物理学における観測と機械学習
- 中性子星の事例 -

福嶋 健二

東京大学大学院理学系研究科物理学専攻

東京大学 知の物理学研究センター (iπ)

今日ここで議論したいこと

脳の構造を変えずに、教え方でもっと頭を良くするにはどうしたらいいか？

(教師あり学習の話)

今日ここで議論したいこと

脳の構造を変えずに、教え方でもっと頭を良くするにはどうしたらいいか？

NN構造をどうデザインするか？
背景にある数理的な原理？定式化？
よりシンプルな(少レイヤーの)DNN

今日ここで議論したいこと

脳の構造を変えずに、教え方でもっと頭を良くするにはどうしたらいいか？

NNのアーキテクチャーを変更せずに訓練データの加工によって学習性能を向上させることはできるか？

今日ここで議論したいこと

NNのアーキテクチャーを変更せずに訓練データの加工によって学習性能を向上させることはできるか？

~~訓練データの最適化はたいへん・・・~~

既存データの生成とプロセス・コストを変えずにできないか？

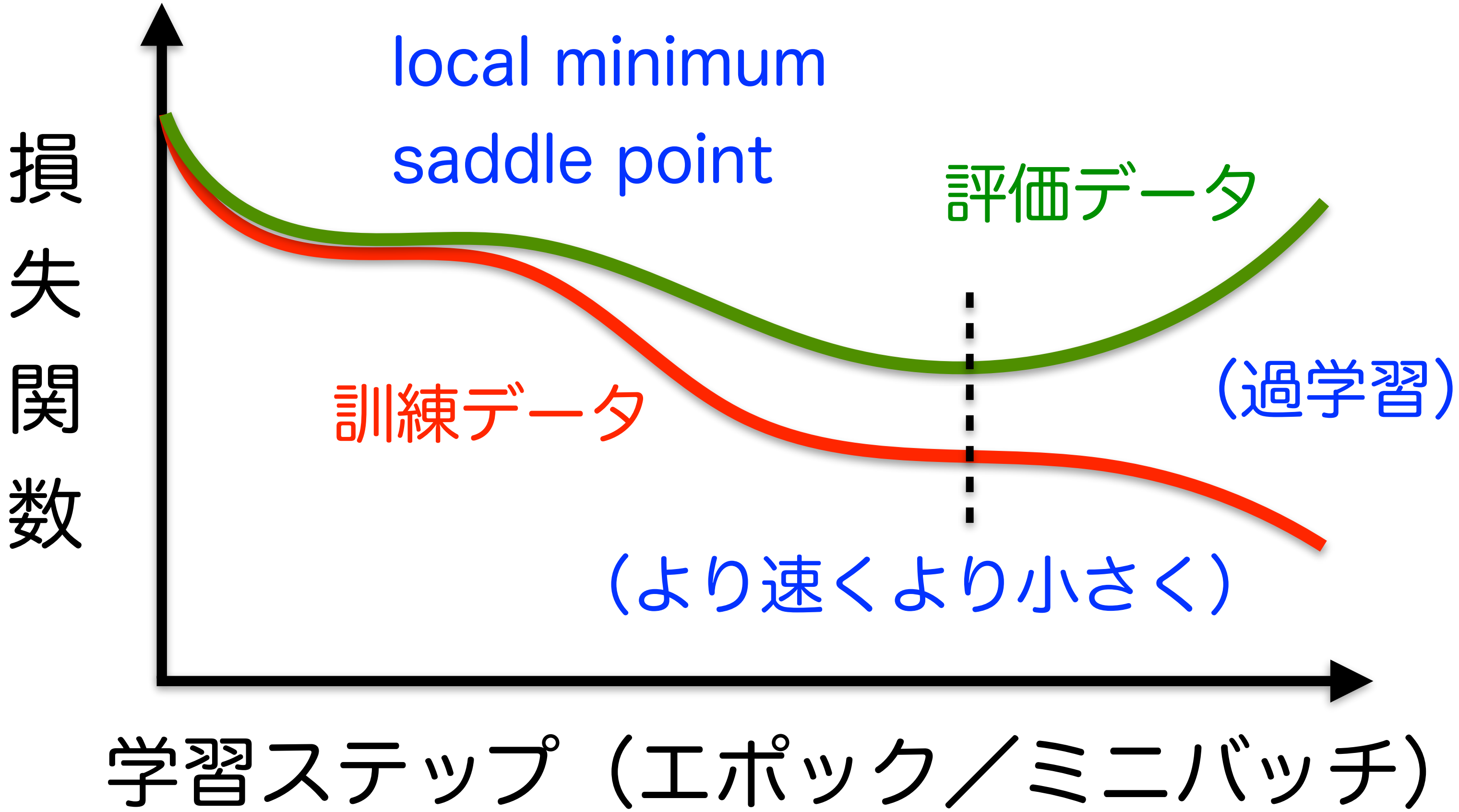
今日ここで議論したいこと

NNのアーキテクチャーを変更せずに
訓練データの加工によって学習性能を
向上させたい

より速く損失関数が小さくなる
損失関数がより小さくなる (local minimum)
過学習をより微調整なく回避できる

今日ここで議論したいこと

（概念図）



今日ここで議論したいこと

（傾向：訓練データを固定したとき）

NNが単純すぎるとlocal minimum
やsaddle pointにつかまりやすい

NNが複雑すぎると表現能力が
高すぎて過学習しやすい

今日ここで議論したいこと

（傾向：訓練データを固定したとき）

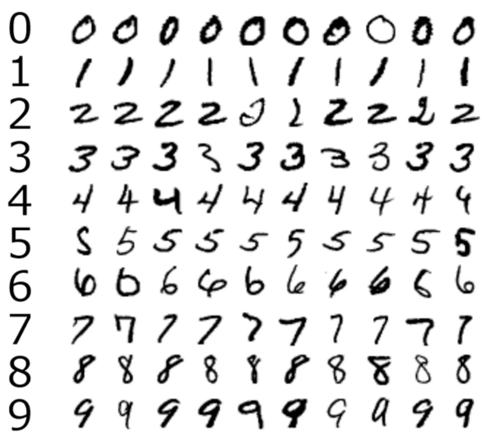
NNを複雑化するのが手っ取り早い・・・
訓練データ生成は計算コストが大きい
（NNを計算コスト削減に使いたかったりする）

NNが複雑すぎると表現能力が高すぎて過学習しやすい

今日ここで議論したいこと

どうしてこんなことを考えるのか

物理でないデータとの違い



(MNIST)

well-definedな正解はもともと存在していない

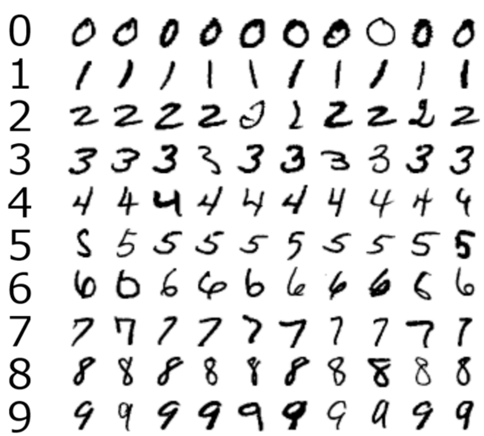
Aさんの6はBさんの0かも？

(cf. 教師なし学習)

今日ここで議論したいこと

どうしてこんなことを考えるのか

物理でないデータとの違い



(MNIST)

人工的に「正解」から程よく揺らいだ訓練データを生成できるか？「正解」の定義や揺らがせ方に強い依存性

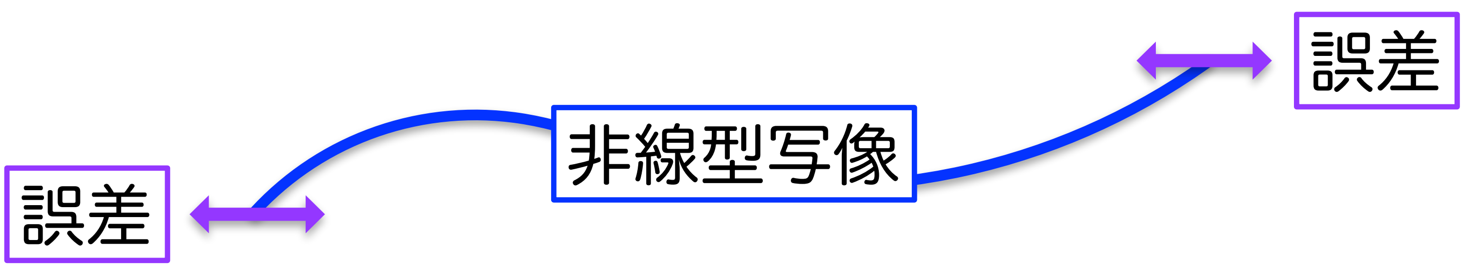
今日ここで議論したいこと

どうしてこんなことを考えるのか

物理で扱うデータの特性

ある（原理的な）正解の周りで精度保証
されて揺らいだ観測データを扱う

中性子星の事例をすぐ後で紹介します



今日ここで議論したいこと

どうしてこんなことを考えるのか

物理でのNN応用の現状

非線型写像　（時短）

非線型写像　（分類）（診断）

今日ここで議論したいこと

どうしてこんなことを考えるのか

物理でのNN応用の現状

物理で扱うデータの特性をまだ十分に深堀りできていないのではないか？

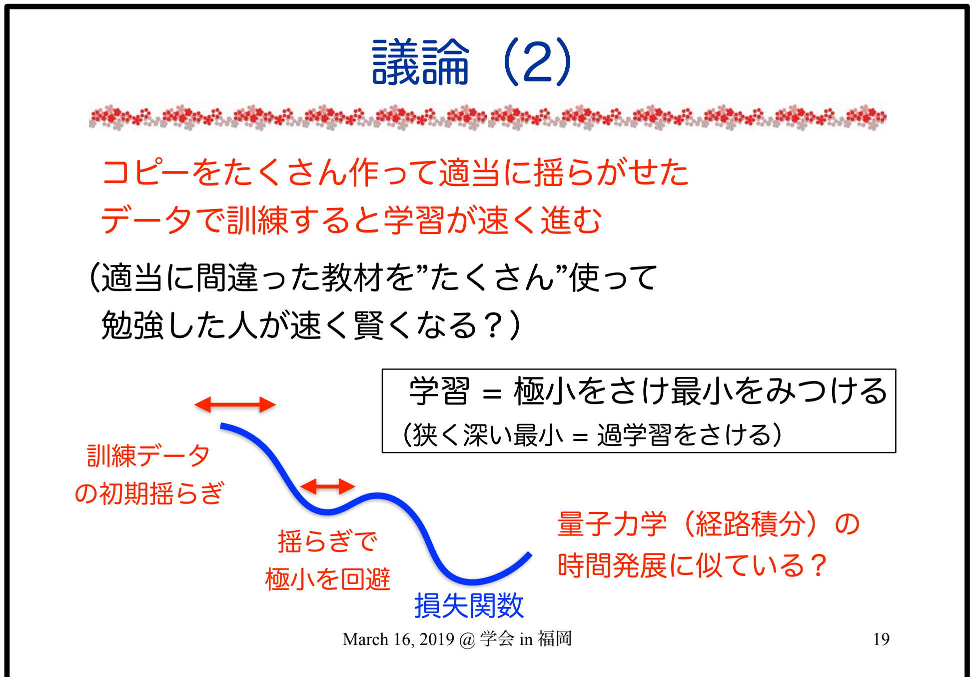
（時短）

（分類）

（診断）

今日ここで議論したいこと

2019年3月の学会にて（機械学習シンポ）



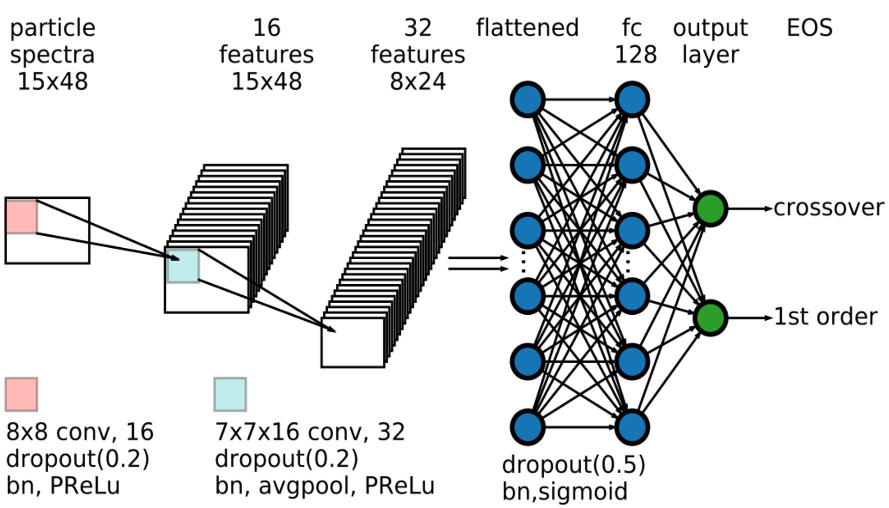
1年前に学会で述べた
speculationについて
数値実験してみた結果
を専門家である
聴衆の皆さんと
議論したい

