



基于截断 1-范数损失函数的鲁棒超限学习机

摘要

对噪声和异常值较敏感、鲁棒性差是超限学习机(ELM)的主要问题.在 1-范数损失函数的基础上,提出截断 1-范数损失函数来抑制噪声和异常值的影响,建立了基于截断 1-范数损失函数的鲁棒 ELM 模型.通过迭代重赋权算法求解对应的优化问题,并利用 4 个模拟数据集和 9 个真实数据集验证模型的有效性.数值实验结果表明,在噪声环境下鲁棒 ELM 的泛化性能优于对比方法,并且具有较强的鲁棒性,尤其是在异常值比例较大的情形下.

关键词

神经网络;超限学习机;鲁棒;截断损失函数;异常值

中图分类号 TP181

文献标志码 A

收稿日期 2019-06-28

资助项目 国家自然科学基金(61833005, 61907033);中国博士后基金(2018M642129)

作者简介

王快妮,女,博士,讲师,研究方向为超限学习机及其应用.wangkuaini1219@sina.com

曹进德(通信作者),男,博士,教授,博士生导师,研究方向为复杂网络动力学分析.jdcao@seu.edu.cn

1 西安石油大学 理学院,西安,710065

2 东南大学 数学学院,南京,211189

0 引言

超限学习机(Extreme Learning Machine, ELM)是由 Huang 等^[1-4]提出的一种有效解决模式识别和回归问题的机器学习方法.作为一种单隐层前馈神经网络,其显著特点是隐层节点的输入权值是随机产生的,并通过求解线性方程组获得输出权值.自提出以来,超限学习机受到机器学习界和优化界学者的广泛关注,在理论研究和应用方面都取得了巨大进展.ELM 的程序设计简单,在学习过程中无需迭代,具有极快的学习速度.已有研究表明,ELM 具有良好的泛化性能和近似逼近能力.ELM 目前已经被成功地应用于生物信息学、图像处理、特征选择等领域^[5-7].

超限学习机采用 2-范数损失函数表示经验风险,是基于训练点误差服从正态分布的假设.然而,很多实际问题并不能保证误差服从正态分布,比如测量仪器、测量手段、统计误差等因素,导致数据集通常含有噪声和异常值^[8-9].在对这些样本数据构建学习模型时,需要模型对这些数据具有鲁棒性,即模型所得的决策函数受数据中噪声的影响较小^[9].超限学习机在训练中忽略了数据集的具体特点,过分强调这类异常值的作用,建立的模型对异常值较敏感、不具有鲁棒性,不能很好地满足实际需求.

超限学习机对噪声和异常值较敏感、鲁棒性差,研究学者从正则化技术、加权形式和鲁棒损失函数等方面提出了许多超限学习机的改进模型.根据 Bartlett 理论,Huang 等^[4]将权值的 2-范数作为正则项添加到 ELM 模型中,提出正则化 ELM 模型,能够有效提升模型的推广性能. Deng 等^[10]将加权最小二乘思想引入到正则化 ELM,建立加权正则化 ELM 模型,数值实验表明该模型对噪声数据具有较强的鲁棒性.为了抑制 ELM 中 2-范数损失函数对噪声数据过分强调,Zhang 等^[11]提出 1-范数损失函数以削弱噪声对 ELM 学习性能的影响,建立基于 1-范数损失函数和 2-范数正则化项的鲁棒 ELM 模型. Horata 等^[12]使用 Huber 损失函数来提升 ELM 模型的鲁棒性,并通过迭代重赋权算法求解实现,然而该模型没有正则化项,容易导致过拟合问题. Chen 等^[13]提出基于迭代重赋权的鲁棒正则化 ELM 统一模型,它采用 4 个损失函数来提升模型鲁棒性,使用 2-范数和 1-范数正则化项避免过拟合问题.然而,上述方法的思想是根据分布信息赋予样本不同的权值,从而降低噪声和异常值的影响.研究人员从鲁棒损

失函数的角度,提出了相关熵损失函数^[14]和可调节损失函数^[15],构建鲁棒 ELM 模型,提升改进模型对异常值的鲁棒性.

随着研究的深入,人们注意到截断损失函数在推广能力和鲁棒性等方面具有学习优势^[14-16].截断损失函数能够直接限制噪声和异常值引起的最大惩罚值,即明确抑制这类样本点对决策超平面的负面影响.因此,建立的模型具有鲁棒性.虽然 1-范数损失函数能够在一定程度上抑制噪声和异常值对 ELM 的影响,但它是误差的线性损失函数,仍会受到较大误差的样本点的影响.因此,本文在 1-范数损失函数的基础上,提出截断 1-范数损失函数,并建立基于该损失函数的鲁棒 ELM 模型 (Robust ELM with truncated 1-norm loss function, RTELM),采用迭代重赋权算法求解相应的优化问题,通过模拟数据集和真实数据集在不同异常值比例情形下的数值实验验证模型的有效性.实验结果表明,提出的 RTELM 具有良好的泛化性能和较强的鲁棒性.

