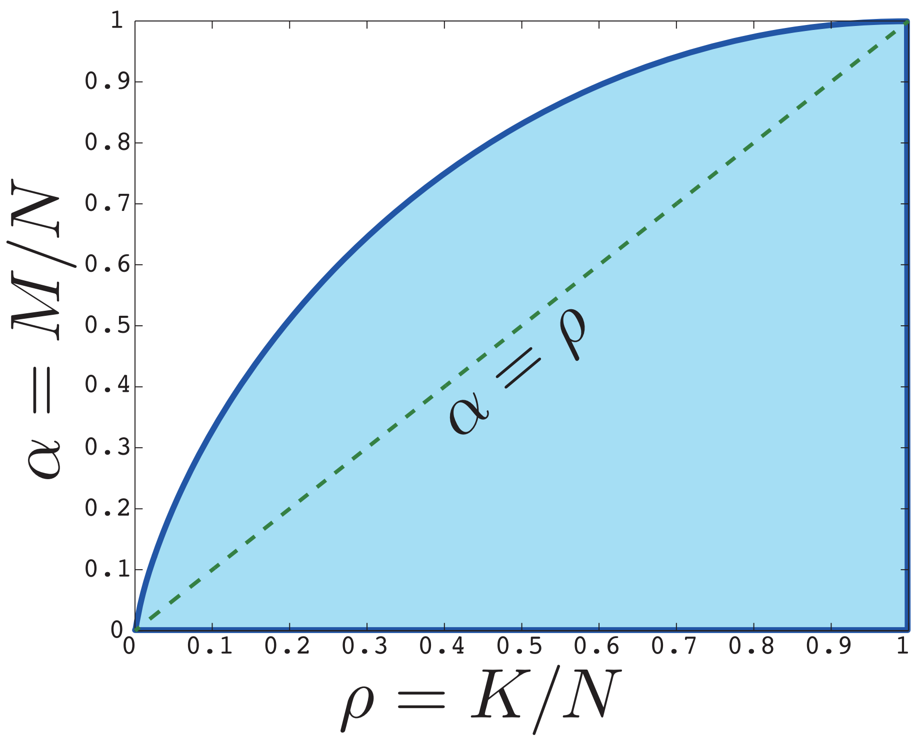
- L1ノルム最小化によるスパース解推定法
  - 前提：A=ガウスランダム行列、原信号はガウスランダム

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

$$\frac{1}{\alpha} = 1 + \sqrt{\frac{\pi}{2}} t e^{\frac{t^2}{2}} \{1 - 2Q(t)\}$$

$$\frac{\rho}{1-\rho} = 2\left(\frac{e^{\frac{t^2}{2}}}{t\sqrt{2\pi}} - Q(t)\right)$$

$$Q(t) = \int_t^{\infty} \frac{e^{\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$



- 解析方法は積分幾何学や統計力学

[D. L. Dohono and J. Tanner: Proc. Nat. Acad. Sci. 102 (2005) 9452]
[Y. Kabashima, T. Wadayama, and T. Tanaka: J. Stat. Mech.: Theor. and Exp. 09 (2009) L09003]