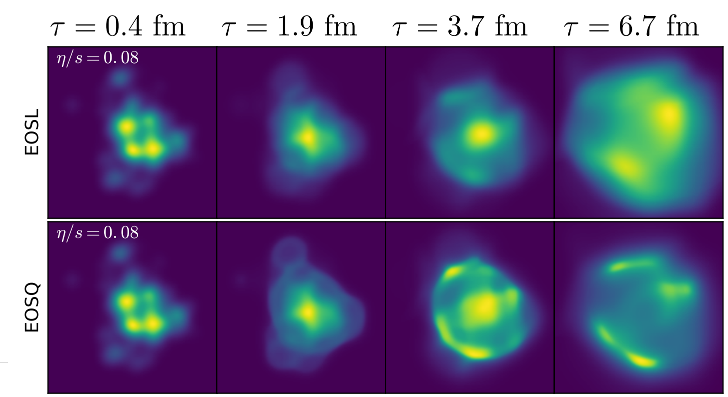
高エネルギー物理での応用事例

【クォークグルーオンプラズマ現象論】

L.-G. Pangのスライドより

相転移の分類





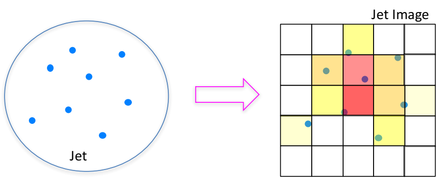
$$l(\theta) = \underbrace{-\frac{1}{N}\sum_{i=1}^{N}[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)]}_{\text{cross entropy loss}} + \underbrace{\lambda||\theta||_2^2}_{\text{L2 regularization}}$$

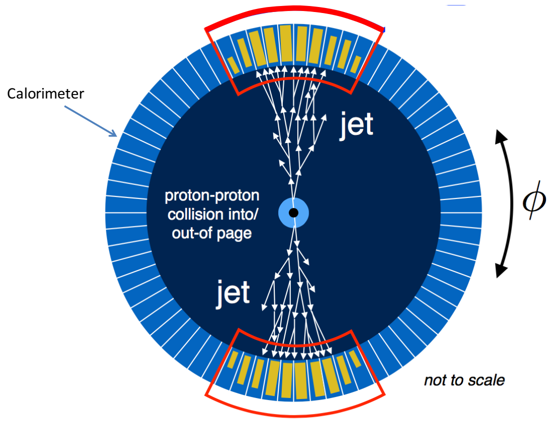
高エネルギー物理での応用事例

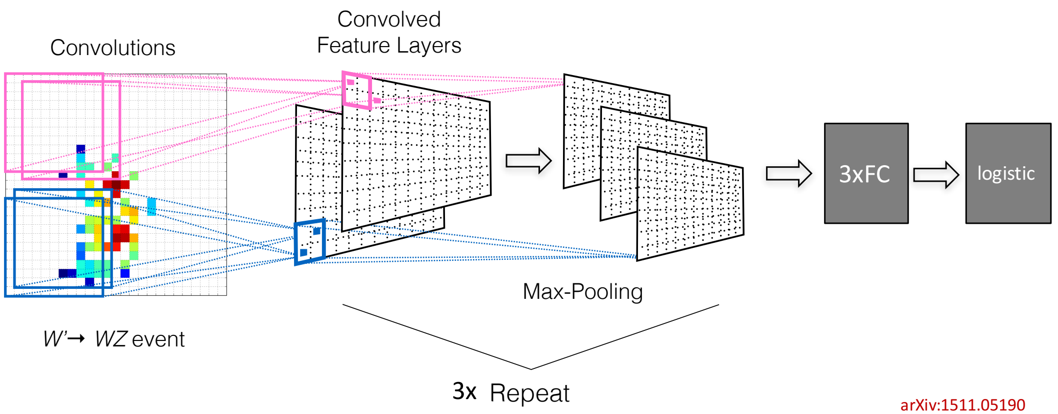
【ジェットの識別】

内部構造の診断

M. Kaganのスライドより抜粋







標準的な方法として市民権を獲得している

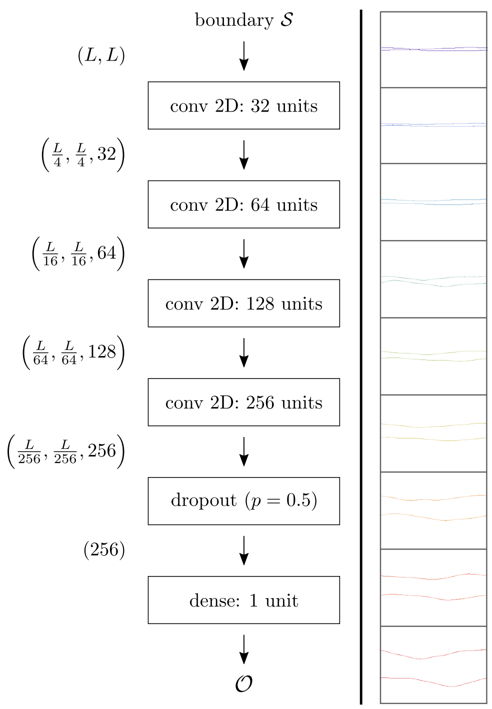
CNNの他、ジェットのシークエンス解析にRNN

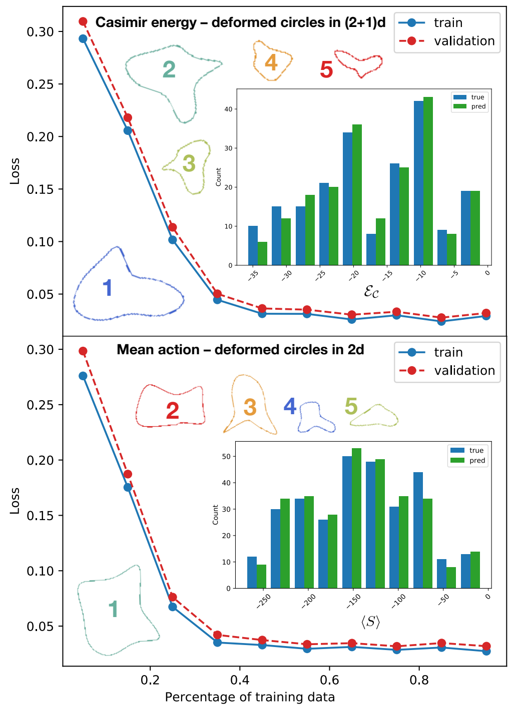
高エネルギー物理での応用事例

【場の量子論の数値シミュレーション】

M. Chernodubの最近の仕事より

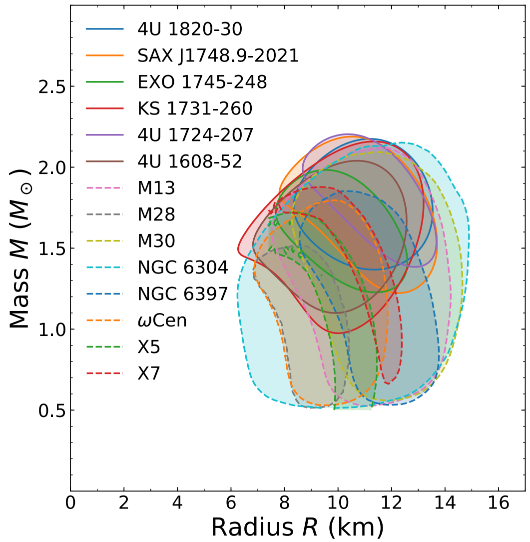
計算高速化

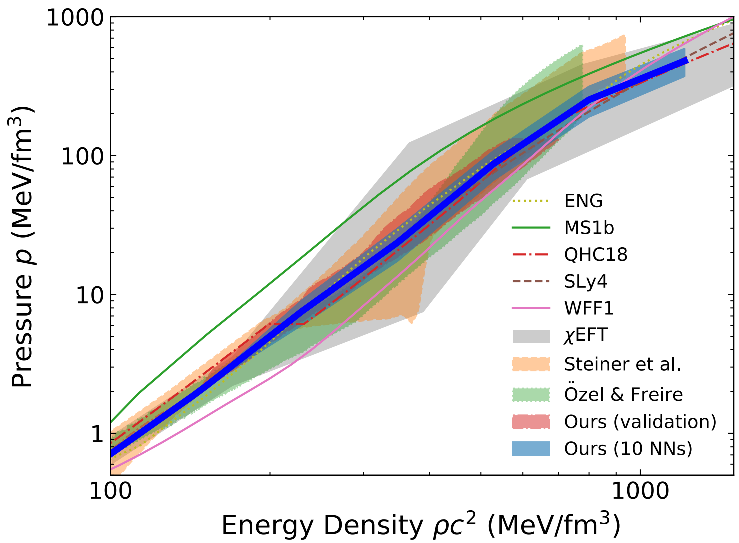




高エネルギー物理での応用事例

【中性子星の状態方程式(EoS)構築】



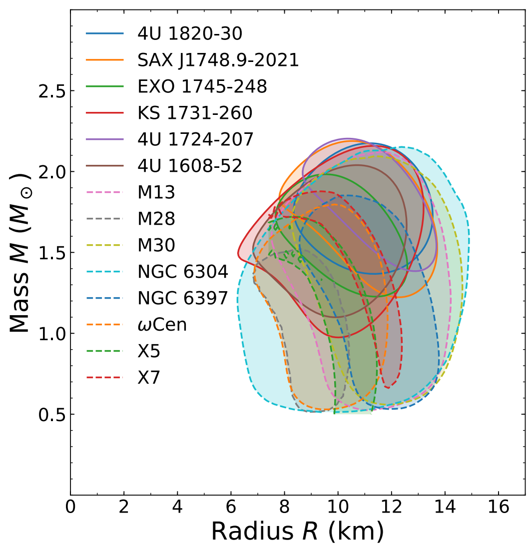


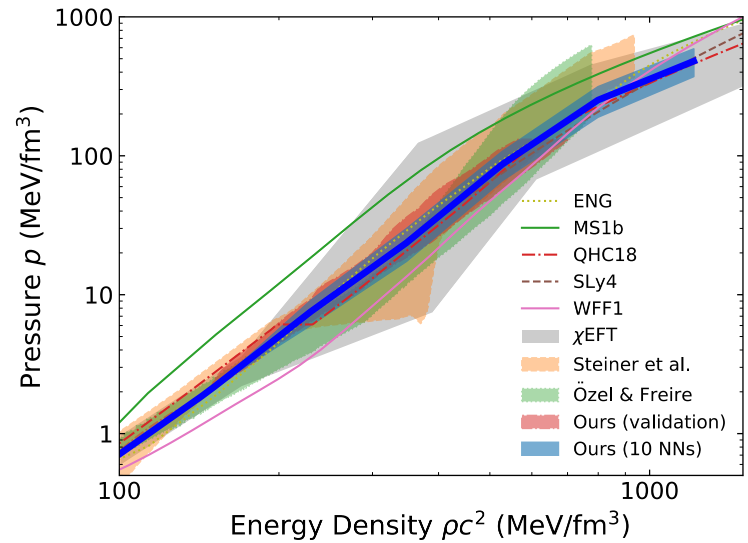
Fujimoto-Fukushima-Murase: 1711.06748 [nucl-th]