1 超限学习机

对回归估计问题,设训练集为 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$,其中 $\mathbf{x}_i \in \mathbf{R}^d, y_i \in \mathbf{R}$. L 个隐节点的 ELM 的输出函数具有如下表达式:

$$f(\mathbf{x}) = \sum_{i=1}^L \mathbf{h}_i(\mathbf{x}) \boldsymbol{\beta} = \mathbf{h}(\mathbf{x}) \boldsymbol{\beta}, \quad (1)$$

其中: $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_L]^T$ 是输出权值向量, $\mathbf{h}(\mathbf{x}) = [\mathbf{h}_1(\mathbf{x}), \mathbf{h}_2(\mathbf{x}), \dots, \mathbf{h}_L(\mathbf{x})]$ 为变量 \mathbf{x} 的隐层输出. ELM 模型^[1-2,4]可表达为如下优化问题:

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{C}{2} \sum_{i=1}^N l_1(e_i), \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{x}_i) \boldsymbol{\beta} = y_i - e_i, \quad i = 1, 2, \dots, N, \end{aligned} \quad (2)$$

其中: $l_1(e_i) = e_i^2$ 是 2-范数损失函数, e_i 表示训练误差变量, C 为正正则化参数. 由最优性条件,式(2)的最优解为

$$\boldsymbol{\beta} = \begin{cases} (\mathbf{H}^T \mathbf{H} + \mathbf{I}/C)^{-1} \mathbf{H}^T \mathbf{y}, & N \geq L, \\ \mathbf{H}^T (\mathbf{H} \mathbf{H}^T + \mathbf{I}/C)^{-1} \mathbf{y}, & N < L. \end{cases} \quad (3)$$

其中: $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$, \mathbf{I} 为单位矩阵, 隐层输出矩阵 $\mathbf{H} = [\mathbf{h}(\mathbf{x}_1), \mathbf{h}(\mathbf{x}_2), \dots, \mathbf{h}(\mathbf{x}_N)]^T$.

2 基于截断 1-范数损失函数的鲁棒超限学习机 (RTELM)

2.1 截断 1-范数损失函数

如上所述,采用 2-范数损失函数的 ELM 对噪声

和异常值较敏感^[3-4].当样本点中存在异常值时,这些点通常具有较大的误差并且远离其他样本点,这导致 ELM 的决策超平面严重偏离原位置,影响 ELM 的学习性能.研究人员提出 1-范数损失函数 $l_2(e_i) = |e_i|$ 减少异常值的影响^[11,13].图 1 给出了 2-范数损失函数和 1-范数损失函数.从图 1 可以看出,在具有较大的误差的样本点处,与 1-范数损失函数相比,2-范数损失函数的夸大作用非常明显.注意到 1-范数损失函数仍然是误差的线性损失函数,较大误差的样本点仍会在 ELM 模型中起到主要作用.

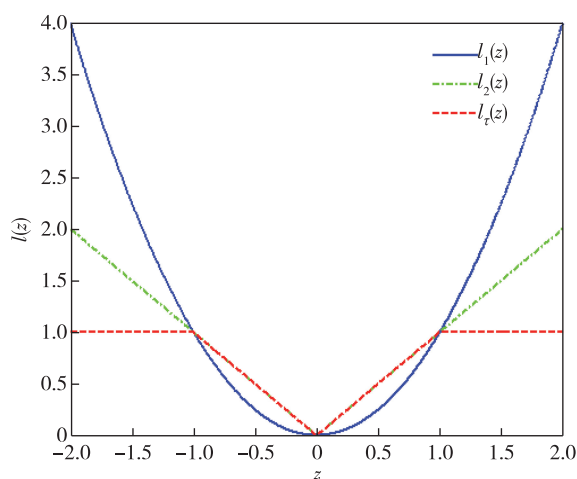


图 1 2-范数损失函数 $l_1(z)$ 、1-范数损失函数 $l_2(z)$ 和截断 1-范数损失函数 $l_\tau(z)$

Fig. 1 2-norm loss function $l_1(z)$, 1-norm loss function $l_2(z)$ and truncated 1-norm loss function $l_\tau(z)$

在 1-范数损失函数的基础上,提出截断 1-范数损失函数:

$$l_\tau(e_i) = \begin{cases} |e_i|, & |e_i| \leq \tau, \\ \tau, & |e_i| > \tau, \end{cases} \quad (4)$$

其中: $\tau > 0$ 为预先给定的异常值的最大惩罚, $l_\tau(e_i)$ 的上界表明当误差超过某定值时,损失函数 $l_\tau(e_i)$ 值为常数,能够抑制异常值的负面影响.