スパースモデリング

Find x by sparsity

スパースモデリング

Find x by sparsity

Make x sparse

どうやって？

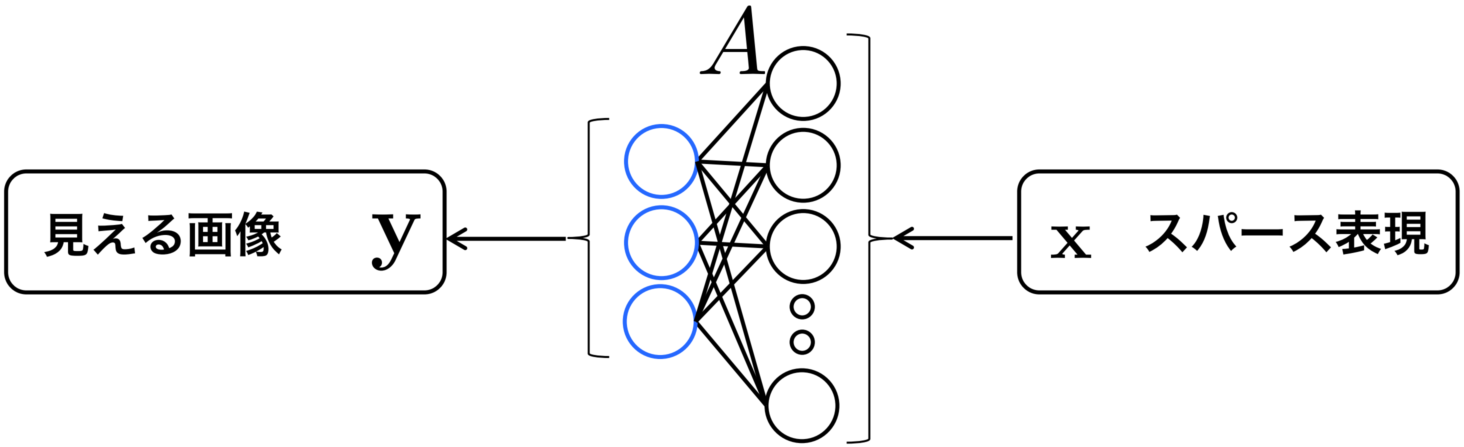
機械学習

$$y = f(\mathbf{x})$$

辞書学習でスパースな基底を

J. Mairal, F. Bach, F. Ponce, and G. Sapiro: ICML (2009)

- ノイズや抜けに対して頑強な基底を獲得



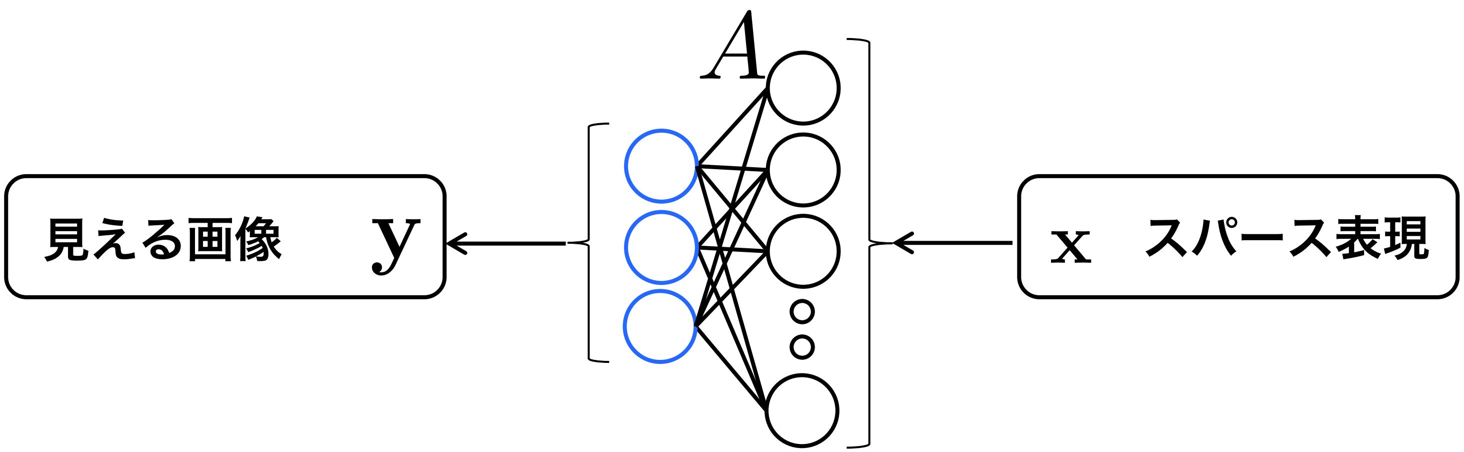
$$\min_{A,\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}$$