Fujimoto-Fukushima-Murase: 1903.03400 [nucl-th]

高エネルギー物理での応用事例

【中性子星の状態方程式(EoS)構築】

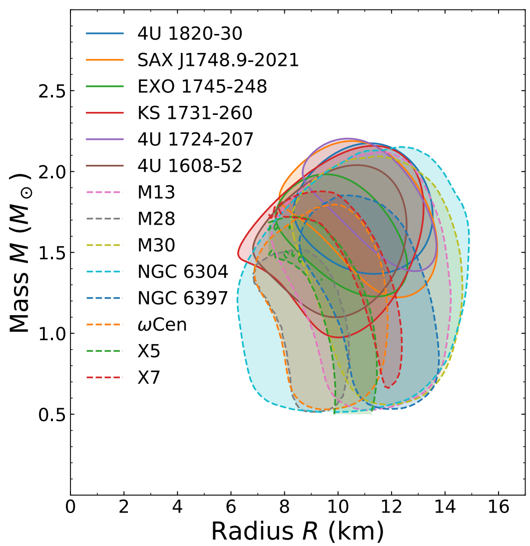


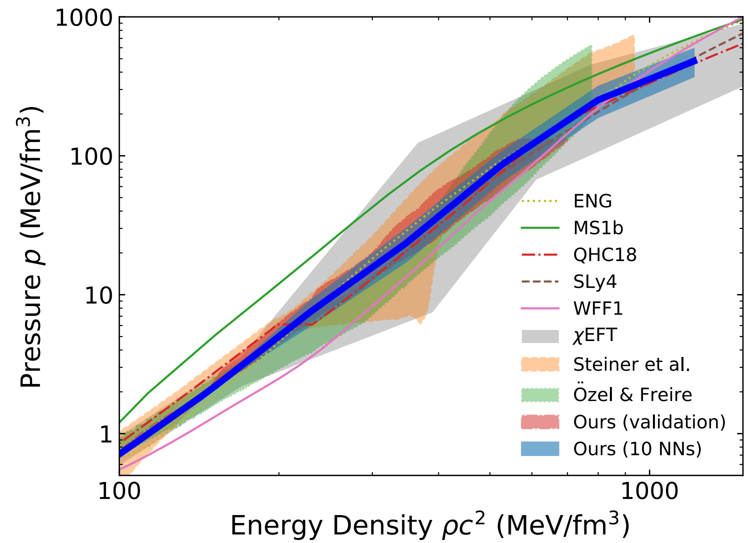


誤差のついた観測データから誤差のついた結果を推定したい（多くの物理に共通の問題）

高エネルギー物理での応用事例

【中性子星の状態方程式(EoS)構築】



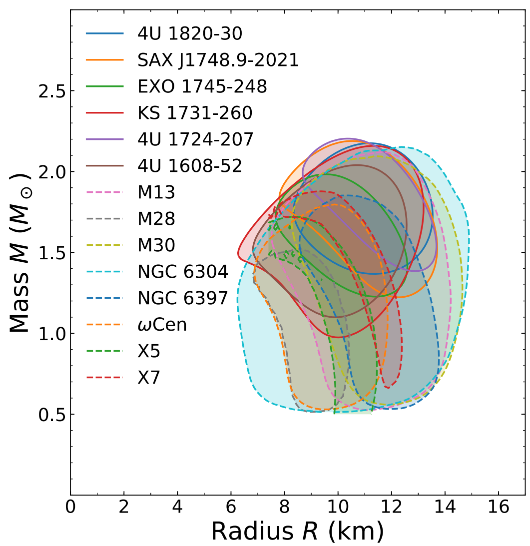


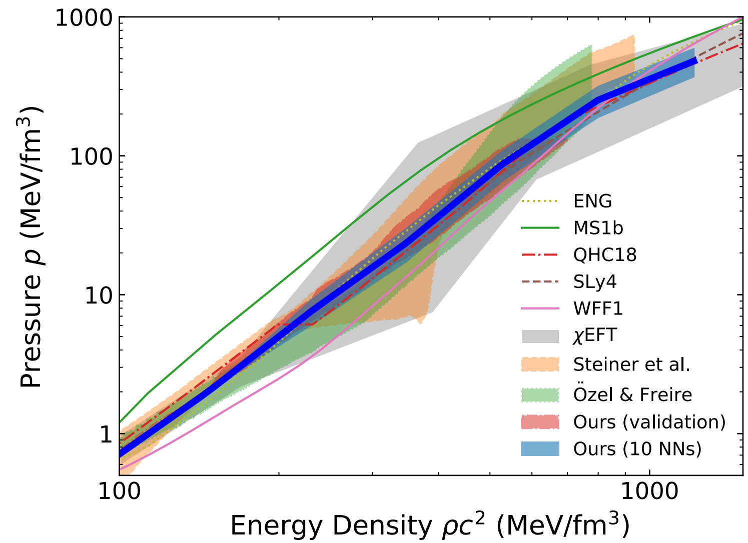
誤差のついた観測データから誤差のついた結果

↑入力の不確かさ　　↑出力の確かさ

高エネルギー物理での応用事例

【中性子星の状態方程式(EoS)構築】



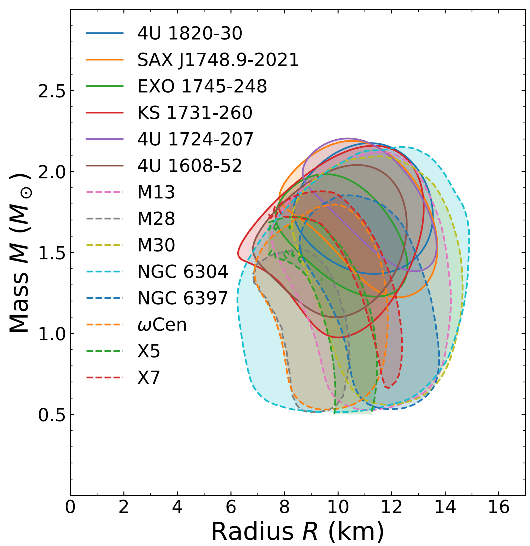


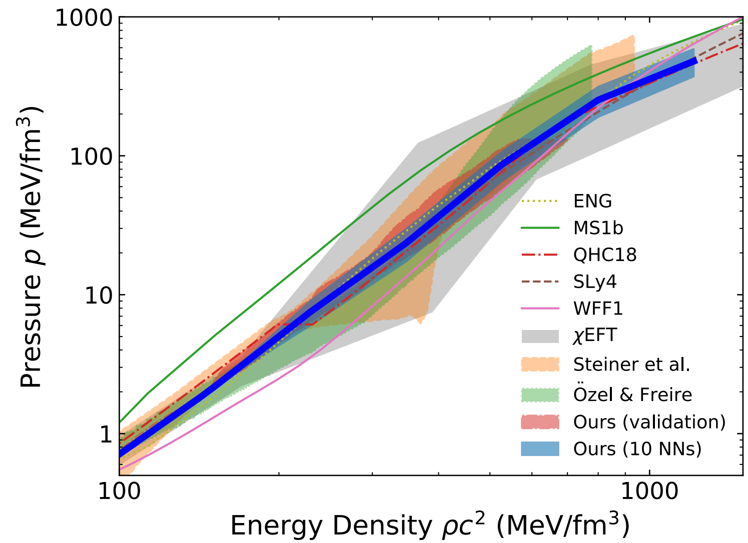
ひとつのアプローチ：確率分布をパラメトライズ

cf. Bayesian解析 / （制限）ボルツマンマシン

高エネルギー物理での応用事例

【中性子星の状態方程式(EoS)構築】



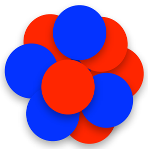


確率分布：Fokker-Planck的

→ 我々のアプローチ：Langevin的 (数値計算に有利)