2.2 基于截断 1-范数损失函数的鲁棒超限学习机 (RTELM)

基于截断 1-范数损失函数,建立能够抑制异常值影响的鲁棒 ELM 模型,其对应的优化模型具有如下形式:

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & \frac{1}{2} \|\boldsymbol{\beta}\|^2 + C \sum_{i=1}^N l_\tau(e_i), \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{x}_i) \boldsymbol{\beta} = y_i - e_i, \quad i = 1, 2, \dots, N. \end{aligned} \quad (5)$$

为了求解优化问题(5),构造如下拉格朗日

函数:

$$\ell(\boldsymbol{\beta}, e_i, \boldsymbol{\alpha}) = \frac{1}{2} \|\boldsymbol{\beta}\|^2 + C \sum_{i=1}^N l_{\tau}(e_i) - \sum_{i=1}^N \alpha_i (\mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} - y_i + e_i). \quad (6)$$

由最优性条件,可得:

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \boldsymbol{\beta} - \sum_{i=1}^N \alpha_i \mathbf{h}^T(\mathbf{x}_i) = 0 \Rightarrow \boldsymbol{\beta} = \mathbf{H}^T \boldsymbol{\alpha}, \quad (7)$$

$$\frac{\partial \ell}{\partial e_i} = C \frac{\partial l_{\tau}(e_i)}{\partial e_i} - \alpha_i = 0 \Rightarrow \alpha_i = C \frac{\partial l_{\tau}(e_i)}{\partial e_i} = C w_i e_i, \quad (8)$$

$$\frac{\partial \ell}{\partial \alpha_i} = \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} - y_i + e_i = 0 \Rightarrow \mathbf{H}\boldsymbol{\beta} - \mathbf{y} + \mathbf{e} = 0, \quad (9)$$

其中:

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T,$$

$$w_i = \frac{\partial l_{\tau}(e_i)}{\partial e_i} / e_i = \begin{cases} 1/\max(|e_i|, 10^{-6}), & |e_i| \leq \tau, \\ 0, & |e_i| \geq \tau, \end{cases}$$

$$\mathbf{e} = [e_1, e_2, \dots, e_N]^T.$$

由式(7)一(9),可得最优解 $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \begin{cases} (\mathbf{H}^T \mathbf{W} \mathbf{H} + \mathbf{I}/C)^{-1} \mathbf{H}^T \mathbf{W} \mathbf{y}, & N \geq L, \\ \mathbf{H}^T (\mathbf{W} \mathbf{H} \mathbf{H}^T + \mathbf{I}/C)^{-1} \mathbf{W} \mathbf{y}, & N < L. \end{cases} \quad (10)$$

对角矩阵 $\mathbf{W} = \text{diag}\{w_1, w_2, \dots, w_N\}$.

2.3 迭代重赋权算法求解 RTELM

输入:训练集 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$,正则化参数 C ,隐层节点数 L ,最大迭代次数 k_{\max} 和预先给定的很小的正数 ρ .计算隐层输出矩阵 \mathbf{H} ,令 $\mathbf{W}^{(0)} = \mathbf{I}$ 和 $k = 1$.

步骤 1: 根据下式,计算 $\boldsymbol{\beta}^{(k)}$,

$$\boldsymbol{\beta}^{(k)} = \begin{cases} (\mathbf{H}^T \mathbf{W}^{(k-1)} \mathbf{H} + \mathbf{I}/C)^{-1} \mathbf{H}^T \mathbf{W}^{(k-1)} \mathbf{y}, & N \geq L, \\ \mathbf{H}^T (\mathbf{W}^{(k-1)} \mathbf{H} \mathbf{H}^T + \mathbf{I}/C)^{-1} \mathbf{W}^{(k-1)} \mathbf{y}, & N < L. \end{cases} \quad (11)$$

由表达式(1),可得 $f^{(k)}(\mathbf{x})$.

步骤 2: 计算 $e_i^{(k)} = y_i - f^{(k)}(\mathbf{x}_i)$,设置对角矩阵 $\mathbf{W}^{(k)}$,其中: $w_i^{(k)}$ 为对角元素.

步骤 3: 由式(11) 可得 $\boldsymbol{\beta}^{(k+1)}$.

步骤 4: 若 $k > k_{\max}$ 或 $\|\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}\| \leq \rho$ 成立,则停机;否则,转步骤 5.

步骤 5: 由式(1) 得到 $f^{(k+1)}(\mathbf{x})$.令 $k = k + 1$,转步骤 2.

3 数值实验

为了验证本文提出的 RTELM 的有效性,对 4 个

模拟数据集和 9 个真实数据集进行数值实验,并与 ELM^[4]、加权 ELM (WELM)^[10]、1-范数 ELM (即异常值鲁棒 ELM, Outliers-Robust ELM, ORELM)^[11] 和迭代重加权 ELM (Iteratively Re-Weighted ELM, IR-WELM)^[12] 算法进行比较.正则化参数 C 的搜索范围为 $\{2^{-19}, 2^{-18}, \dots, 2^{18}, 2^{19}\}$, RTELM 中参数 τ 的搜索范围为 $\{0.05, 0.10, 0.15, \dots, 0.85, 0.90, 0.95\}$. 在实验中,选取 Sigmoid 函数为激活函数,隐节点个数和最大迭代次数分别设置为 500 和 20.这里采用均方根误差 (RMSE, 其量值记为 E_{RMS}) 评价算法的学习性能^[15]:

$$E_{\text{RMS}} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2},$$

其中: m 为测试样本点个数, y_i 和 \hat{y}_i 分别表示输出的真实值和预测值.