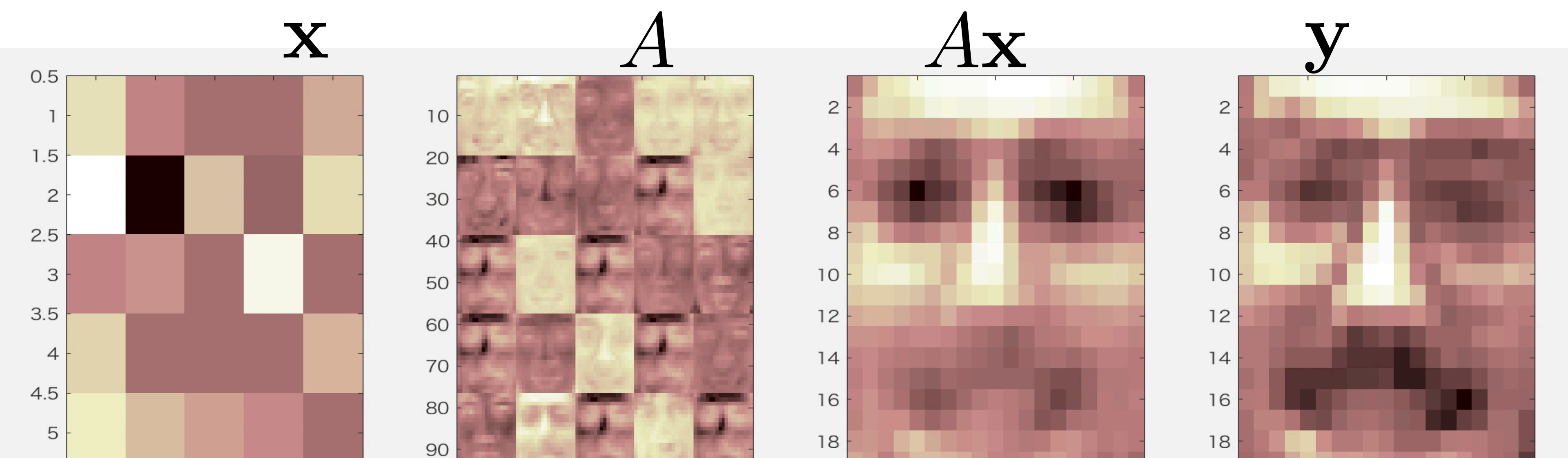
- スパースなxとなるようなAを求める
- yが与えられているだけで良い

辞書学習でスパースな基底を

J. Mairal, F. Bach, F. Ponce, and G. Sapiro: ICML (2009)

- ノイズや抜けに対して頑強な基底を獲得





圧縮センシング実践編

何をすれば良いのか？？

- **最適化問題を解けば良い！**
  - 解きたい最適化問題

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x}$$

  - **罰金法**により等価な最適化問題

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}$$

    - LASSO（Least Absolute Shrinkage and Selection Operators)
    - 絶対値関数が含まれるので難しそう＝**単なる印象**にすぎない

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{v}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}$$

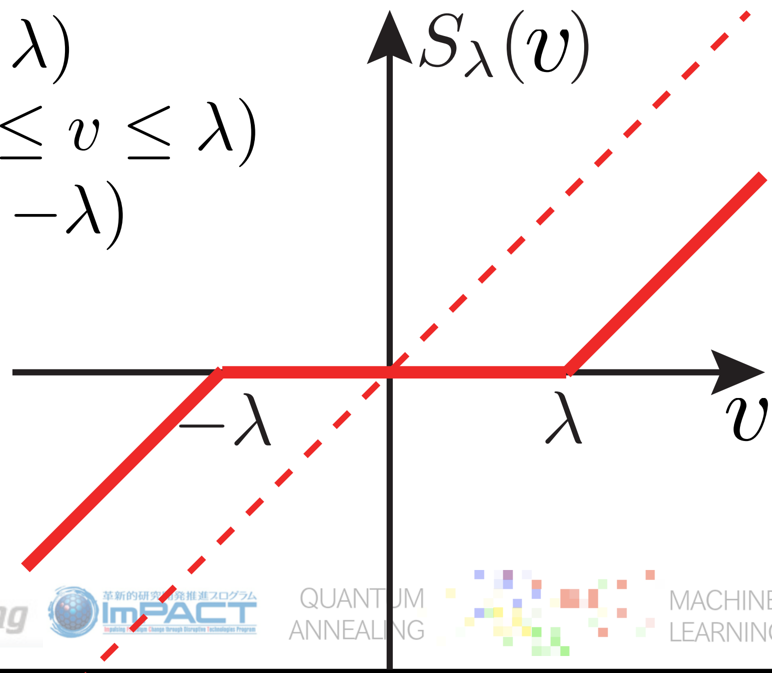
1変数なら解ける！

- **軟判定しきい値関数の導入**

$$\min_x \left\{ \frac{1}{2}(x - v)^2 + \lambda |x| \right\}$$

- **この最小化問題の解は、以下の軟判定しきい値関数で与えられる.**

$$x^* = S_\lambda(v) = \begin{cases} v - \lambda & (v > \lambda) \\ 0 & (-\lambda \le v \le \lambda) \\ v + \lambda & (v < -\lambda) \end{cases}$$