中性子星の物理

【原子核＝陽子・中性子の多体系】



Fe~Niが最も安定

重い原子核は核分裂

中心部の密度ほぼ一定（飽和）

標準核密度 $\rho_0 \simeq 2.7 \times 10^{17}\ \mathrm{kg/m^3}$

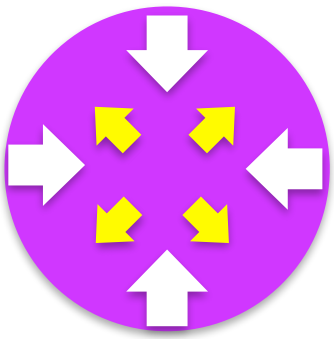
【中性子星＝超巨大原子核】

重力で束縛

静水圧条件で構造決定

中心部の密度

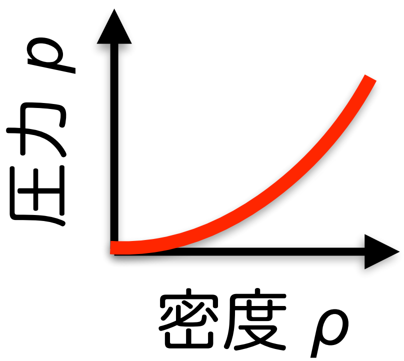
$\rho \gtrsim 5\rho_0$



重力と**圧力**の釣合い

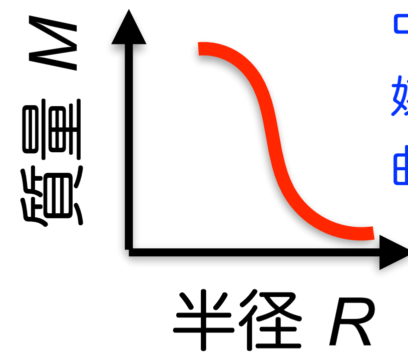
中性子星の物理

状態方程式



1 対 1 対応

M-R関係

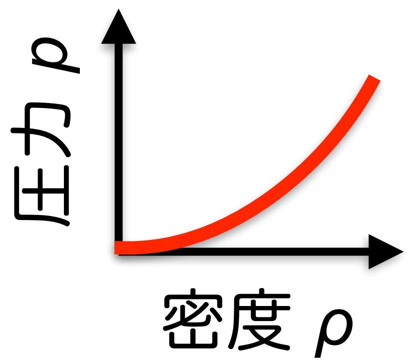


中心部の密度を
媒介変数にした
曲線を描く

中性子星の物理の専門家になると計算しなくても M-R関係を見ただけで、元の状態方程式の性質を推定できるようになる（つまり経験的に学習する）

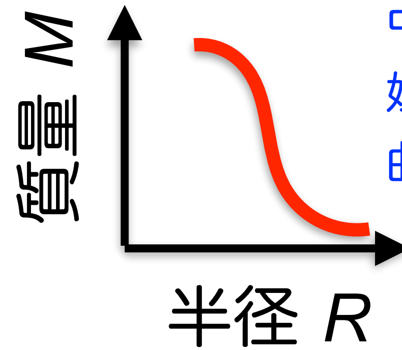
中性子星の物理

状態方程式



1 対 1 対応

M-R関係

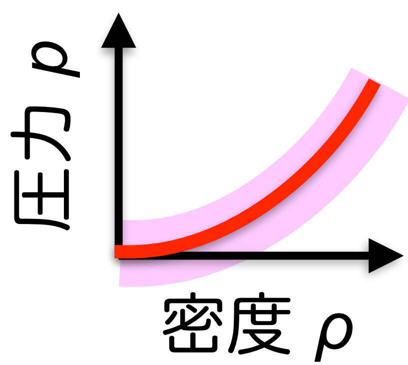


中心部の密度を
媒介変数にした
曲線を描く

「学習」を専門家でなく機械にさせたい

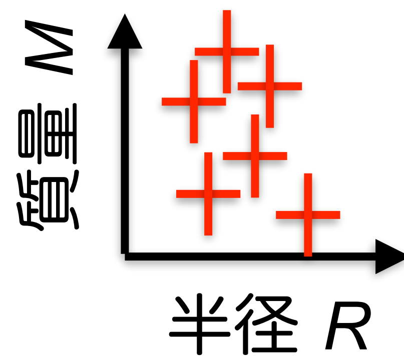
中性子星の物理

状態方程式



~~1対1対応~~

M-R関係



観測データは
有限個＋誤差

尤もらしい答え ⟵ 限られた観測データ

Bayesian解析が威力を発揮

Bayesian解析について

ベイズの定理

規格化

$$\underbrace{P(A|B)}_{\text{観測のもとでのEoSの確率分布}}\cancel{P(B)} = \underbrace{P(B|A)}_{\text{Likelihood}}\underbrace{P(A)}_{\text{prior モデル}}$$

観測のもとでの
EoSの確率分布

Likelihood

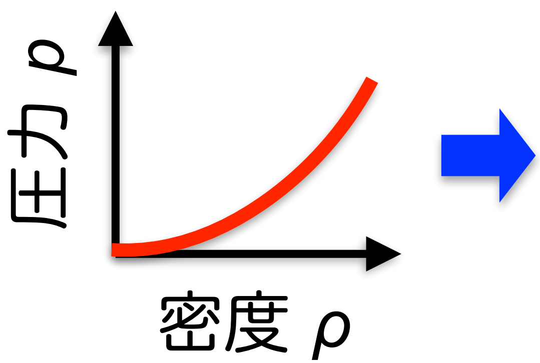
prior
モデル

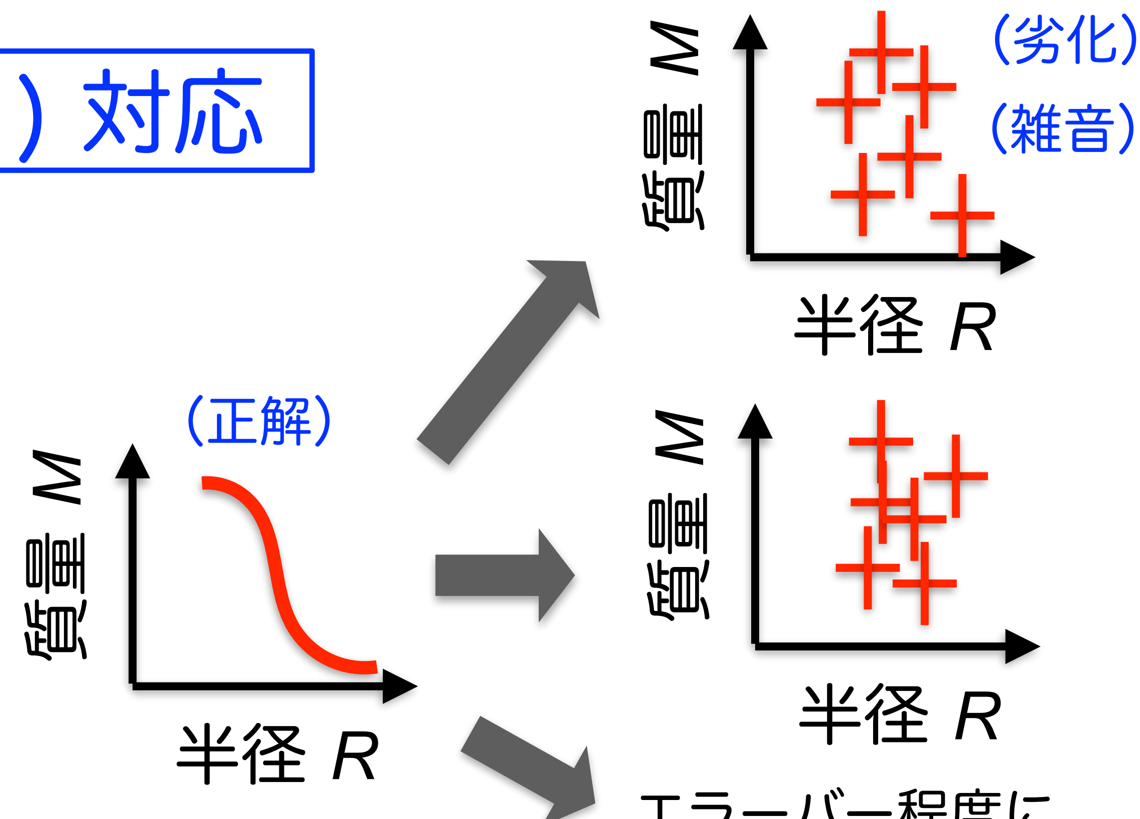
観測数(中性子星の数)が増えればLikelihoodが中心極限定理で鋭い関数になりprior依存性は消える・・・が
現状ではprior依存性はかなり大きい