3.1 模拟数据实验

在模拟实验部分,模拟数据集由 $y = \text{sinc}(x)$ 函数生成^[15].在实验中,噪声分布来自正态分布 $N(0, 0.15^2)$ 、 $N(0, 0.3^2)$ 和均匀分布 $U[-0.15, 0.15]$ 、 $U[-0.3, 0.3]$.为了验证 RTELM 的鲁棒性,对未添加异常值、添加异常值的模拟数据集进行训练学习,这些异常值是由集合 $[y_{\min}, y_{\max}]$ 的随机值添加到某些训练样本的输出值上而生成的.对于每个数据集,都进行 10 次独立实验,实验结果(均方根误差 \pm 标准偏差, $\text{RMSE} \pm \text{Std}$) 是 10 次实验的平均结果,如表 1 所示,其中用黑体表示实验所得的最好结果.

由表 1,在未添加异常值的情形下,ELM 获得了最好的预测精度,RTELM 的学习结果较差.然而,在添加异常值的情形下,ELM 的预测精度下降,RTELM 的评价准则 RMSE 均排名第一,且 RTELM 的预测精度明显优于其他对比方法.因此,RTELM 能够有效处理受噪声和异常值干扰的数据集,具有较强的鲁棒性.

图 2 给出在不同分布噪声干扰下,添加异常值情形下的各算法所得的回归曲线.由图 2 可知,在添加异常值情形下,这些算法所得的回归曲线都不同程度地偏向异常值,导致所得曲线偏离了原始曲线.同时,在 4 种情形下,我们发现 RTELM 所得的回归曲线总是最接近原始曲线.

3.2 真实数据集实验

在本小节中,采用 9 个真实数据集进一步验证

表 1 模拟数据集实验结果

Table 1 Experimental results on simulated datasets

噪声类别	异常值情形	ELM (RMSE±Std)	WELM (RMSE±Std)	ORELM (RMSE±Std)	IRWELM (RMSE±Std)	RTELM (RMSE±Std)
$N(0,0.15^2)$	未添加	0.033 2±0.010 4	0.034 1±0.009 0	0.040 2±0.008 2	0.034 2±0.008 9	0.040 1±0.008 2
	添加	0.155 4±0.016 4	0.125 3±0.018 3	0.095 0±0.015 6	0.112 0±0.017 9	0.070 7±0.017 3
$N(0,0.3^2)$	未添加	0.065 3±0.014 7	0.068 8±0.014 5	0.071 3±0.018 5	0.069 8±0.015 8	0.065 7±0.020 3
	添加	0.170 7±0.017 9	0.160 8±0.018 2	0.153 0±0.023 1	0.158 9±0.019 8	0.149 4±0.024 4
$U[-0.15,0.15]$	未添加	0.021 3±0.005 0	0.021 3±0.005 0	0.030 0±0.005 5	0.021 3±0.005 0	0.029 5±0.005 7
	添加	0.152 7±0.007 8	0.098 6±0.016 4	0.075 3±0.013 1	0.084 3±0.016 9	0.047 7±0.011 9
$U[-0.3,0.3]$	未添加	0.040 8±0.006 7	0.040 8±0.006 7	0.057 7±0.007 0	0.040 8±0.006 7	0.057 7±0.007 0
	添加	0.155 0±0.013 6	0.132 8±0.011 4	0.120 2±0.013 8	0.124 8±0.012 9	0.092 5±0.016 6