何をすれば良いのか？？

- これなら解ける
  - 仮に以下の最適化問題であれば、**微分なし**で解ける

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \left\| \mathbf{x} - \mathbf{v} \right\|_2^2 + \lambda \left\| \mathbf{x} \right\|_1 \right\}$$

  - 軟判定しきい値関数の導入

$$x_i^* = S_\lambda(v_i) = \begin{cases} v_i - \lambda & (\lambda \leq v_i) \\ 0 & (-\lambda < v_i < \lambda) \\ v_i + \lambda & (v_i \leq -\lambda) \end{cases}$$

  - もともとの最適化問題の代わりに**解きやすい形**にしよう

最近の進展：ADMM [Alternating Direction of Multiplier method)

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- 勾配法に頼らない高速なアルゴリズム
  - 解きたい問題が2つ以上の関数の組み合わせとする

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + g(\mathbf{x}) \right\}$$

- Ex) LASSO

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \qquad g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$$

CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

最近の進展：ADMM [Alternating Direction of Multiplier method)

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- 勾配法に頼らない高速なアルゴリズム
  - 解きたい問題が2つ以上の関数の組み合わせとする

$$\min_{\mathbf{x}} \{ f(\mathbf{x}) + g(\mathbf{x}) \}$$

  - Ex) LASSO

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \qquad g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$$

  - あえてふたつの変数の問題に分割（Spilitting）

$$\min_{\mathbf{x},\mathbf{z}} \{ f(\mathbf{x}) + g(\mathbf{z}) \} \quad \text{s.t.} \quad \mathbf{x} = \mathbf{z}$$

CREST　Sparse Modeling　革新的研究開発推進プログラム ImPACT　QUANTUM ANNEALING　MACHINE LEARNING

最近の進展：ADMM [Alternating Direction of Multiplier method)

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- 勾配法に頼らない高速なアルゴリズム
  - 解きたい問題が2つ以上の関数の組み合わせとする

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + g(\mathbf{x}) \right\}$$

  - Ex) LASSO

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \qquad g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$$

  - あえてふたつの変数の問題に分割（Spilitting）

$$\min_{\mathbf{x},\mathbf{z}} \left\{ f(\mathbf{x}) + g(\mathbf{z}) \right\} \quad \mathrm{s.t.} \quad \mathbf{x} = \mathbf{z}$$

  - 拡張ラグランジュ法を利用する！(未定乗数：h、罰金係数：ρ)

$$\min_{\mathbf{x},\mathbf{z},\mathbf{h}} \left\{ f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{h}^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}$$

CREST Sparse Modeling ImPACT QUANTUM ANNEALING MACHINE LEARNING

最近の進展：ADMM [Alternating Direction of Multiplier method)

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- 勾配法に頼らない高速なアルゴリズム
  - 解きたい問題が2つ以上の関数の組み合わせとする

$$\min_{\mathbf{x}} \{ f(\mathbf{x}) + g(\mathbf{x}) \}$$

  - Ex) LASSO

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \qquad g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$$

  - 交互にふたつの変数の最適化問題を解く

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \mathbf{h}^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}$$

$$\min_{\mathbf{z}} \left\{ g(\mathbf{z}) + \mathbf{h}^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}$$

  - ラグランジュ未定乗数を更新 $\mathbf{h} = \mathbf{h} + \rho(\mathbf{x} - \mathbf{z})$

最近の進展：ADMM [Alternating Direction of Multiplier method)

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- 勾配法に頼らない高速なアルゴリズム
  - 解きたい問題が2つ以上の関数の組み合わせとする

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + g(\mathbf{x}) \right\}$$

    - Ex) LASSO

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \qquad g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$$

  - 交互にふたつの変数の最適化問題を解く

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \mathbf{h}^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}$$

