

ランジュバン動力学を用いた 過剰パラメータモデルの汎化ギャップ推定

奥野 彰文 統計思考院 助教

※本研究は矢野恵佑准教授(統計数理研究所・数理推論系)との共同研究であり <https://arxiv.org/abs/2112.03660> にてプレプリントを公開済です。

1. 背景

昨今ではDeep Neural Network (DNN)をはじめ、多数のパラメータを持つ過剰パラメータモデルがよく用いられる。

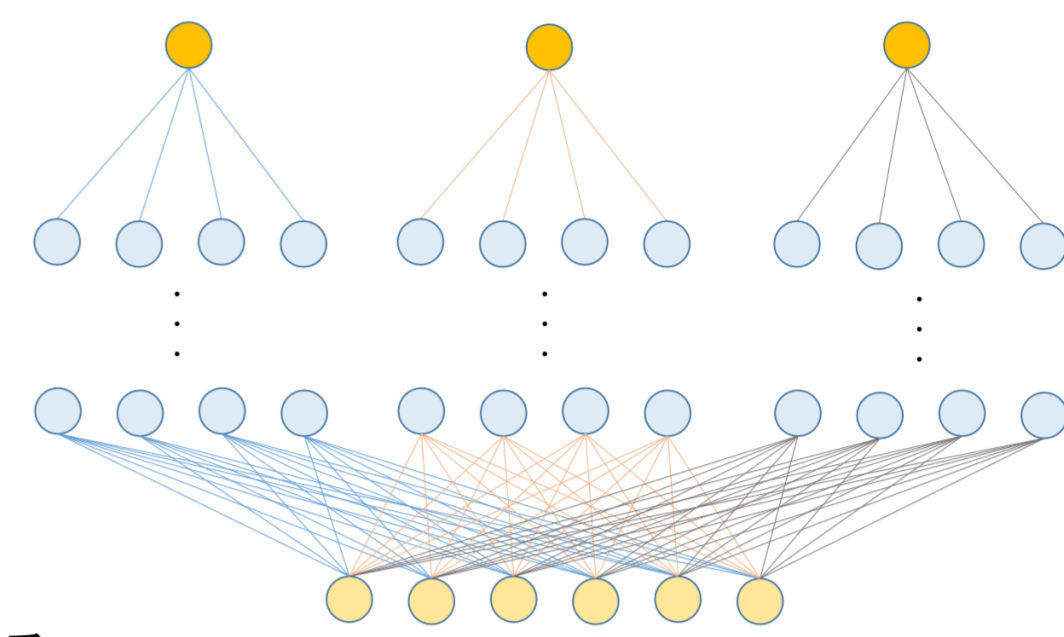
予測モデルのパラメータ数が多くなるとトレーニングデータに過適合しやすく、

汎化ギャップ = テストデータとトレーニングデータに対する当てはまりの差

が大きくなりやすい。そこで汎化ギャップをうまく推定することは応用上重要な問題となっている。

DNNの汎化ギャップ推定でよく用いられるCross-validation (CV)は計算コストが高く、また歴史的によく知られるAICなどの情報量規準はCVを近似するが、DNNなど特異なモデルに適用することができない。