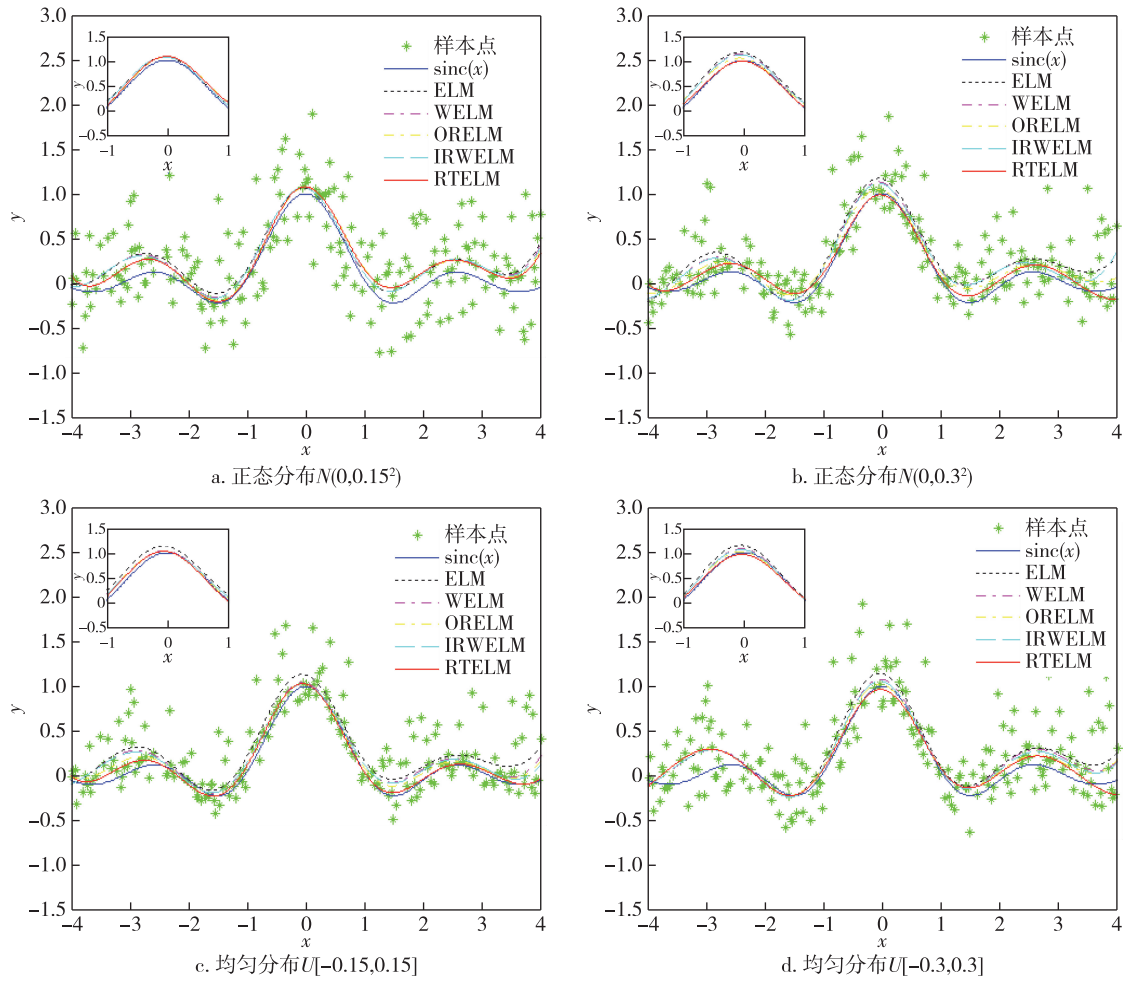


图 2 模拟数据集正态分布 $N(0,0.15^2)$ 、 $N(0,0.3^2)$ 和均匀分布 $U[-0.15,0.15]$ 、 $U[-0.3,0.3]$

噪声下,添加异常值情形下的回归曲线

Fig 2 Regression curves from simulated datasets corrupted by noises from normal distribution $N(0,0.15^2)$, $N(0,0.3^2)$, and uniform distribution $U[-0.15,0.15]$, $U[-0.3,0.3]$ with outliers

所提出的 RTELM 的有效性.在数据准备中,异常值的生成方式与模拟数据实验部分相同,且测试样本集不添加任何噪声.为了充分说明 RTELM 对添加了

异常值的数据集的学习性能,这里分别对异常值比例为 0%、5%、10%、15%、20%、30% 和 40% 的数据集进行数值实验.在训练每个数据集时,都进行了 10

次独立实验,实验的平均结果及算法排名如表 2 所示.表中用黑体表示最好的实验结果.

从表 2 可以看出,在 0% 异常值比例情况下,ELM 在数据集 Pollution、Triazines 和 BH 上的 RMSE 排名第一;在其他数据集上,这 5 种对比算法没有明显的差别.当训练样本包含异常值时,ELM 的预测精度总是最差的,这表明 ELM 对异常值较敏感.改进模型 WELM、ORELM 和 IRWELM 都能在一定程度上提升 ELM 的预测精度.在异常值比例为 5%、10% 和 15% 的情形下,在除 Pyrim、Autompg、Diabetes 和 Triazines 之外的数据集上,与其他算法相比较,RTELM 获得了最好的泛化性能.在其余的数据集上(除 Diabetes 和 Triazines 外),RTELM 仍然具有排名第二的

学习性能.当数据集包含较大异常值比例(20%、30% 和 40%) 时,RTELM 在除 Pollution、Autompg 和 MCPU 之外的数据集上都获得了排名第一的 RMSE 值,在 Autompg 和 MCPU 数据集上的学习性能排名第二.

同时,随着异常值比例的增加,ELM 和 WELM 的预测精度急剧下降,本文提出的 RTELM 具有相对稳定的更高的预测精度.这是因为 RTELM 采用截断 1-范数损失函数来限定异常值所造成的最大损失,在训练过程中异常值的权值设置为 0,使得异常值不参与最小化目标函数,从而异常值在训练过程中不起主要作用,达到抑制异常值影响的目的.