我々(藤本・福嶋・村瀬）のアイデア

$\boxed{1\ 対\ n_s(>1)\ 対応}$

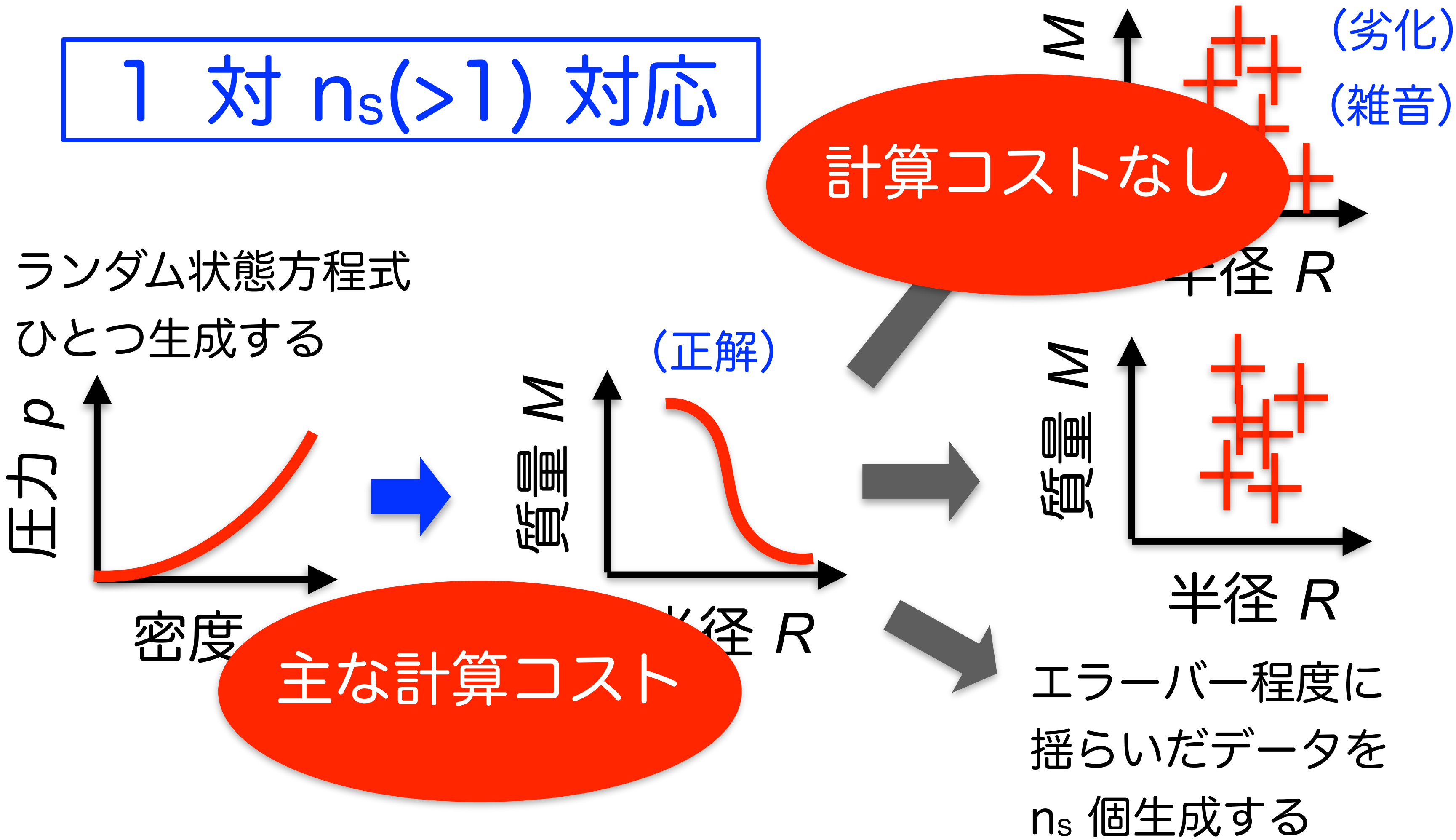
ランダム状態方程式
ひとつ生成する





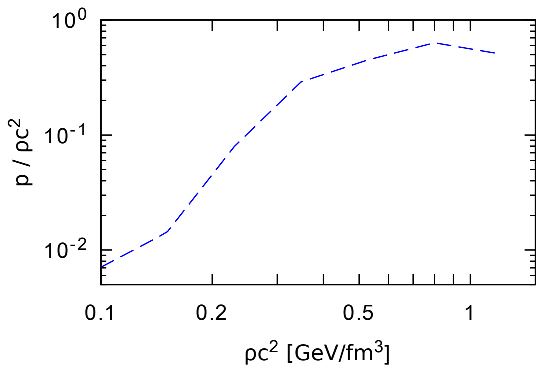
エラーバー程度に
揺らいだデータを
n_s 個生成する

我々(藤本・福嶋・村瀬）のアイデア



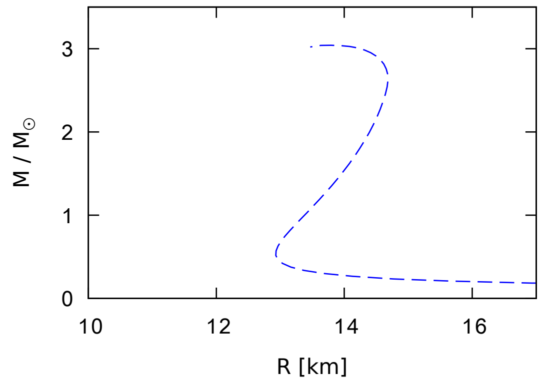
我々(藤本・福嶋・村瀬）のアイデア

正解の状態方程式



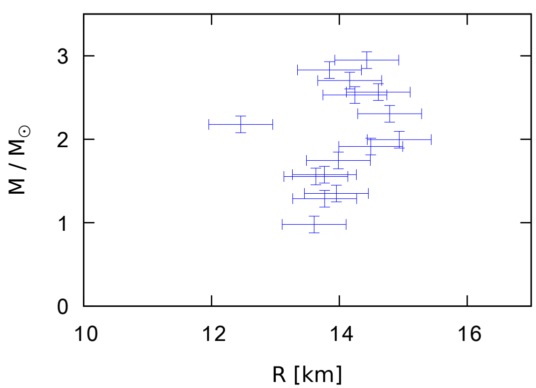
を

対応する*M-R*関係



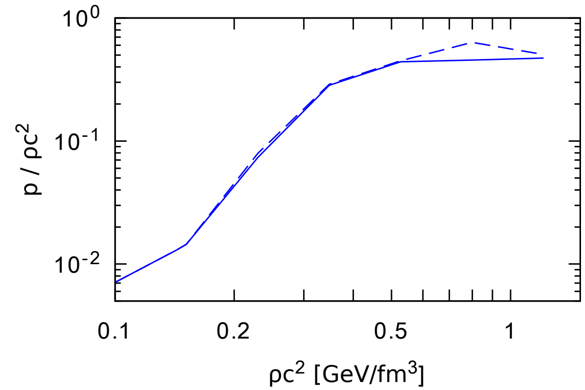
ではなくて

その周りでサンプリングしたデータ



から推定

評価データによる検証



Fujimoto-Fukushima-Murase (2018)

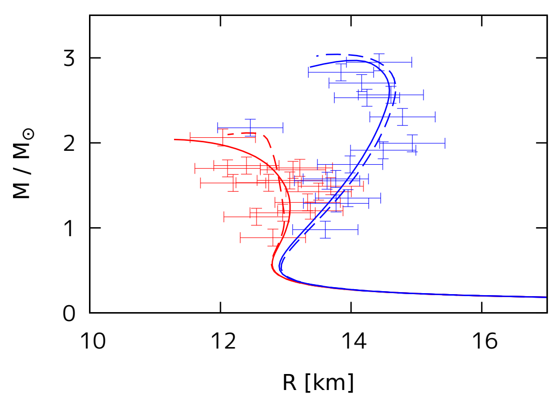
我々(藤本・福嶋・村瀬）のアイデア

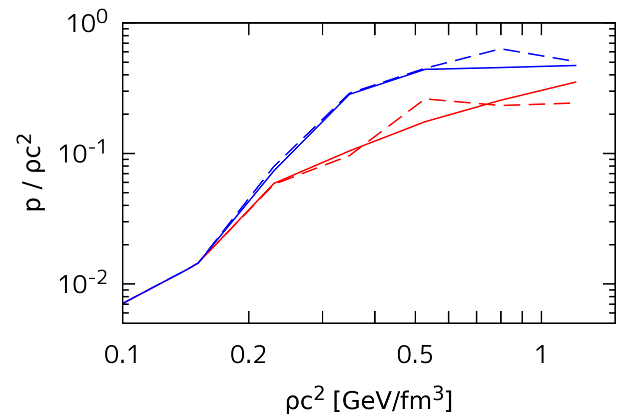
(計算の詳細：正解の分かっている検証実験)

Layer index	Nodes	Activation
0	30	N/A
1	60	ReLU
2	40	ReLU
3	40	ReLU
4	5	tanh

訓練データ

200 x 100 (n_s)





疑似発生した観測データ → EoS推定 → MRに戻す(破線)

→ 正解と比較する(実線)

Mass ($M_\odot$)	0.6	0.8	1.0	1.2	1.4	1.6	1.8
RMS (km)	0.16	0.12	0.10	0.099	0.11	0.11	0.12

簡単に性能を出せたので驚いた・・・

我々(藤本・福嶋・村瀬）のアイデア

Bayesian解析との比較

事前確率 (prior) 依存性はどこに？

EoSを特徴付けるパラメータのランダムな選び方

観測された中性子星の数が極端に少なかったら
学習の効果が少なく出力されるEoSは初期値の
分布の影響 (prior依存性) を強く受ける

損失関数の選び方も不定性の起原

困難保存の法則

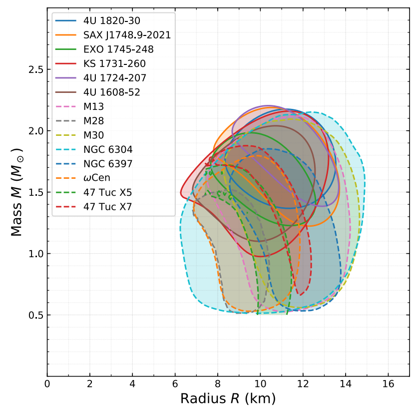
我々(藤本・福嶋・村瀬）のアイデア

（計算の詳細：実際の観測データを用いた解析）

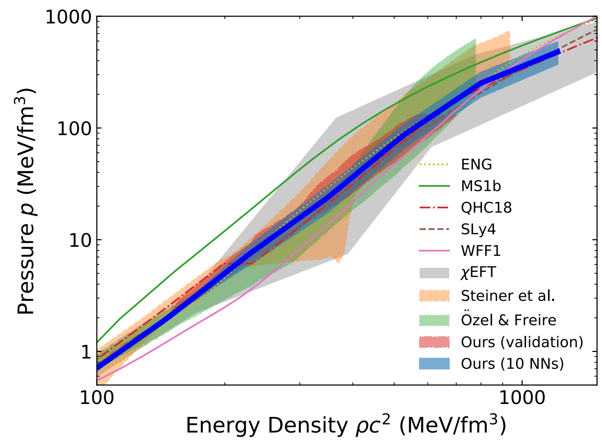
Layer	Number of neurons	Activation function
0 (Input)	56	N/A
1	60	ReLU
2, 3	40	ReLU
4 (Output)	5	tanh

訓練データ

500 x 100 x 100 (n_s)



14個の中性子星の
観測の1σ領域



入力(観測)の誤差は訓練データのノイズコピー

出力(推定)の誤差はどうやって評価するか？

我々(藤本・福嶋・村瀬）のアイデア

出力(推定)の誤差はどうやって評価するか？

専門家に質問 → それは今後研究したい
でも我々は今すぐ答えが欲しい・・・

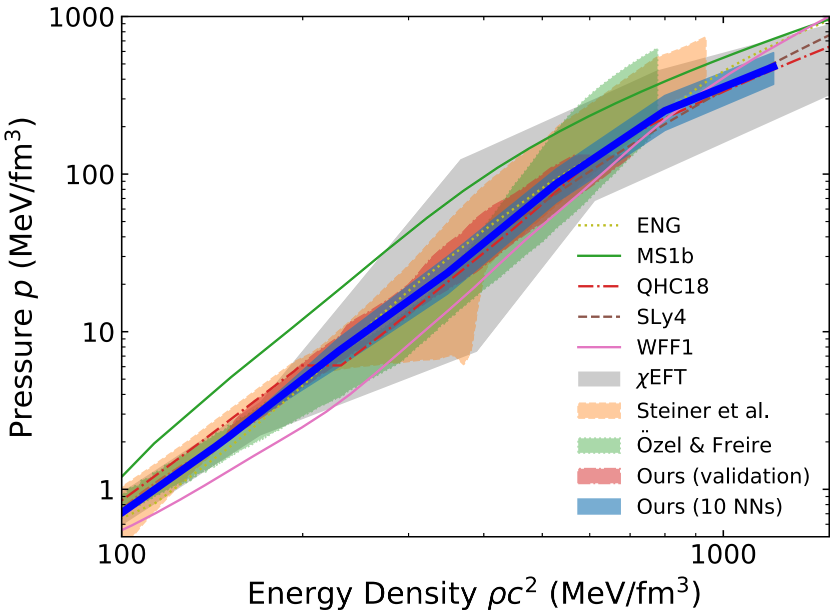
ひとつの知られている方法：Bootstrap法

異なる訓練データセットから独立な予測モデルを複数構成して出力の揺らぎから信頼幅を推定する

※ Bootstrap法は重複を許してランダムに再標本化するサンプリング方法のこと (信頼幅推定は用途)

我々(藤本・福嶋・村瀬）のアイデア

出力(推定)の誤差はどうやって評価するか？



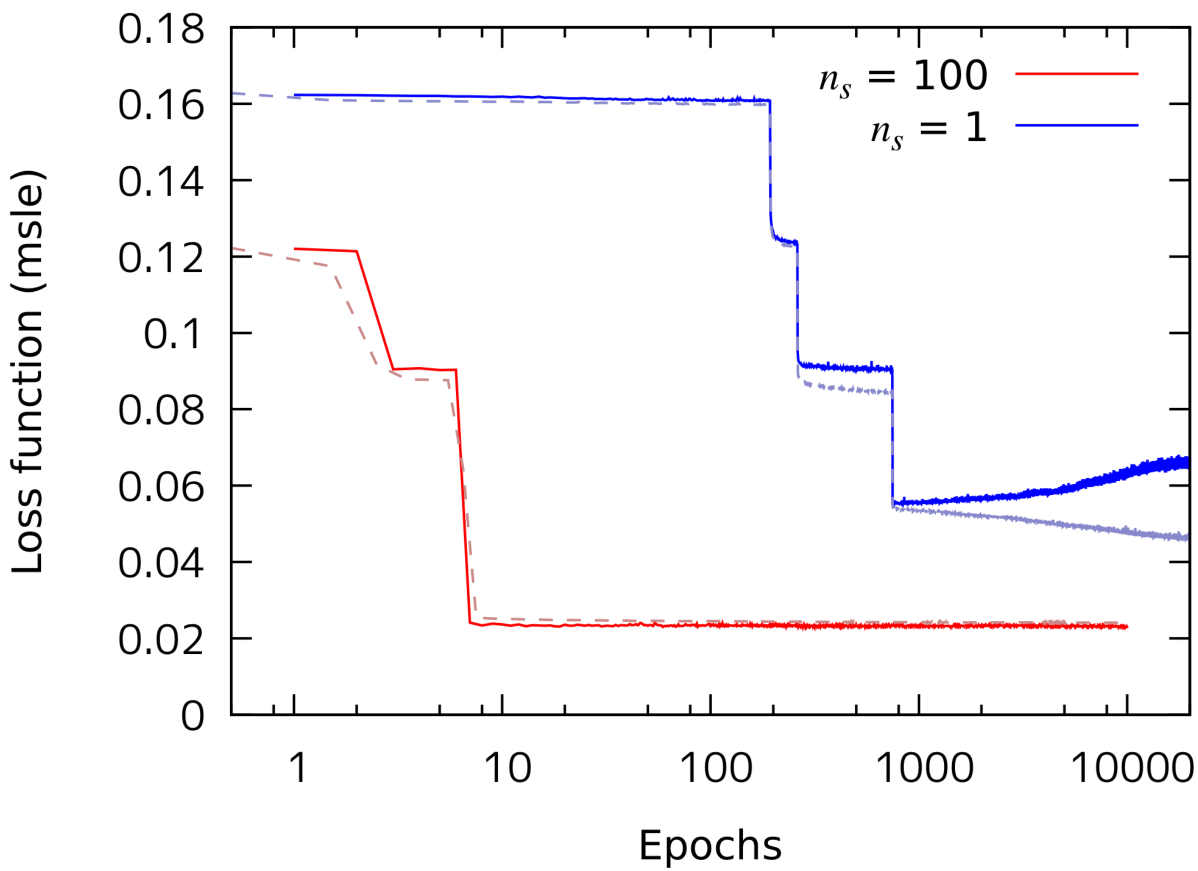
Bootstrapで得られる信頼幅はひとつの枠組みの中での出力の再現性の確からしさを表す

constraintが弱ければ乱数的に大きく揺らぐはずなので指標として悪くない・・・気がする

物理では統計誤差と系統誤差がある
系統誤差は(かなり)過小評価している可能性あり

今日ここで議論したいこと

Fujimoto-Fukushima-Murase (2018)



