P-92

基礎科学

深層学習の過剰適合に対する正則化手法の解析

■ はじめに

近年，訓練データ数より遥かに多いパラメータを持つ深層学習モデルを使用して，訓練誤差を最小にするようにパラメータを更新する．しかし，訓練誤差が0になるまで学習を続けると，過剰信頼(overconfidence)状態になってしまい，テストデータに対する精度が低くなってしまう．この問題を過剰適合(overfitting)と呼び，深層学習モデルを学習する上で重要な課題である．本論文では，過剰適合を防ぐために提案されたLabel Smoothing[1]とFlood Regularization[2]に焦点を当てる．この2つの手法は，アプローチは異なるが，最小化したい損失関数に正則化を加えるという点で共通している．本論文では，それぞれの正則化の効果を解析し，過剰適合に対するより良い正則化について議論する．

■ 関連研究

1. Label Smoothing[1]

多クラス分類において，教師ラベルをソフトにして過学習を防ぐ手法である．正解クラスは1，それ以外は0のone-hotベクトル$\boldsymbol{y}$の代わりに，$\boldsymbol{y}$と一様分布によって平滑化されたソフトなラベルを使用し，以下の損失関数を最小化するようにパラメータ更新を行う．

$$\mathcal{L}_{ls} = -(1-\lambda)\sum_{i=1}^{K} t_i \log p_i - \lambda \sum_{i=1}^{K} \log\ p_i$$

t_iはone-hotベクトルであり，p_iはiクラスに対するソフトマックス確率を表している．これにより，正解クラスに対する過剰な予測を防ぐ．

2. Flood Regularization[2]

過剰適合を防ぐため，訓練誤差をある閾値bよりも小さくならないような正則化手法である．訓練誤差がbを下回った場合，損失の符号が反転し，パラメータの更新方向の符号が反転することで，先鋭解への到達を回避している．損失関数は以下の通りである．

$$\mathcal{L}_{ls} = \left| -\sum_{i=1}^{K} t_i \log p_i - b \right| + b$$

非常にシンプルな手法であり，識別だけではなく，回帰にも使用できる．

■ 解析結果

クロスエントロピー損失において，ロジット(ソフトマックス関数への)と正則化の性質について確認する．1つのロジットを用いた2クラス分類の場合を考える．この時のロジットと損失・勾配のグラフは図1のようになる． 通常のクロスエントロピー損失はロジットが大きくなるにつれて，損失と勾配が0に近づく．しかし，正の勾配を一度も取らないため，学習を続けるとロジットが無限大に向かう．一方，Label SmoothingとFlood Regularizationは正の勾配を取る．この2つの正則化の違いは，ロジットが負の部分での傾き，滑らかに勾配が上昇するか否かと正の勾配が減少するか否かである．Label Smoothingはロジットが大きくなるにつれて，勾配が滑らかに大きくなり続ける．一方，Flood Regularizationはある地点で急激に勾配の符号が反転し，ロジットが大きくなりすぎると勾配が減衰する．

また，実際の多クラス分類のベンチマークで実験を行い，ロジットを分析した．使用したデータはCIFAR-10，モデルはResNet34を使用し，200エポックの学習を行った．各手法における，ロジットの分散の平均，テストデータに対する損失と正答率について確認した(表1)．どちらの正則化手法も通常より精度は高い．しかし，ロジットの分散が小さくなるのはLabel Smoothingであり，大幅にテスト損失が下がるのはFlood Regularizationである．

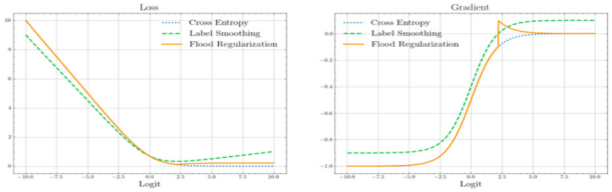

図 1: ロジットと正則化の関係

	Variance of Logit (avg)	Test Loss (avg)	Test Acc.
Normal	18.64	0.265	94.7
Label Smoothing[1]	**3.42**	0.212	**95.13**
Flood Regularization[2]	6.23	**0.164**	94.88

表 1：CIFAR-10 での実験における各手法の評価

[1] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[2] Ishida, Takashi, et al. "Do we need zero training loss after achieving zero training error?." *arXiv preprint arXiv:2002.08709* (2020).

代表発表者 **福田　幸平（ふくだ　こうへい）**
所　　属 **広島大学先進理工系科学研究科**
先進理工系科学専攻情報科学プログラム
問合せ先 **〒739-8527 広島県東広島市鏡山1丁目4番1号**
TEL: 090-3656-9962
Email: m231319@hiroshima-u.ac.jp

■キーワード: (1)機械学習
(2)正則化
(3)過剰適合
■共同研究者: 小林　匠(産総研)
栗田　多喜夫(広島大、産総研)