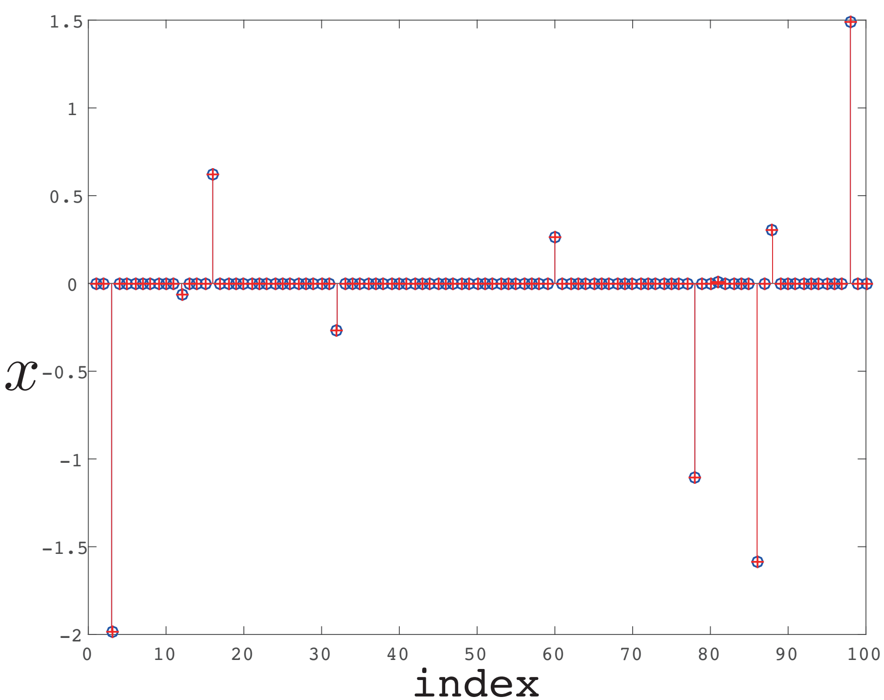
$$\min_{\mathbf{z}} \left\{ \lambda \|\mathbf{z}\|_1 + \mathbf{h}^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}$$

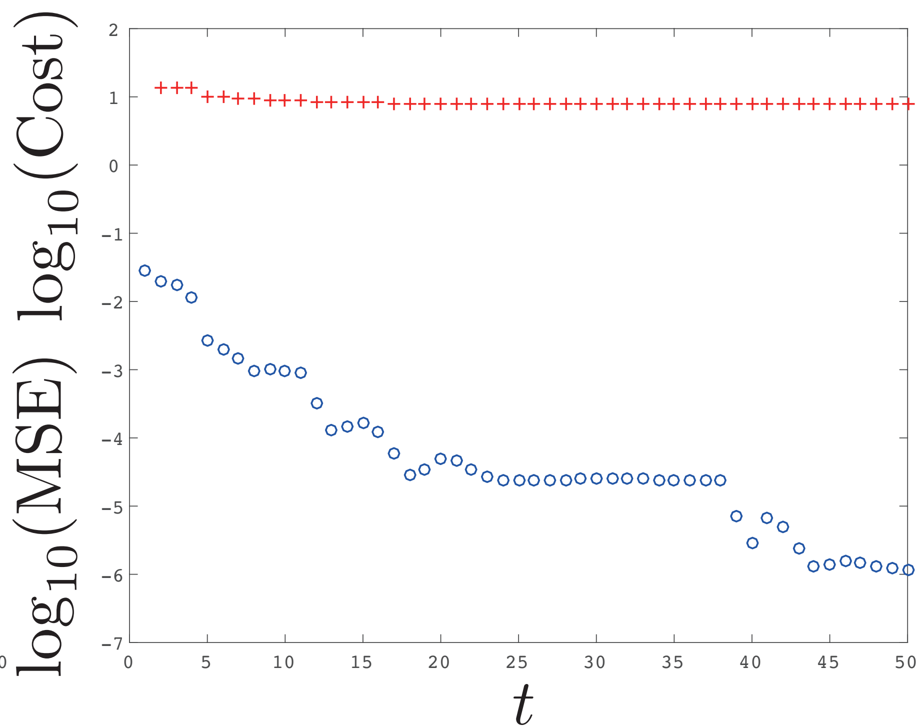
  - ラグランジュ未定乗数を更新 $\mathbf{h} = \mathbf{h} + \rho(\mathbf{x} - \mathbf{z})$

最近の進展：ADMM [Alternating Direction of Multiplier method]

