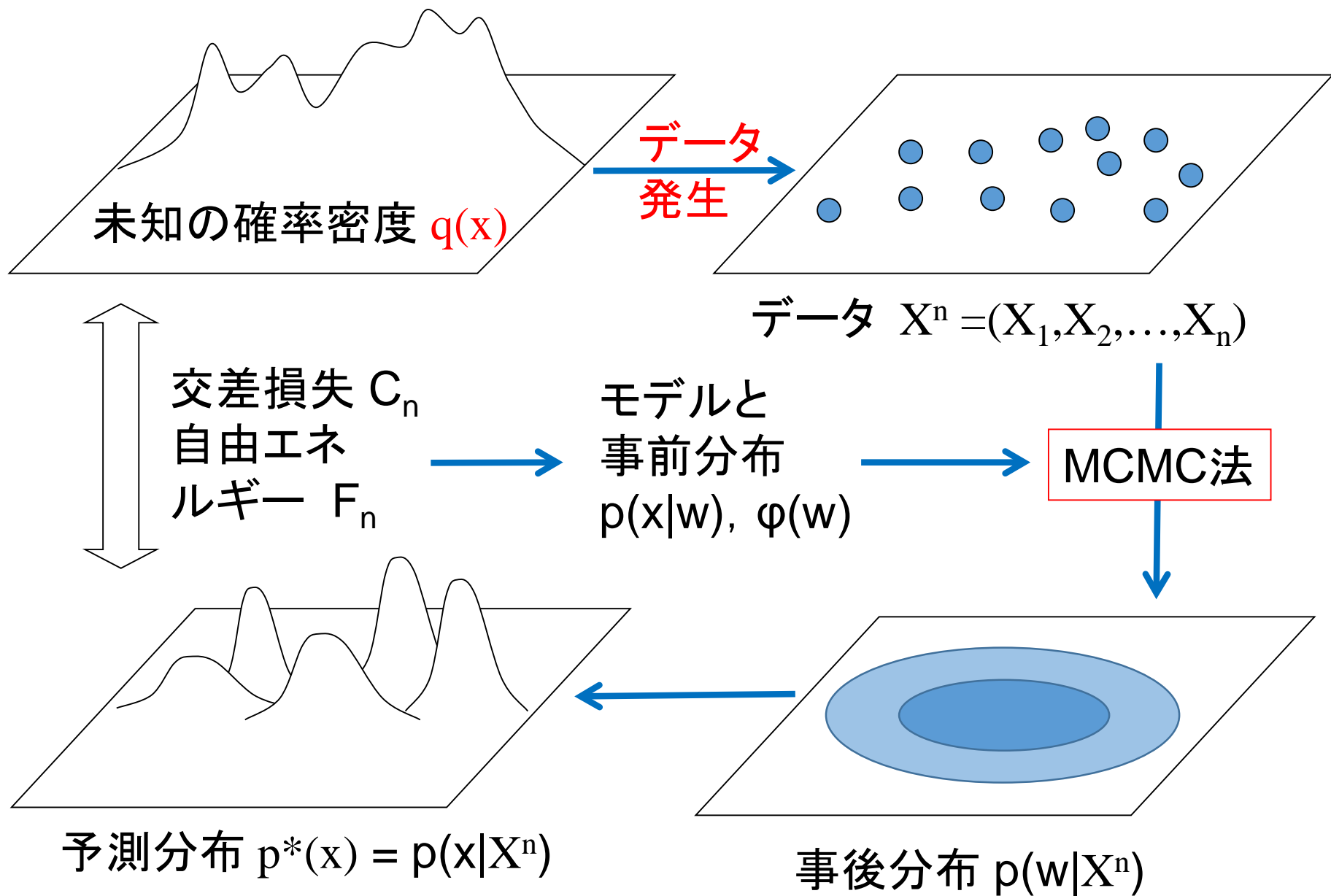


学習理論の練習 8

復習

統計的学習の物語



学習に関する量 まとめ

真のエントロピー $S = - \int q(x) \log q(x) dx$

事後分布 $p(w|X^n) = (1/Z) \varphi(w) \prod_i p(X_i|w)$

周辺尤度 $Z(X^n) = \int \varphi(w) \prod_i p(X_i|w) dw$

自由エネルギー $F_n = - \log Z(X^n)$

予測分布 $p^*(x) = E_w[p(x|w)] = p(x|X^n)$

汎化損失 $G_n = - E_x[\log p(X|X^n)]$

交差損失 $C_n = (1/n) \sum_{i=1}^n \log E_w[1/p(X_i|w)]$

恒等式

$$D(q(x) \parallel p(x|X^n)) = G_n - S$$

$$D(q(x^n) \parallel Z(x^n)) = E[F_n] - nS,$$

$$E[C_n] = E[G_{n-1}]$$

$$E[G_{n-1}] = E[F_n] - E[F_{n-1}].$$

学習理論とは

- (1) $E[G_n]$ は n のどんな関数か
- (2) G_n と C_n はどんな確率変数か
- (3) $E[F_n]$ は n のどんな関数か
- (4) F_n はどんな確率変数か

研究課題に向けて

もちろん研究の課題はたくさんあります。ここで紹介するものだけではありません。初めて研究を始めるときには何をどのくらい行ったら研究になるのかがわかりにくいと思いますので、例をあげて説明します。研究課題がみつからずに困るようでしたら問題を出すこともできます。

ここ数年ほど、ものすごく多くの人が機械学習の研究をするようになってきましたので、「ちょっとした思いつき」は類似研究が行われている可能性が高いです。オリジナリティの出し方については十分に検討する必要があると思います。