過学習

過学習を回避

今日ここで議論したいこと

【簡単な用語の復習】

学習：損失関数を勾配に沿って落としていく

勾配計算方法

バッチ学習：訓練データを全部使って計算
（時間がかかる）

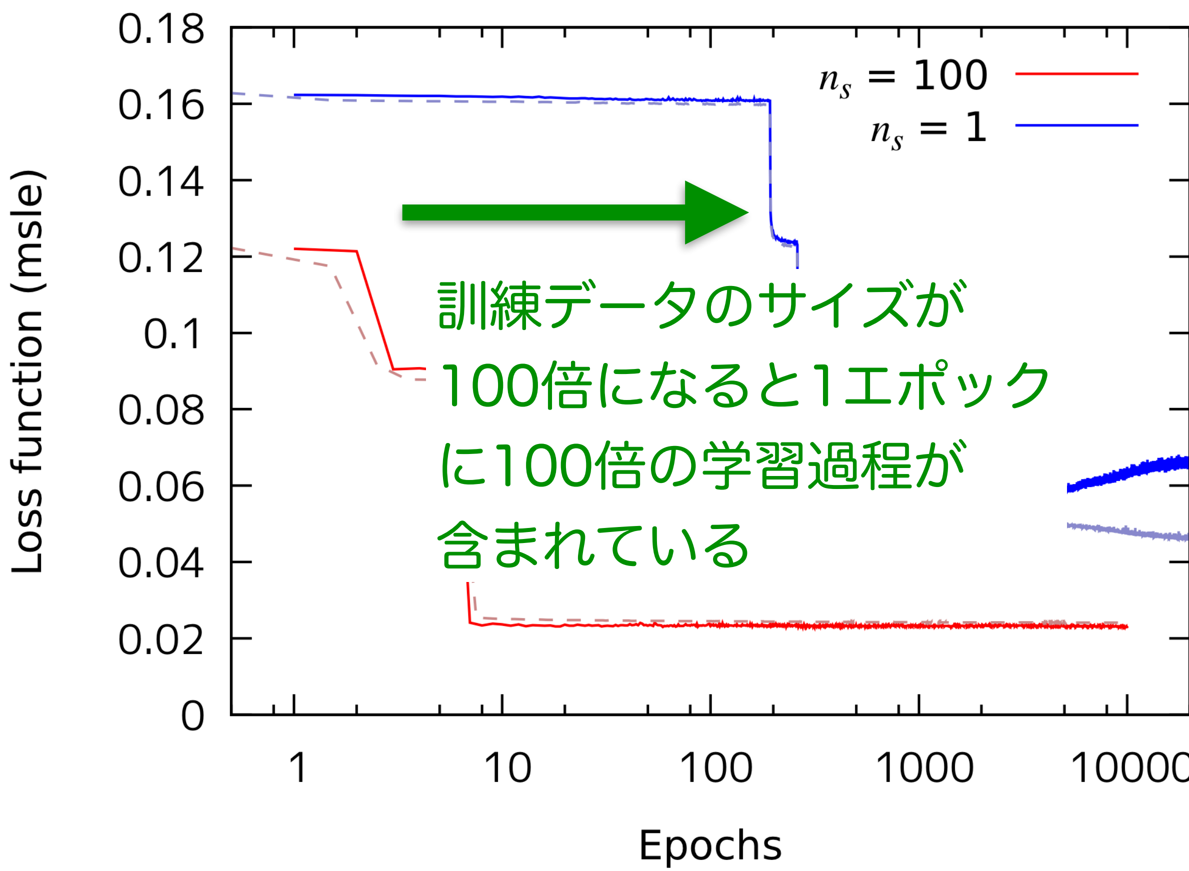
オンライン学習：訓練データひとつずつ計算
（揺らぎが大きい）

ミニバッチ学習：訓練データからランダムに
ミニバッチを選んで計算

エポック：訓練データ全体を使った１プロセス

今日ここで議論したいこと

Fujimoto-Fukushima-Murase (2018)



過学習

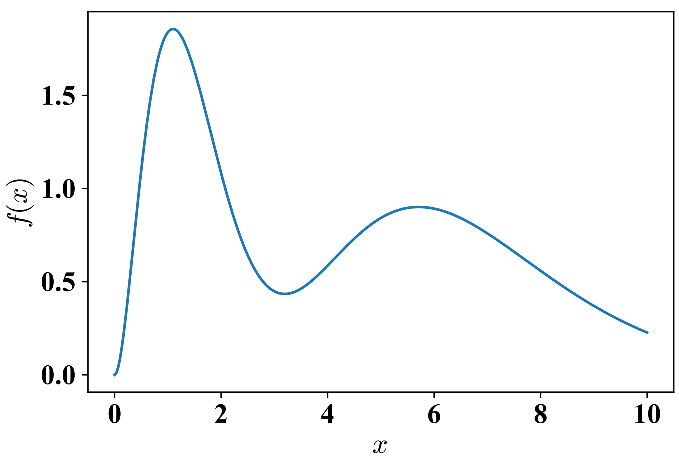
過学習を回避

注目したいもの

今日ここで議論したいこと

【物理屋の発想】

ほとんど自明とも言える最もシンプルな
状況設定を使ってアイデアをテストする



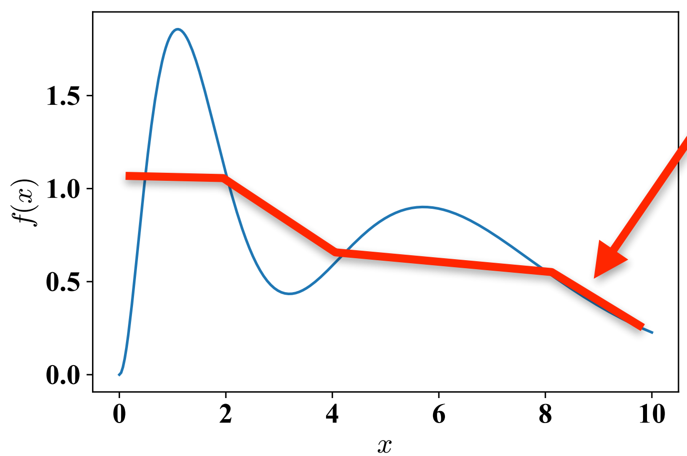
$$f(x) = (10x^2 - 6x^3 + x^4)\, e^{-x}$$

・inputもoutputも1変数
・関数フィットの問題
（雰囲気はスペクトル
関数っぽくして・・・）
・劣化＋雑音は手で処理

今日ここで議論したいこと

【物理屋の発想】

ほとんど自明とも言える最もシンプルな
状況設定を使ってアイデアをテストする



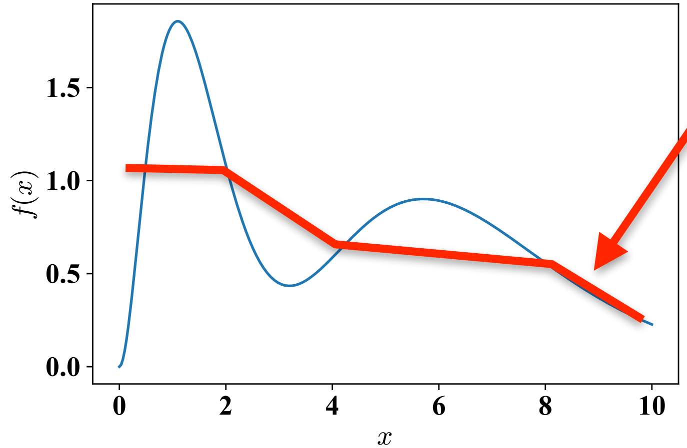
$$f(x) = (10x^2 - 6x^3 + x^4)\, e^{-x}$$

結構local minimumに落ちる
ReLuは全然ダメ(dying ReLu)
sigmoidの方がずっとよい
我々はLeaky ReLuを使う

今日ここで議論したいこと

【物理屋の発想】

物理の問題ではf(x)の計算コストが高い
プロトタイプで時間をかけず試行・実験



$$f(x) = (10x^2 - 6x^3 + x^4)\, e^{-x}$$

結構local minimumに落ちる
ReLuは全然ダメ(dying ReLu)
sigmoidの方がずっとよい
我々はLeaky ReLuを使う

今日ここで議論したいこと

【わざと過学習を起こす設定にしておく】

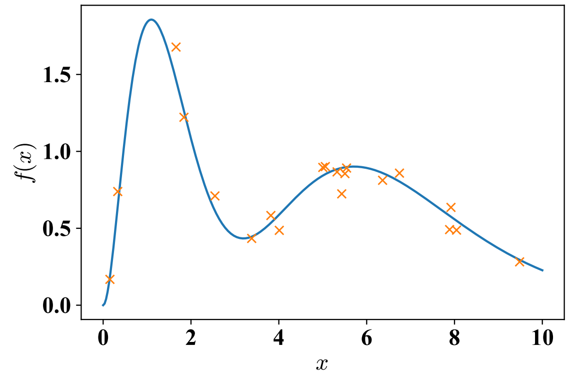
1-2-2-1: 13パラメータ
1-4-4-1: 33パラメータ
1-9-9-1: 118パラメータ

“極端”な状況設定

１層だとlocal minimumに落ちる

２層だとほぼフィットできる

訓練データが十分ないと過学習



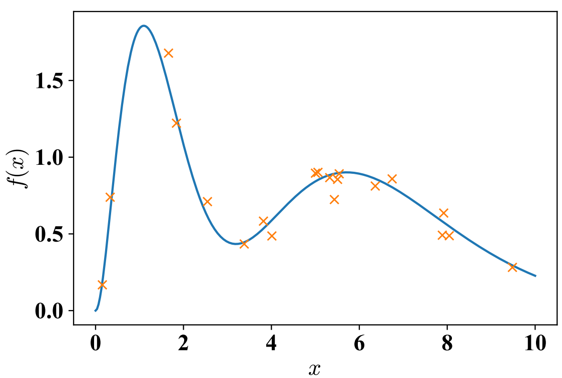
ランダムなxを20個えらぶ → f(x)を計算（計算コスト大）
それぞれのxにランダムノイズdxをのせる

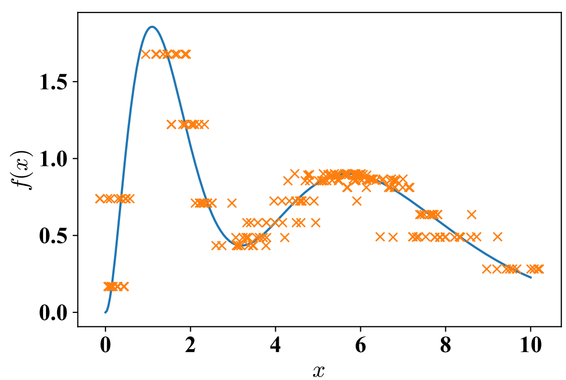
今日ここで議論したいこと

【わざと過学習を起こす設定にしておく】

“極端”な状況設定

ランダムなxを20個えらぶ → f(x)を計算（計算コスト大）
それぞれのxにランダムノイズdxをのせる
n_s個のランダムdxで訓練データかさ増し（計算コストゼロ）