表 2 真实数据集的实验结果

Table 2 Experimental results on real-world datasets

数据集	噪声比例/%	ELM (RMSE±Std)	WELM (RMSE±Std)	ORELM (RMSE±Std)	IRWELM (RMSE±Std)	RTELM (RMSE±Std)
Bodyfat	0	0.002 6±0.001 9(5)	0.002 1±0.002 1(4)	0.002 0±0.002 1(2)	0.002 0±0.002 1(2)	0.002 0±0.002 1(2)
	5	0.018 5±0.001 9(5)	0.002 7±0.001 9(4)	0.002 1±0.002 1(2.5)	0.002 1±0.002 1(2.5)	0.002 0±0.002 2(1)
	10	0.018 7±0.002 2(5)	0.002 7±0.001 9(4)	0.002 1±0.002 1(2.5)	0.002 1±0.002 1(2.5)	0.002 0±0.002 1(1)
	15	0.020 1±0.001 9(5)	0.002 8±0.001 9(4)	0.002 2±0.002 1(3)	0.002 1±0.002 1(2)	0.002 0±0.002 2(1)
	20	0.019 4±0.002 6(5)	0.004 2±0.001 8(4)	0.002 2±0.002 1(2.5)	0.002 2±0.002 1(2.5)	0.002 0±0.002 1(1)
	30	0.020 1±0.002 2(4)	0.020 1±0.002 2(4)	0.002 5±0.001 9(2)	0.020 1±0.002 2(4)	0.002 0±0.002 1(1)
	40	0.020 7±0.003 9(4)	0.020 7±0.003 9(4)	0.007 0±0.002 4(2)	0.020 7±0.003 9(4)	0.002 0±0.002 1(1)
Pollution	0	35.503 2±6.679 7(1)	37.074 4±4.761 2(5)	35.999 4±6.138 0(2.5)	36.918 1±4.115 9(4)	35.999 4±6.138 0(2.5)
	5	56.334 5±6.961 9(5)	38.209 1±5.981 6(4)	37.892 6±5.968 7(3)	36.897 1±5.325 1(2)	36.722 9±6.415 2(1)
	10	58.605 7±7.377 6(5)	38.821 5±4.705 8(3)	39.442 7±6.743 0(4)	37.340 0±4.748 9(2)	36.345 8±6.652 4(1)
	15	57.542 9±8.224 8(5)	42.566 1±4.660 6(3)	43.229 7±6.388 8(4)	37.817 0±5.815 3(2)	36.586 5±5.936 5(1)
	20	60.851 4±7.864 2(5)	45.950 2±6.345 5(3)	46.700 3±7.371 8(4)	37.466 3±6.074 3(2)	37.065 9±6.274 2(1)
	30	58.398 8±8.496 5(3.5)	58.398 8±8.496 5(3.5)	56.891 4±7.954 7(1)	58.398 8±8.496 5(3.5)	58.398 8±8.496 5(3.5)
	40	64.714 5±10.242 7(3.5)	64.714 5±10.242 7(3.5)	61.330 3±7.829 2(1)	64.714 5±10.242 7(3.5)	64.714 5±10.242 7(3.5)
Servo	0	0.753 1±0.130 5(5)	0.680 6±0.164 0(1)	0.706 1±0.167 4(3)	0.714 7±0.151 2(4)	0.703 8±0.173 7(2)
	5	0.860 0±0.125 6(5)	0.727 8±0.181 9(3)	0.717 6±0.180 4(2)	0.752 0±0.157 3(4)	0.705 0±0.191 0(1)
	10	1.023 0±0.107 2(5)	0.834 5±0.106 9(4)	0.773 1±0.173 2(2)	0.808 9±0.188 1(3)	0.733 5±0.179 7(1)
	15	1.182 9±0.159 4(5)	0.958 0±0.115 1(4)	0.842 7±0.154 4(2)	0.867 4±0.126 0(3)	0.774 1±0.174 2(1)
	20	1.277 0±0.143 8(5)	1.084 2±0.160 7(4)	1.007 6±0.157 9(3)	0.912 0±0.153 7(1)	0.916 7±0.184 2(2)
	30	1.558 3±0.089 6(5)	1.344 8±0.114 3(4)	1.149 4±0.213 3(2)	1.281 4±0.123 0(3)	1.005 0±0.213 1(1)
	40	1.565 7±0.109 2(5)	1.564 1±0.082 4(4)	1.445 8±0.363 9(2)	1.563 6±0.082 7(3)	1.287 5±0.194 3(1)
Pyrim	0	0.112 3±0.020 9(5)	0.106 5±0.029 8(2)	0.109 0±0.026 6(4)	0.107 4±0.030 2(3)	0.105 5±0.027 5(1)
	5	0.115 9±0.022 9(5)	0.108 1±0.032 8(3)	0.109 7±0.029 2(4)	0.106 9±0.031 2(1)	0.107 3±0.029 5(2)
	10	0.136 0±0.021 4(5)	0.111 5±0.033 4(3)	0.111 9±0.030 2(4)	0.108 2±0.034 2(2)	0.105 7±0.031 3(1)
	15	0.139 3±0.027 1(5)	0.121 8±0.032 6(4)	0.117 4±0.034 1(3)	0.109 8±0.035 8(2)	0.108 7±0.032 3(1)
	20	0.138 9±0.027 1(5)	0.132 8±0.026 3(4)	0.119 3±0.035 8(3)	0.117 9±0.037 2(2)	0.113 6±0.034 5(1)
	30	0.154 8±0.009 8(4)	0.154 7±0.029 6(3)	0.135 4±0.041 5(2)	0.156 6±0.030 5(5)	0.113 5±0.037 9(1)
	40	0.158 8±0.018 8(3)	0.164 1±0.015 8(4)	0.138 8±0.019 2(2)	0.165 5±0.016 5(5)	0.119 5±0.040 0(1)