卒業研究・修士研究の題材について

(1) 学習理論の数学

学習理論に現れる数理を解明する。

⇒ 複素関数論、代数幾何、ゼータ関数、超関数、経験過程など

(2) 学習理論の統計学

学習理論に現れる統計学の問題を考えて、実データに応用する。

⇒ モデル選択、統計的検定、ハイパーパラメータ最適化、
交差検証、情報量規準、ブートストラップ、潜在変数...

(3) 学習理論の計算科学

学習理論に現れる計算上の課題を解決して実装可能にする。

⇒ マルコフ連鎖、平均場近似、相転移、ランジュバン方程式、
交差検証・周辺尤度の計算、双有理不変量の計算...

学習モデル いろいろ この他にもたくさん。

(1) かんたんなモデル

2項分布、ポアソン分布、正規分布、線形回帰、判別分析など実務的。
数学・統計学・計算科学はほぼできているが、課題も残されている。

(2) 階層型神経回路網

数学としてわかっている部分は多くはない。統計学としてまだ設計可能ではない。計算上の問題が多数ある。大流行。数理はほぼない。
研究すべき課題は多数あるが、簡単ではないかもしれない。

(3) 混合〇〇分布

数学としてある程度わかっている。統計学としてかなり設計可能になってきた。計算上の問題も解決されつつあり実務に使われている。
問題は豊富にあるが、既存研究も多数ある。

具体的な例を使って
研究分野の違いを説明する

簡単なニューラルネットの例

$w = \{(a_k, b_k) ; k=1, 2, \dots, K\}$ パラメータ

$Y = \sum_{k=1}^K a_k \tanh(b_k x) + N(0, 1^2)$ モデル

$$p(y|x, w) = 1/(2\pi)^{1/2} \exp[- (1/2) \{ y - \sum_{k=1}^K a_k \tanh(b_k x) \}^2]$$