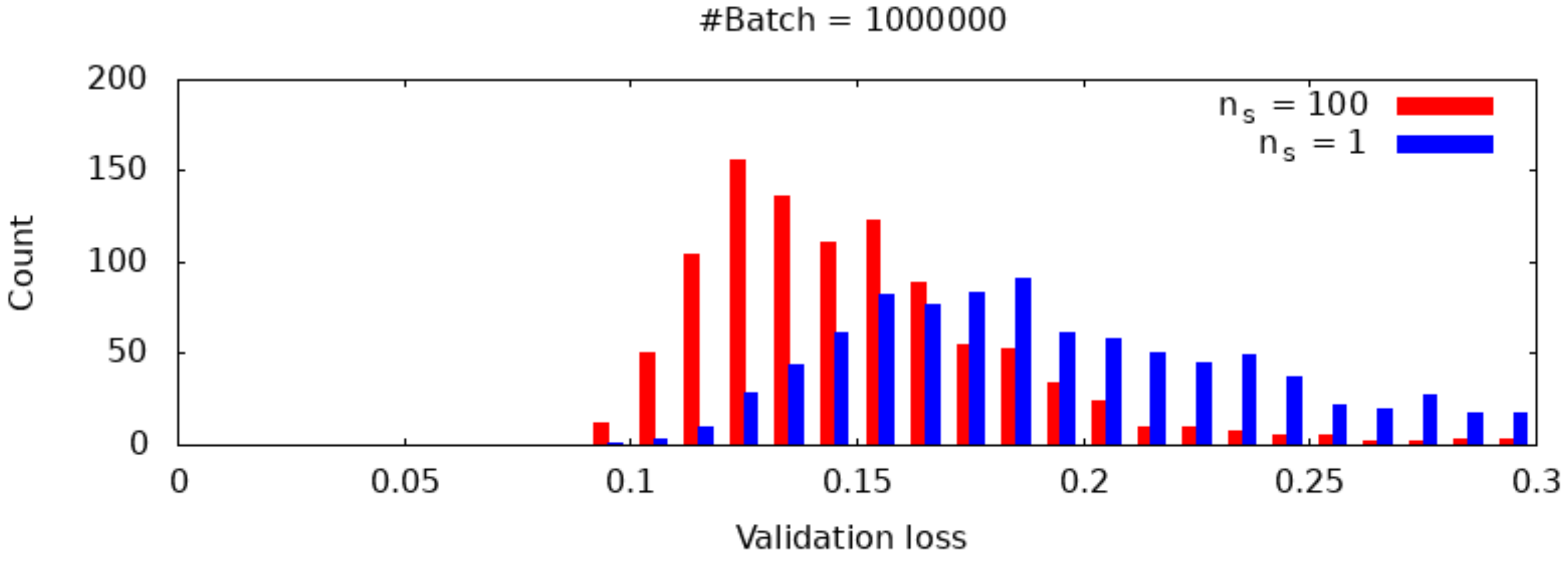


中性子星のとき
観測誤差を扱った
方法のモデル化

今日ここで議論したいこと

【わざと過学習を起こす設定にしておく】

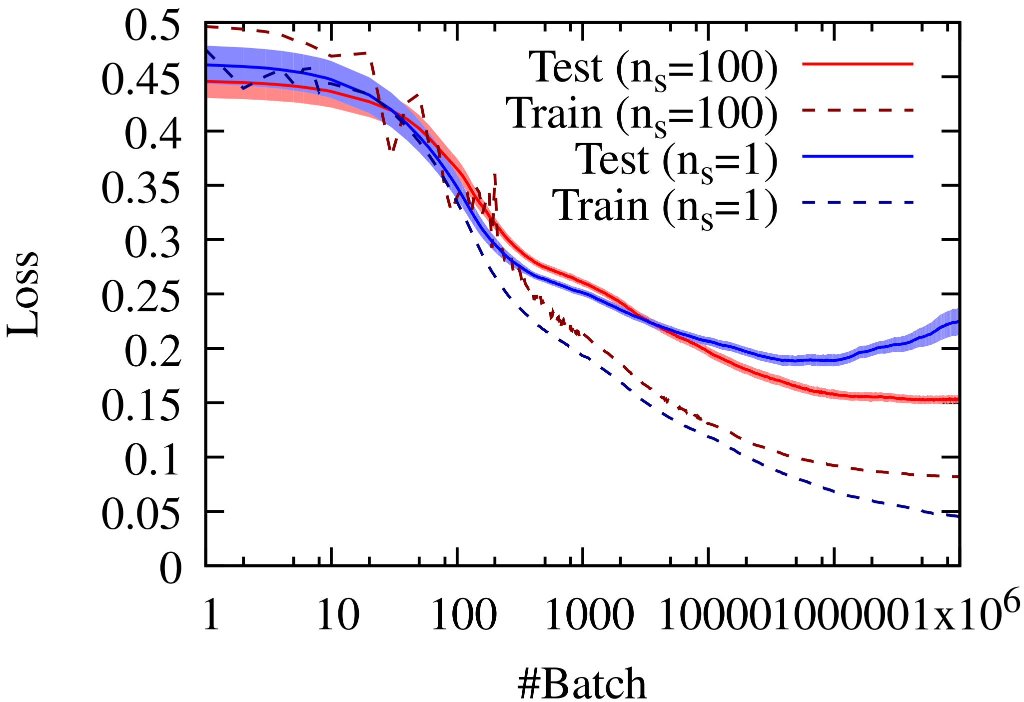
1-4-4-1で過学習した場合



今日ここで議論したいこと

【わざと過学習を起こす設定にしておく】

1-4-4-1で過学習した場合



- 100回訓練した平均
- 実際はn_s=5程度で十分
- 横軸はエポックではない

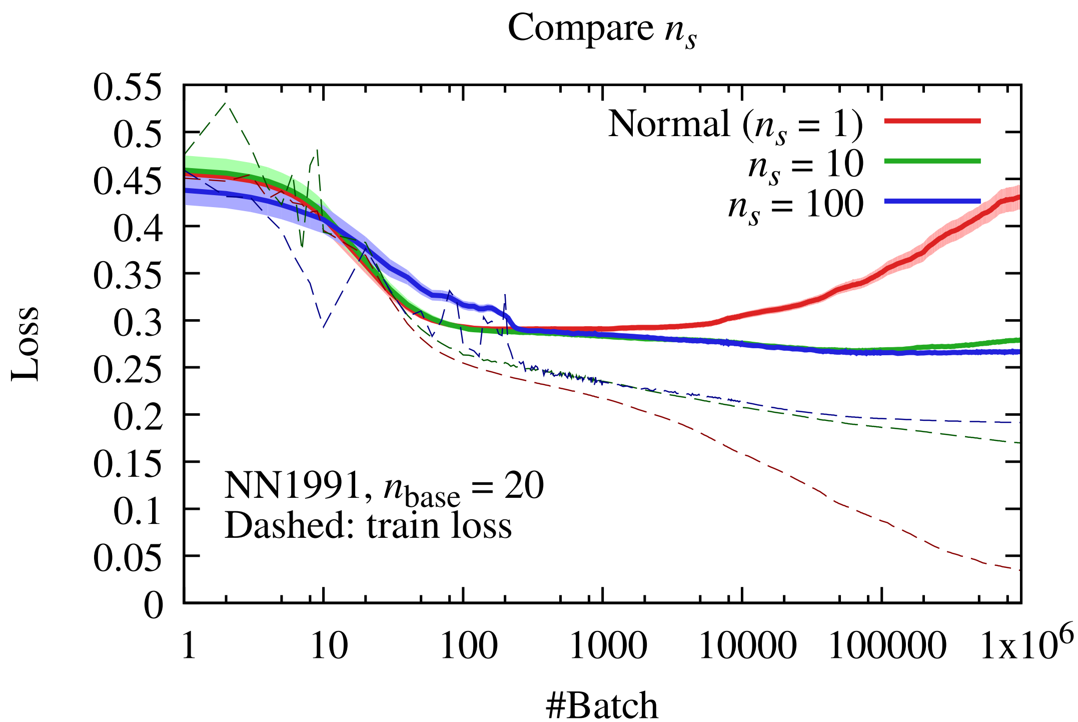
過学習

を抑制！

今日ここで議論したいこと

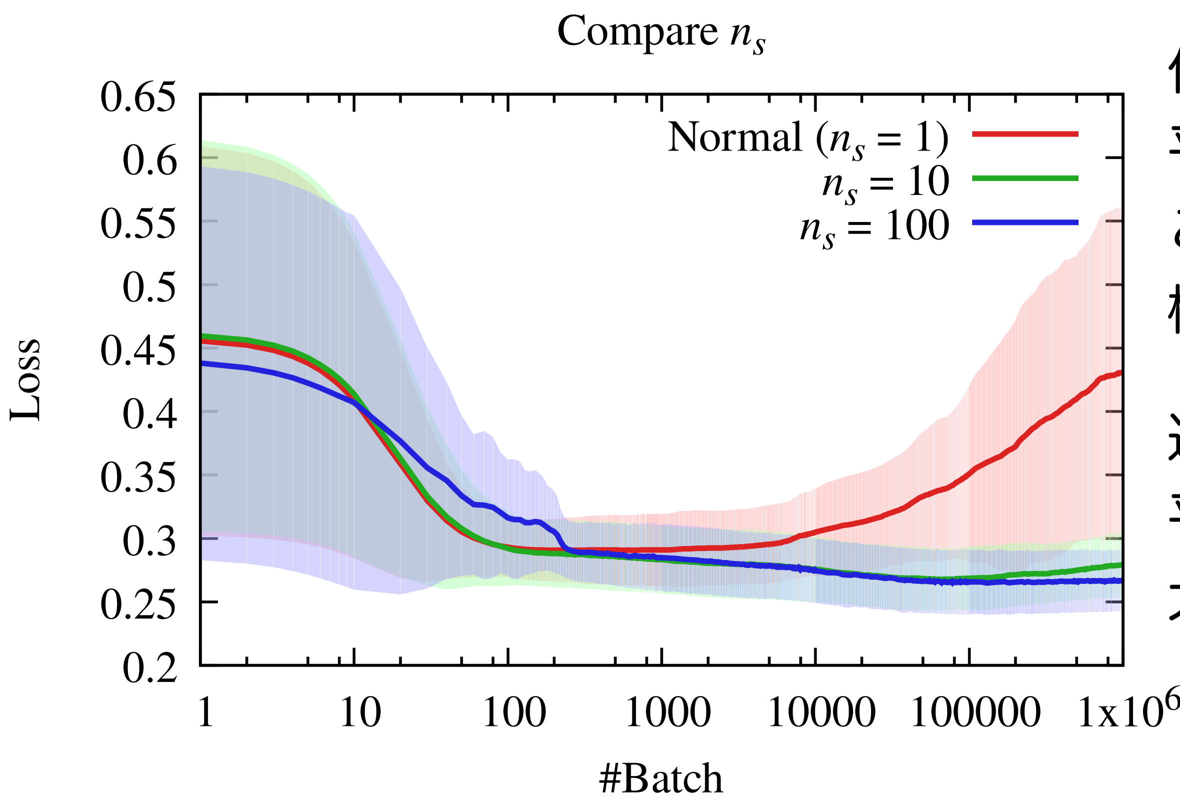
【物理屋らしくもっと極端な設定にする】

1-9-9-1で酷い過学習した場合



今日ここで議論したいこと

【物理屋らしくもっと極端な設定にする】



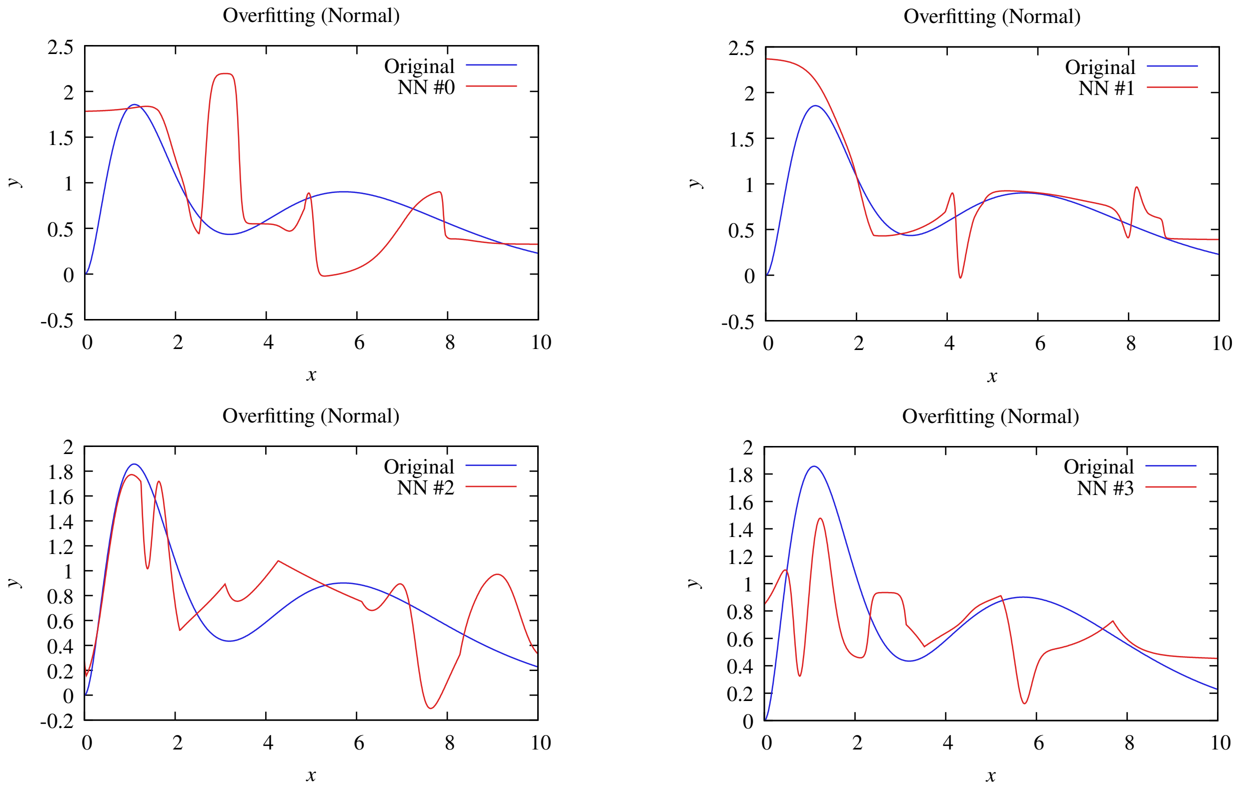
他のグラフでは幅は
平均値の幅を示す
このグラフでの幅は
標準偏差を表す

過学習が抑えられると
平均だけでなく個別でも
大きく外すことがない

今日ここで議論したいこと

【物理屋らしくもっと極端な設定にする】

過学習すると何が起きるのか



今日ここで議論したいこと

dxの揺らぎはbの揺らぎと等価

$$f_i(\boldsymbol{W}_i \cdot \boldsymbol{x}_i + \boldsymbol{b}_i)$$

$$\begin{aligned} &\boldsymbol{W}_i \cdot (\boldsymbol{x}_i + \boxed{\delta \boldsymbol{x}_i}) + \boldsymbol{b}_i \\ &= \boldsymbol{W}_i \cdot \boldsymbol{x}_i + (\boldsymbol{W}_i \cdot \boxed{\delta \boldsymbol{x}_i} + \boldsymbol{b}_i) \end{aligned}$$

{W_i, b_i}はもともと乱数的に発生していたのでは？

（学習は{W_i, b_i}の初期値に強く依存する）

Heの初期化 / Xavierの初期化

今日ここで議論したいこと

dxの揺らぎはbの揺らぎと等価

$$f_i(\boldsymbol{W}_i \cdot \boldsymbol{x}_i + \boldsymbol{b}_i)$$