续表

数据集	噪声比例/%	ELM (RMSE±Std)	WELM (RMSE±Std)	ORELM (RMSE±Std)	IRWELM (RMSE±Std)	RTELM (RMSE±Std)
Autompg	0	2.693 3±0.237 2(2)	2.659 5±0.228 7(1)	2.719 4±0.246 1(3.5)	2.767 0±0.273 4(5)	2.719 4±0.246 1(3.5)
	5	3.514 2±0.230 3(5)	2.758 7±0.192 0(3)	2.745 9±0.272 0(2)	2.766 9±0.274 1(4)	2.724 1±0.248 1(1)
	10	4.348 4±0.299 7(5)	2.874 0±0.189 6(4)	2.817 5±0.201 6(3)	2.747 9±0.183 5(1)	2.748 0±0.229 6(2)
	15	5.462 9±0.368 8(5)	2.966 5±0.153 4(4)	2.906 2±0.144 3(3)	2.792 3±0.221 4(1)	2.857 8±0.191 6(2)
	20	6.116 9±0.335 6(5)	3.223 6±0.183 7(4)	2.950 7±0.144 4(3)	2.811 3±0.200 8(1)	2.821 5±0.192 3(2)
	30	8.017 4±0.331 1(5)	7.627 1±0.496 6(4)	3.317 7±0.372 3(2)	7.072 6±0.478 8(3)	2.919 2±0.203 2(1)
	40	8.292 9±0.302 8(5)	8.220 2±0.342 6(3.5)	4.762 9±0.736 7(2)	8.220 2±0.342 6(3.5)	2.936 2±0.195 2(1)
Diabetes	0	0.584 3±0.094 1(3)	0.582 6±0.090 9(2)	0.596 1±0.105 3(4)	0.582 5±0.090 7(1)	0.598 6±0.103 8(5)
	5	0.653 4±0.099 6(5)	0.576 4±0.094 1(2)	0.601 5±0.079 1(4)	0.574 3±0.089 2(1)	0.586 2±0.082 9(3)
	10	0.653 5±0.104 8(5)	0.596 4±0.106 5(3)	0.616 2±0.068 6(4)	0.579 1±0.098 4(1)	0.590 4±0.078 7(2)
	15	0.658 5±0.121 8(5)	0.587 8±0.106 7(2)	0.633 0±0.102 6(4)	0.580 7±0.098 7(1)	0.588 7±0.082 6(3)
	20	0.658 0±0.117 0(4)	0.688 1±0.088 4(5)	0.635 4±0.104 7(3)	0.625 4±0.099 2(2)	0.592 8±0.088 9(1)
	30	0.663 0±0.126 3(3)	0.676 1±0.121 3(4)	0.650 0±0.110 0(2)	0.687 7±0.128 8(5)	0.618 7±0.105 4(1)
	40	0.670 3±0.125 0(2)	0.673 3±0.125 3(3)	0.684 5±0.108 9(5)	0.673 9±0.125 4(4)	0.649 1±0.107 2(1)
Triazines	0	0.148 5±0.017 4(1)	0.149 8±0.019 9(2)	0.151 6±0.020 7(4)	0.150 6±0.021 4(3)	0.152 2±0.022 0(5)
	5	0.156 6±0.018 8(5)	0.149 0±0.020 6(1)	0.153 3±0.020 3(4)	0.150 5±0.020 8(2)	0.152 6±0.020 9(3)
	10	0.159 7±0.017 0(5)	0.150 6±0.020 0(1)	0.154 6±0.021 1(4)	0.151 2±0.020 9(2)	0.154 0±0.021 3(3)
	15	0.160 2±0.015 0(5)	0.152 5±0.020 6(2)	0.157 3±0.021 2(4)	0.151 5±0.020 5(1)	0.154 9±0.021 3(3)
	20	0.160 5±0.014 9(5)	0.159 1±0.020 7(4)	0.158 5±0.016 7(3)	0.154 1±0.017 2(1)	0.157 9±0.017 9(2)
	30	0.164 8±0.012 8(5)	0.161 0±0.018 4(4)	0.159 7±0.016 2(2)	0.160 1±0.017 0(3)	0.158 3±0.017 2(1)
	40	0.161 2±0.015 3(5)	0.167 2±0.019 8(5)	0.160 0±0.017 9(2)	0.165 7±0.018 8(4)	0.159 5±0.017 1(1)
MCPU	0	56.701 0±27.139 9(5)	56.144 2±25.647 6(4)	53.447 3±28.929 9(3)	52.409 7±21.207 8(1)	53.374 2±25.260 9(2)
	5	78.343 5±25.097 0(5)	53.410 2±22.587 0(3)	54.912 5±23.467 4(4)	53.380 7±23.142 2(2)	51.883 8±22.033 6(1)
	10	114.634 2±7.792 4(5)	61.273 8±23.176 8(4)	55.579 3±19.952 9(3)	55.119 0±24.354 7(2)	53.241 8±21.519 6(1)
	15	116.736 1±16.258 2(5)	68.448 7±23.762 2(4)	60.350 1±19.825 9(3)	58.897 4±25.131 8(2)	56.086 7±20.542 1(1)
	20	145.811 7±10.594 2(5)	74.893 4±14.385 1(4)	68.324 3±14.176 4(3)	64.563 3±23.607 5(1)	67.163 6±20.515 5(2)
	30	158.828 6±48.695 4(5)	118.806 8±18.448 4(4)	74.113 8±14.195 8(2)	91.813 1±14.142 1(3)	72.008 3±16.061 0(1)
	40	158.181 1±46.562 9(4)	159.421 6±44.142 6(5)	89.463 1±20.981 3(2)	137.959 9±23.596 9(3)	77.590 0±33.576 4(1)
BH	0	3.290 5±0.247 0(1)	3.374 2±0.433 5(2)	3.403 0±0.414 6(3)	3.447 3±0.461 3(5)	3.417 5±0.425 4(4)
	5	4.390 5±0.421 5(5)	3.475 0±0.432 7(2)	3.476 6±0.447 3(3)	3.503 5±0.556 4(4)	3.468 7±0.457 0(1)
	10	5.181 9±0.361 9(5)	3.701 7±0.579 7(4)	3.634 5±0.521 1(3)	3.600 5±0.678 3(2)	3.564 7±0.502 6(1)
	15	6.412 6±0.526 0(5)	3.891 2±0.572 2(4)	3.766 4±0.551 8(3)	3.645 1±0.454 3(2)	3.636 1±0.569 3(1)
	20	7.650 2±0.297 3(5)	4.530 8±0.546 4(4)	4.018 3±0.532 9(3)	3.958 3±0.892 9(2)	3.771 8±0.697 6(1)
	30	8.996 1±0.469 2(5)	8.230 8±0.542 7(4)	4.890 5±0.738 8(2)	7.004 0±0.647 2(3)	4.180 4±0.660 7(1)
	40	9.189 6±0.385 4(5)	9.109 0±0.399 0(3.5)	6.557 2±0.384 2(2)	9.109 0±0.399 0(3.5)	4.578 1±1.039 2(1)
平均排名		4.460 3	3.444 4	2.865 1	2.611 1	1.634 9