$\varphi(w) =$ 十分大きなコンパクト集合上で非零の分布

$\{ (X_i, Y_i) ; i=1, 2, \dots, n \}$ データ

$q(x) q(y|x)$ 真の分布

研究例と答例

数学の問題。データを発生している分布についての仮定を設定し
汎化損失の漸近挙動を求めよ。

答例。真の分布が $q(x) p(y|x,0)$ のとき $E[G_n]=S+(2/3)(1/n)+o(1/n)$ である。

統計の問題。データが与えられたとき、汎化損失を最小にする K を求めよ。

答例。交差損失を各 K について計算し、最小となる K にすればよい。
もっと良い方法が考案できればもっと良い。

計算の問題。データが与えられたとき、予測分布と交差損失を計算する
アルゴリズムを作れ。

答例。ハミルトン法を実装して事後分布を数値的に実現すれば、交差
損失を計算できる。もっと良い方法が作れればもっとよい。

解説(1) 数学

答例。真の分布が $q(x) p(y|x,0)$ であるとき $E[G_n]=S+2/(3n)+o(1/n)$ である。

(真の分布・学習モデル・事前分布)が与えられているときに、汎化損失と自由エネルギーの漸近挙動を求めるための数学的基盤はある。しかし実際にその方法を適用するためには個々の課題について理論研究が必要である。

- 統計的正則モデル — パラメータ次元/2
- 縮小ランク回帰 — 完全に解決されている(青柳美輝)
- 混合正規分布 — 上界(山崎啓介)
- 非負値行列分解 — 上界(林直輝)
- 混合ポアソン分布 — 上下界(佐藤件一郎)
- 階層的神経回路網 — 上界はわかっている...

研究は鉛筆とノート。イデアル、ブローアップ、射影空間、超関数の漸近展開、経験過程など数学すべてに渡る知識と、どこまでも果てしない計算の砂漠をさすらいことになる。つらくて厳しい努力の果てに代数幾何や超関数が実問題の解決に力を発揮する瞬間に出会うというプライスレスな体験ができるが社会からは理解されにくい。万難を排して数学をやりぬく気持ちの人に限定。

解説(2) 統計学

答例。交差損失を各 K について計算し、最小となる K にすればよい。

複数の学習モデルの候補の中から最も適切なものを選ぶことを「統計的モデル選択」という。統計学では標準的な研究課題のひとつであり、目標として汎化損失の最小化だけでなく、モデル選択の一致性が設定されるときもある。方法としては交差損失だけでなく各種の情報量規準が検討される。また適用するモデルやデータの性質に依存する課題もある。

基本的な研究スタイルはアルゴリズムの着想をソフトウェアで実現したり実問題への適用の検討になる。もしも理論的な研究ができればすばらしいが、時間が限られている状況で新しいアイデアを思いつくことを必要とする目標はリスクであるため、安全な目標の設定は相談になります。

今日では「データサイエンティスト」や「機械学習エンジニア」という職業もあり、統計学を学んだ人は産業界からとても強く求められています。もしも統計学を人生の中心にしていきたいのであれば、研究テーマを進めるだけでなく、基礎からきちんと学びましょう。人工知能が作る未来で活躍できる仕事です。

解説(3) 計算科学

答え。ハミルトン法を実装して事後分布を数値的に実現すれば、交差損失を計算できる。

学習理論で現れる概念には、膨大な演算量を必要とするためコンピュータで算出することが容易でないものが多い。事後分布の実現や自由エネルギーの計算が基本的な例である。計算ができないと学習が実現できず、自由エネルギーも不明なままなので、実務における重要性は社会から高く評価されている。MCMC召喚士(MCMCを自由自在に操れる才能に恵まれた人)は機械学習で実世界に挑戦するパーティには必ず1人は必要になります。

研究のスタイルとしては物理系のシミュレーションと似ていて、ハミルトン関数が与えられたときの平衡状態を実現することと数理的には等価である。ただし極めて多くの研究があるので、オリジナリティの出し方については、よく考える必要がある。安全な目標の設定にあたっては相談になります。