$$\begin{aligned}&\boldsymbol{W}_i \cdot (\boldsymbol{x}_i + \boxed{\delta \boldsymbol{x}_i}) + \boldsymbol{b}_i \\ &= \boldsymbol{W}_i \cdot \boldsymbol{x}_i + (\boldsymbol{W}_i \cdot \boxed{\delta \boldsymbol{x}_i} + \boldsymbol{b}_i)\end{aligned}$$

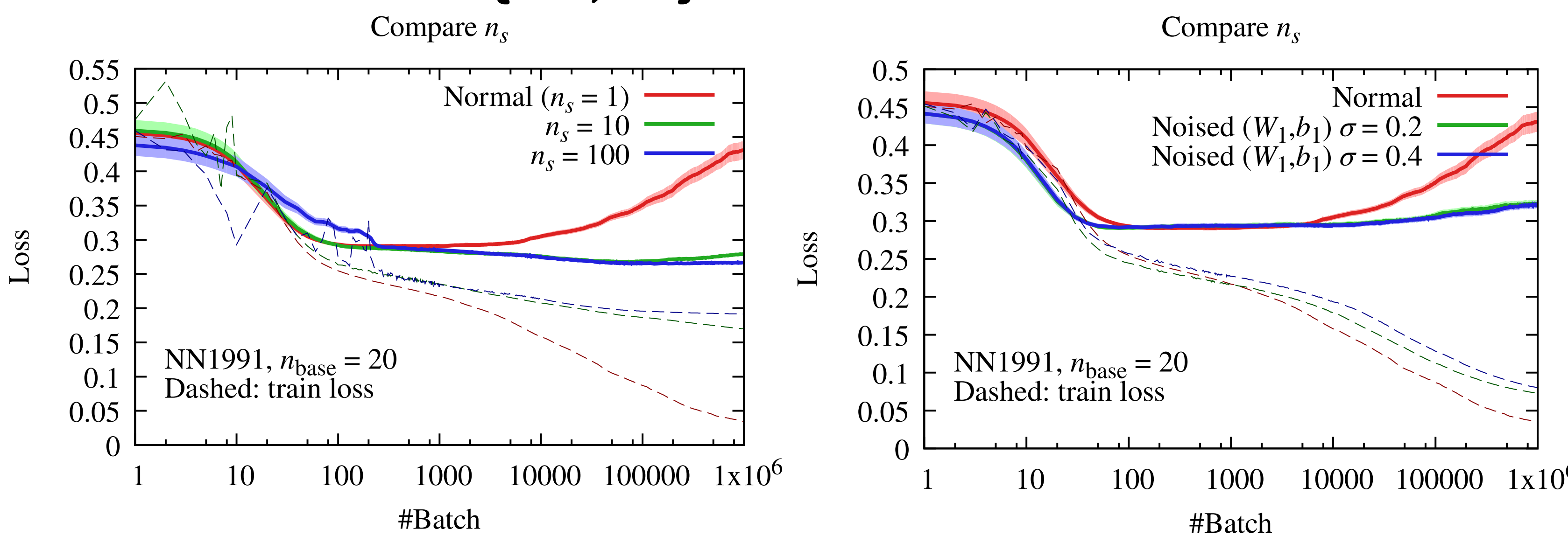
パラメータの初期化は学習の最初だけ

dxによる揺らぎはミニバッチ毎（勾配計算毎）

ミニバッチ毎に{W_i, b_i}を揺らがせたらどうなるか？

今日ここで議論したいこと

ミニバッチ毎に{Wi, bi}を揺らがせたらどうなるか？



１層目の{W, b}を揺らがせるとほぼ同じ結果を得る

（全ての層の{W, b}を揺らがせるとさらに少し改善）

今日ここで議論したいこと

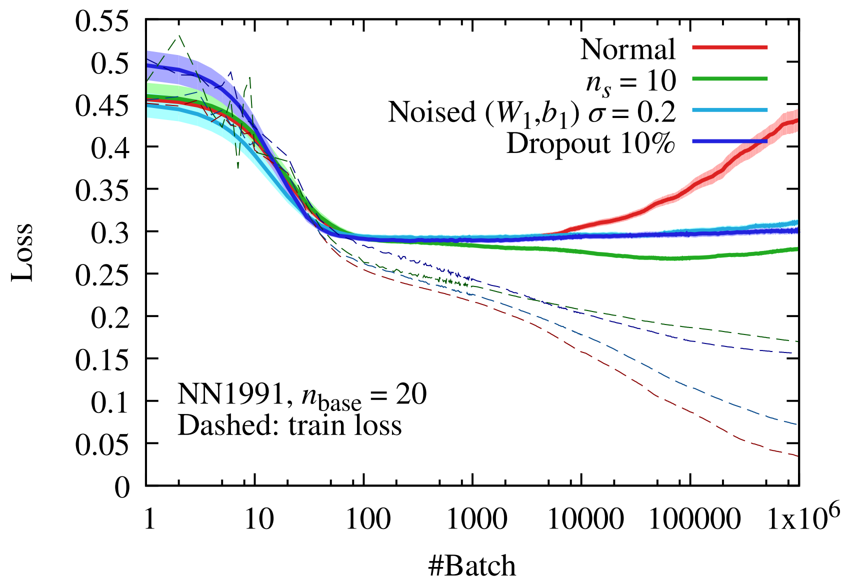
ほぼコストゼロで乱数的に訓練データを複製
NNのbの揺らぎを「外から」コントロールできる

脳の構造を変えずに、教え方でもっと頭を良くするにはどうしたらいいか？

部分的に達成できた？

今日ここで議論したいこと

脳の構造を変えれば過学習を回避する方法はある



1-9-(dropout)-9–1

dropoutは実装するのはいちばんシンプルな方法

n_sのデータ複製はある種のアンサンブル学習に実効的に対応している(かも)

いまの問題だとdropoutは１層目の{W, b}の揺らぎとほぼ同じ過学習抑制効果

今日ここで議論したいこと

我々の方法の持つ利点

- NNの構造には手を加えない
- dropout等のように学習の各プロセスで乱数発生させる必要がないので学習スピードはいちばん速い
- 物理の観測データはもともとこうやって与えられている

皆さまへのお願い

We are wondering:

・こういう問題設定は意味ある研究になっているのだろうか？

・先行研究は絶対にあるに違いない・・・

・異分野すぎて研究のまとめ方がサッパリ分からない・・・

Your comments/feedback are welcome!

まとめ

物理のデータ = 真値＋観測誤差

□物理の問題には原理的な正解が存在する

□データは正解の周りである精度で揺らいでいる

中性子星の事例

□訓練データにノイズ入り複製を作る

□Bootstrap法で信頼幅を推定 (系統誤差を過小評価)

過学習を抑制する新しい手法としての可能性

□ノイズ入り複製はパラメータにバッチ揺らぎを入れることに相当：dropoutと同等の性能を確認