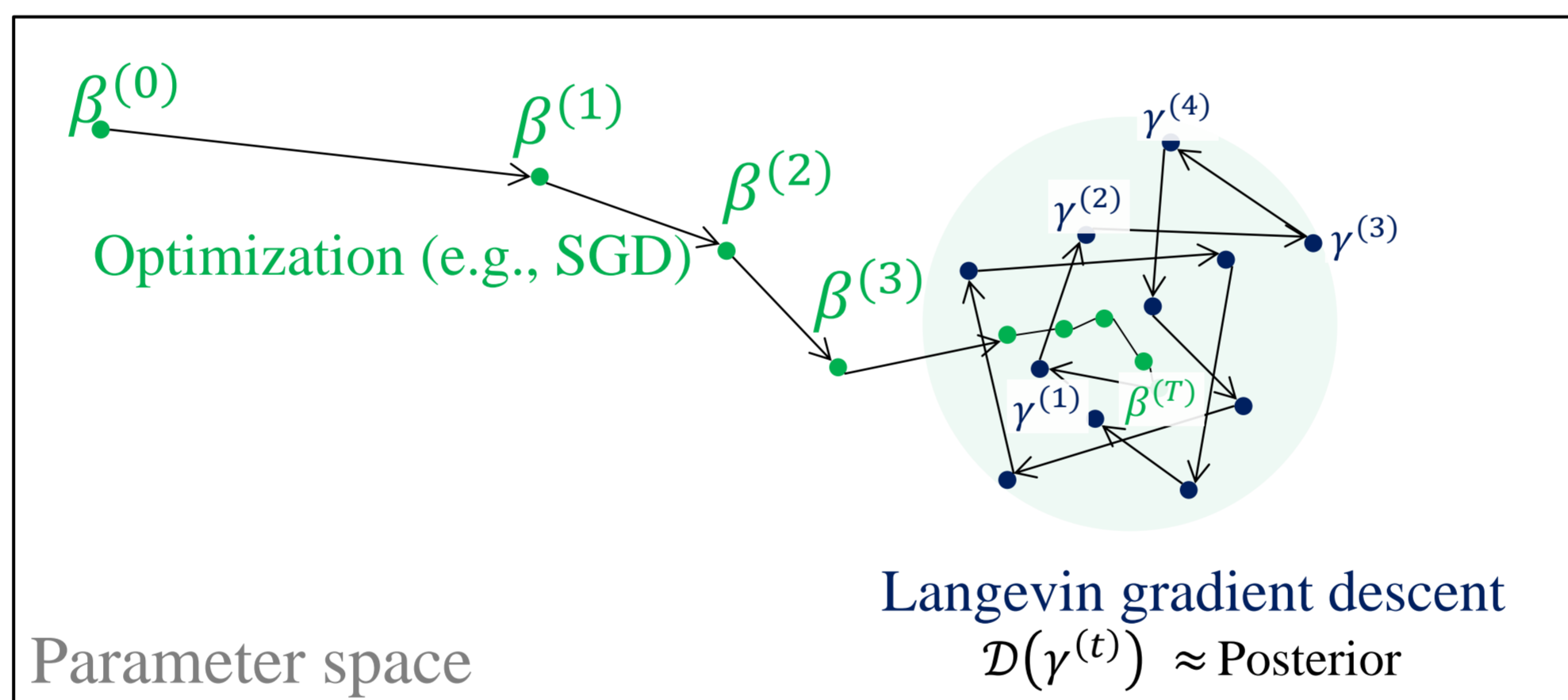
Watanabe (JMLR2010)により特異なモデルに適用できる情報量規準WAICが提案されたが、(理論的問題)過剰パラメータモデルに適用できるか不明であり、(計算的問題)WAICの計算に必要な事後分布をDNNに対して単純に計算するのはコストが大きい、という問題があった。



2. 我々の結果

(理論的貢献) 過剰パラメータモデルの線形近似として高次元線形回帰を考え、WAICの一致性を示した。

(計算的貢献) Langevin動力学により事後分布を近似し、WAICを効率的に計算する方法を提案した。



$$\text{Langevin方程式: } \gamma^{(t+1)} = \gamma^{(t)} - \frac{1}{4} \delta \kappa_n \frac{\partial \ell_\alpha(\gamma^{(t)})}{\partial \gamma} + \delta^{1/2} e^{(t)}, \text{ where } e^{(t)} \sim N(0, I_d).$$

Langevin方程式による事後分布の近似は損失関数の一次勾配しか使わず高速に計算できるほか、勾配情報を利用するPyTorchなどNNのwrapperと互換性が非常に高く実装が極めて容易である。(すでに実装されている勾配法に正規ノイズを加えるだけで実現できる)

表: 1000~2000 weightsを持つoverparameterized NN (灰色)での実験。
 d が入力次元, M が中間素子数, $\bar{\Delta}$ が汎化ギャップ。

| | $M = 50$ | | $M = 100$ | | $M = 150$ | |
|----------|--------------|----------------|--------------|----------------|--------------|----------------|
| | LFV | $\bar{\Delta}$ | LFV | $\bar{\Delta}$ | LFV | $\bar{\Delta}$ |
| $d = 5$ | 6.43 ± 0.96 | 4.26 | 6.49 ± 0.52 | 4.41 | 7.30 ± 0.80 | 4.60 |
| $d = 10$ | 11.03 ± 1.28 | 8.75 | 12.91 ± 1.54 | 9.18 | 13.56 ± 1.14 | 9.64 |
| $d = 15$ | 16.78 ± 1.39 | 17.64 | 18.93 ± 1.57 | 18.60 | 20.13 ± 2.07 | 19.46 |

正確に値を当てることは難しいが、おおよそ汎化ギャップを推定できている。

なお提案法は過剰パラメータモデルを一度しか訓練する必要がなく、CVなど複数回の学習を必要とする方法に比べ計算量の点で大きなアドバンテージがある。