MCMCは科学技術計算や人文社会科学でも必要になりますので、会社や独立法人などで研究開発の仕事をしたい人に推奨します。

具体的な例を使って
課題を説明する 続き

数学の例：学習曲線と実対数閾値

汎化損失

$$E[G_n] = S + \lambda/n + o(1/n)$$

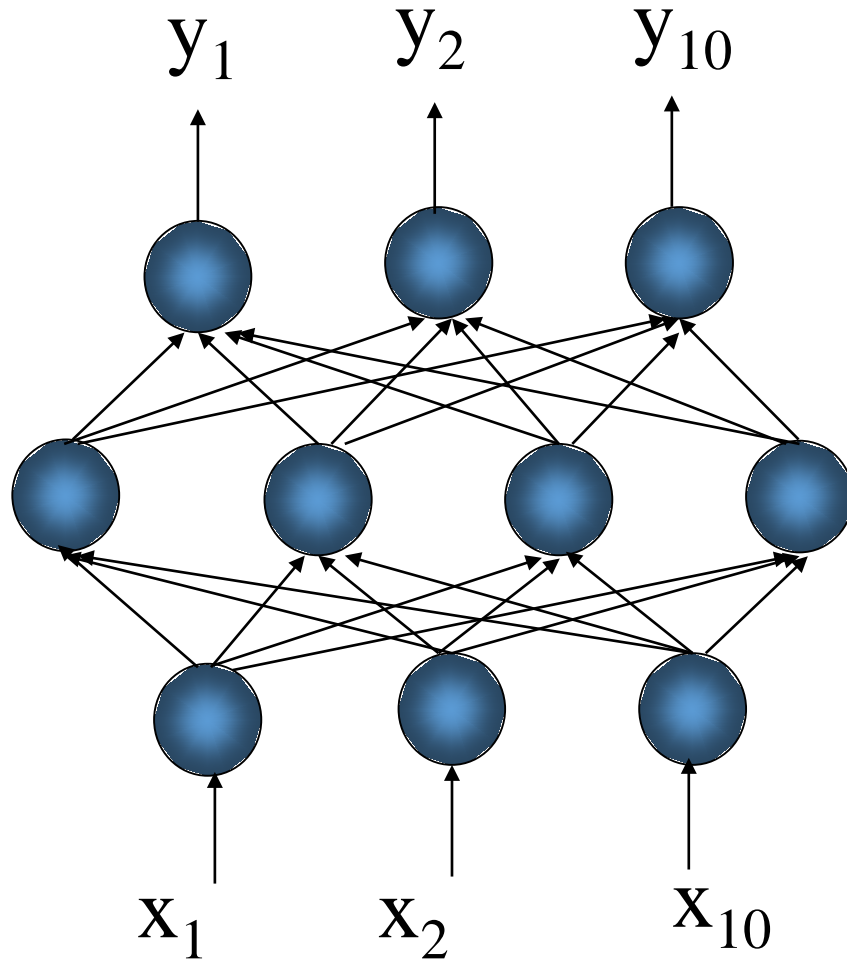
$$E[F_n] = nS + \lambda \log n + o(\log n)$$

S

$$S = - \int q(x) q(y|x) \log q(y|x) dx dy$$

n

統計学の例：モデル選択



真:

$$10 \rightarrow 5 \rightarrow 10$$

候補のモデル:

