

活性化関数と回帰関数の性能と人の認知からの評価

統計数理研究所 柳本 武美

1. 序

深層学習の枠組み統計的方法と軌を一にしている。活性化関数はその重要な構成要素であるが、その選択は必ずしも良く研究されていない。形式的には類似した回帰関数との対比させ機械学習での損失関数と併せて、解析的な性能と人の認知のための道具の視点から研究している。

2. ReLU 関数と logistic 関数

中間層での代表的な活性化関数は、ReLU 関数: $\max\{x, 0\}$ と logistic 曲線 $e^x/(1 + e^x)$ であり、前者が有望視されている。前者は、非有界な関数で計算の負担が小さい関数として使われている。関数としての単純さが計算量の軽減をもたらしているが、一方で人の認知機能の過程と関連していることが期待される。後者はその多次元版が、最終層でも使われて softmax 関数と呼ばれる。これは指数分布族における双対な平均母数と自然母間の変換関数 [1] と見られ、一般化線形モデルでの自然連結回帰関数として汎用されている。しかし、この事実は最終層で活性化関数としての利用を促すが、中間層での利用を支える事実ではない。また生体の反応、特に五感のインパルスの分布、との関連はなさそうである。むしろ正規分布の分布関数の方がもっともらしい。

3. Hockey stick 回帰関数

ReLU 関数と同じ関数が回帰分析では、hockey stick 解析として用いられる [2]。導入の動機はモデルの簡明さである。しかし、接点に格別な意味を持たせている点が問題である [3]。特に、一般市民に対する安全量と関連していたからである。この点は ReLU 関数と異なり、接点に格別の意味はない。一方で、モデルの意味の取りやすさは、回帰関数として魅力的である。定数値と線形関数の組み合わせであるから、最も簡便な形である。人の認知過程との類似性が示唆される。機械学習においても損失関数として用いられていて、hinge 関数と呼ばれる。また、Mathematica 等では ramp 関数と呼ばれるが、近年の一般語としての ramp とは異なる。

4. Ramp 関数族

その基本形が $\max\{0, \min\{x, 1\}\}$ となる ramp 関数は、簡便でその分布族への拡張も柔軟で容易である。深層学習では hard sigmoid として登場したが、特に logistic 関数とは関係なく誤差関数を機械学習の損失関数としては ramp 関数とよばれる [4]。Logistic 関数に代表される sigmoid 関数とは特別な関係はない。知られている分布族では、一般 Cauchy 分布 [5] の裾の軽い極限分布導出できる。その簡便さと共に拡張の容易さが特徴である。その意味で、魅力的な活性化関数の族を定義する。試験的な数値実験によればその性能は多様であるが、ramp 関数は比較的良好な性能を示す [6]。計算量が大きく軽減できる長所もある。

一方で回帰関数としての利用が見込まれる。問題の一つは二つの折点の解釈にあるが、特別な意味を持たせなければ無理がない。この点は hockey stick 回帰関数を利用する場合と同じである。折点人が人にとって理解が容易である長所を用いた解析が期待される。

文献：[1] Huzurbazar, V.S. (1956). Sankhya, 17, 217-220. [2] Hasselblad, V. et al. (1973). Read at 101st Meet. Amer. Pub. Health A. [3] Yanagimoto, T. and Yamamoto, E. (1979). Env. Health Persp., 32, 193-199. [4] Collobert, R.F. et al. (2006). J. Machine L. Res., 7, 1687-1712. [5] Rider, P.R. (1957). Ann. Inst. Statist. Math., 9, 215-223. [6] 大草孝介 (2018). 個人的情報.