S. Boyd, et al. Foundation and Trends in Machine Learning, 3 (2010) 1

- ADMM（コードは講義ノートに掲載中）

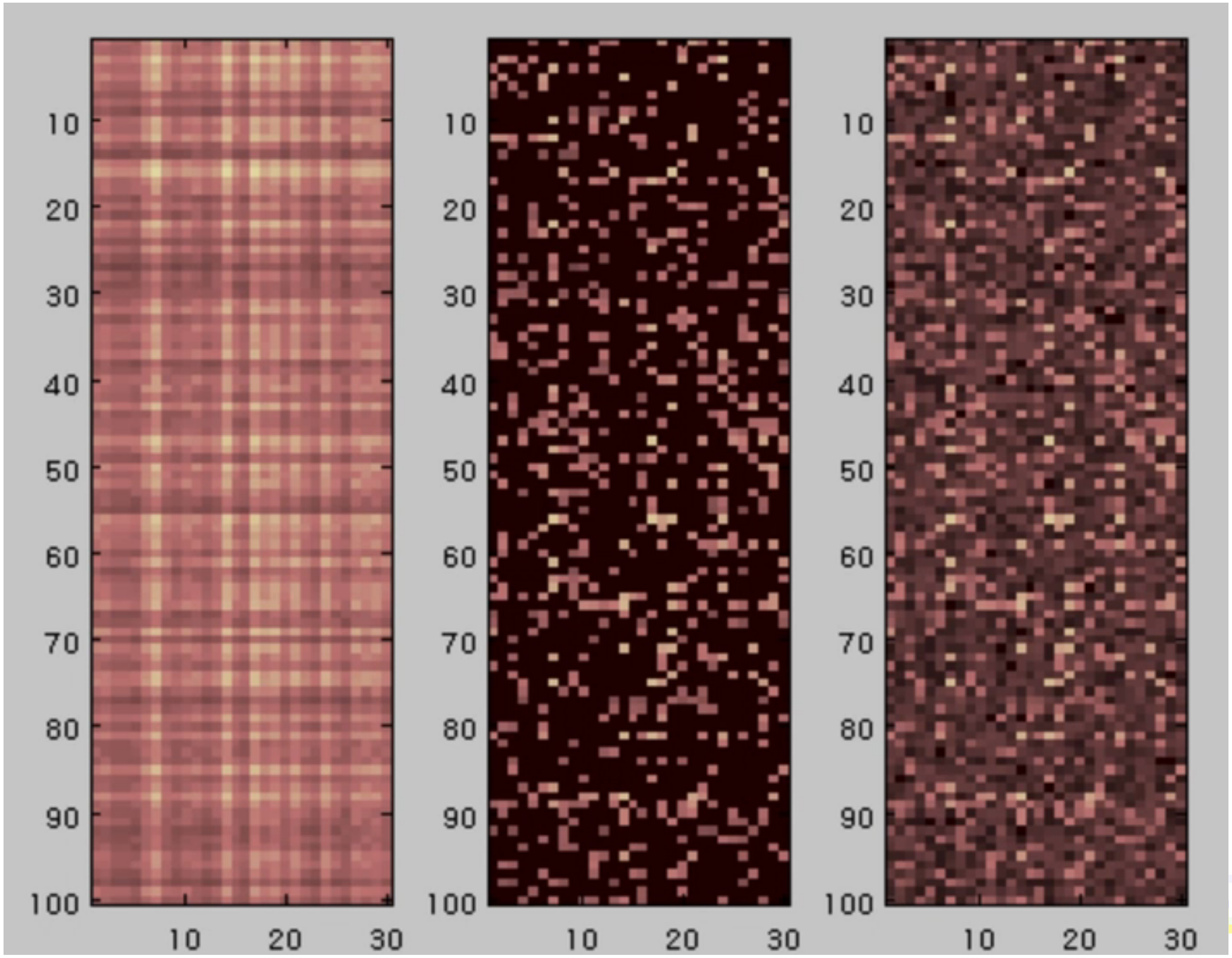




更なる戦略

低ランク性

サービスデータ＋スパースモデリング



まとめ

- **機械学習という現代の魔法**
  - 深層学習
    - 単純な繰り返しによる複雑な関数 f の近似
  - スパースモデリング
    - Find x by sparsity (圧縮センシング)
    - Make x sparse (辞書学習)
  - L1ノルム
    - スパースな解を当てることができる
    - 数学や統計物理による数理的保障がある
  - ADMM (alternating direction of multiplier method)
    - 拡張ラグランジュ法による最適化手法
    - 複数のコスト関数があっても良い
- **鍵は最適化問題**
  - 世界各国で最適化専門のマシン・チップ登場