$$10 \rightarrow (1, 3, 5, 7, 9) \rightarrow 10$$

学習データ $n=200$

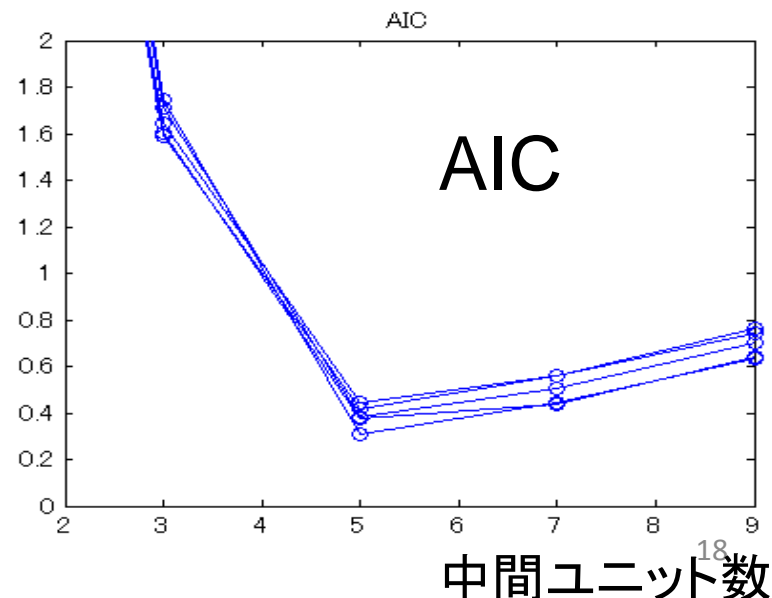
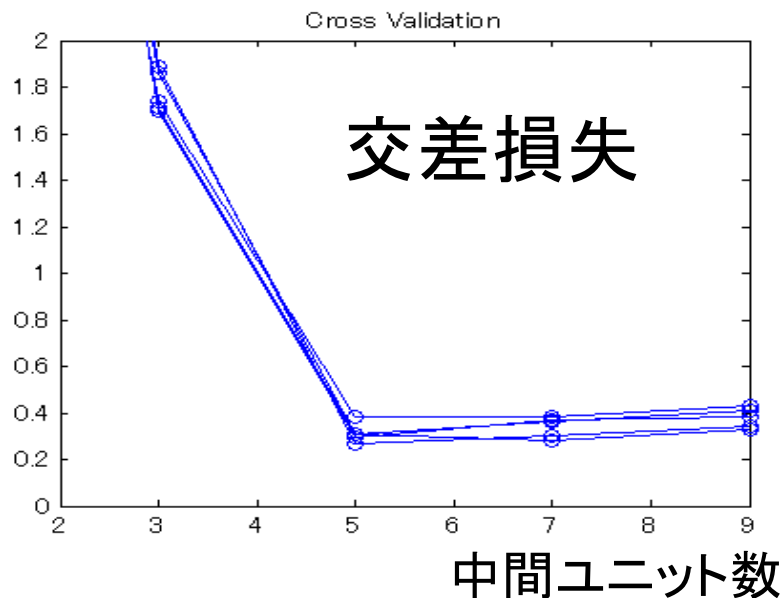
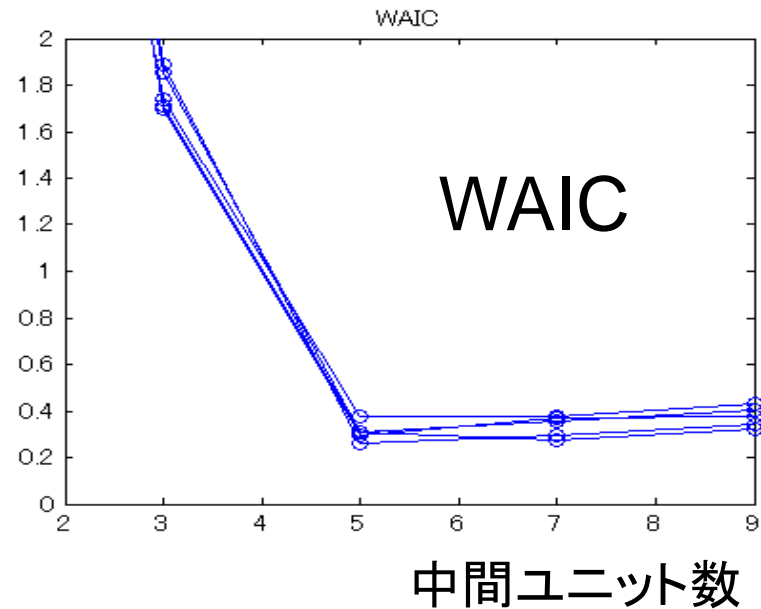
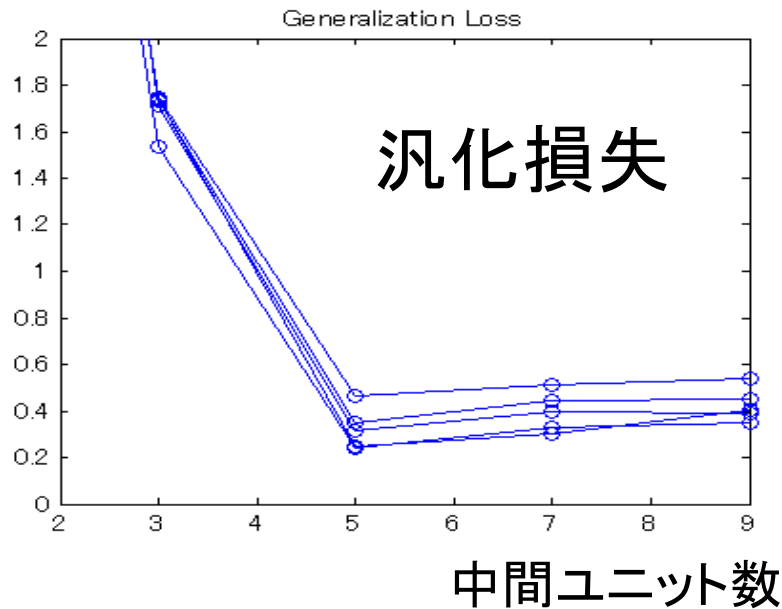
テストデータ =1000

事後分布は

ランジュバン方程式

(雑音あり最急降下)

10組の独立な学習データセットでの結果



マルコフ連鎖を作る例

真: $x=(x_1, x_2)$

$$g(x) = \exp(-x_1^2 - x_2^2 - x_1 x_2)$$

$$q(y|x) = g(x)^y (1-g(x))^{1-y}$$



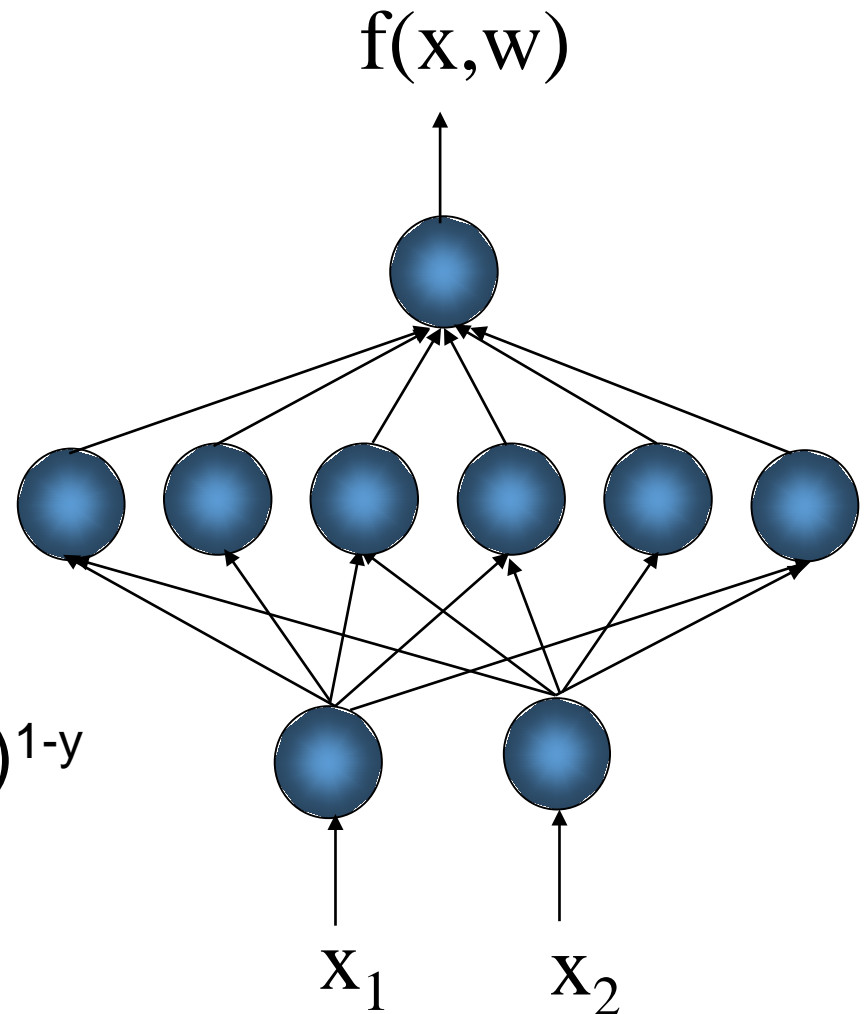
モデル:

$f(x, w)$: Neural Network

$$p(y|x, w) = f(x, w)^y (1-f(x, w))^{1-y}$$

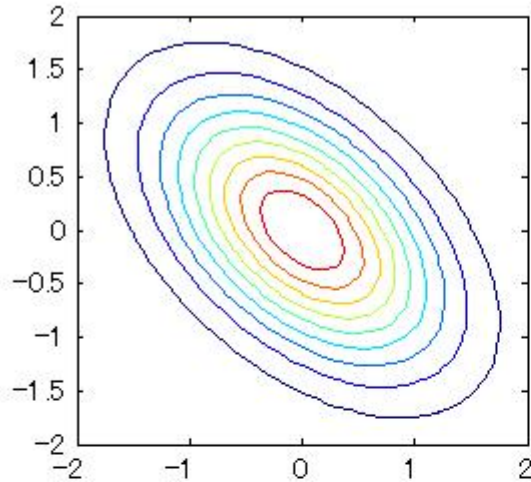


事後分布 $\propto \varphi(w) \prod p(Y_i|X_i, w)$

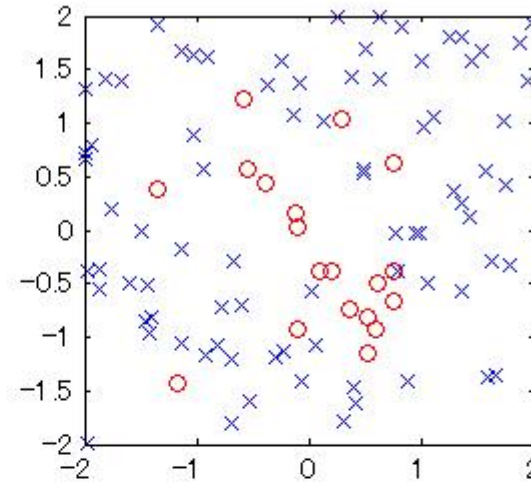


計算科学の例：MCMC法で事後分布を数値的に作って予測分布を求める

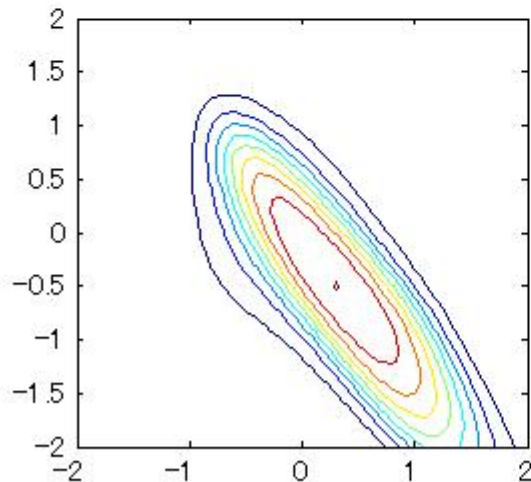
真の
条件
つき
確率



データ



予測
分布



MCMC法で
パラメータの
事後分布を作成