4 结束语

在许多实际应用中,数据集通常包含噪声和异常值.ELM 在这类数据集上的泛化性能和鲁棒性较差.为了抑制噪声和异常值的负面影响,提升 ELM 的鲁棒性,本文提出基于截断 1-范数损失函数的鲁棒 ELM 模型,利用迭代重赋权算法求解相应的优化问

题.在训练过程中,每一步迭代都将异常值的权值设置为 0.因此,异常值不参与最小化目标函数的过程,从而能够提升 ELM 的鲁棒性.4 个模拟数据集和 9 个真实数据集的实验结果表明,RTELM 获得了更优和相对稳定的泛化能力和鲁棒性,尤其是在异常值比例较大的情形下.

参考文献

References

- [1] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine; a new learning scheme of feedforward neural networks [C] // 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. no. 04CH37541), Budapest, Hungary, 2004: 985-990
- [2] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine; theory and applications [J]. Neurocomputing, 2006, 70(1/2/3): 489-501
- [3] Huang G B, Chen L, Siew C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes [J]. IEEE Transactions on Neural Networks, 2006, 17(4): 879-892
- [4] Huang G B, Zhou H M, Ding X J, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part b(Cybernetics), 2012, 42(2): 513-529
- [5] 姜琳颖, 余东海, 石鑫. 基于加权极限学习机的肿瘤基因表达谱数据分类 [J]. 东北大学学报(自然科学版), 2017, 38(6): 798-803
JIANG Linying, YU Donghai, SHI Xin. Tumor microarray gene expression data classification based on weighted extreme learning machine [J]. Journal of Northeastern University(Natural Science), 2017, 38(6): 798-803
- [6] 柯道, 邹嘉伟, 杜明智, 等. 基于蒙特卡罗数据集均衡与鲁棒性增量极限学习机的图像自动标注 [J]. 电子学报, 2017, 45(12): 2925-2935
KE Xiao, ZOU Jiawei, DU Mingzhi, et al. The automatic image annotation based on monte-carlo data set balance and robustness incremental extreme learning machine [J]. Acta Electronica Sinica, 2017, 45(12): 2925-2935
- [7] Li Q, Liu Y, Wang S J, et al. Image classification using low-rank regularized extreme learning machine [J]. IEEE Access, 2019, 7: 877-883
- [8] Hampel F R, Ronchetti E M, Rousseeuw P J, et al. Robust Statistics [M]. Wiley, 2005. DOI: 10. 1002/9781118186435
- [9] Frenay B, Verleysen M. Classification in the presence of label noise: a survey [J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(5): 845-869
- [10] Deng W N, Zheng Q H, Chen L. Regularized extreme learning machine [C] // 2009 IEEE Symposium on Computational Intelligence and Data Mining, Nashville, TN, USA, 2009: 389-395
- [11] Zhang K, Luo M X. Outlier-robust extreme learning machine for regression problems [J]. Neurocomputing, 2015, 151: 1519-1527
- [12] Horata P, Chiewchanwattana S, Sunat K. Robust extreme learning machine [J]. Neurocomputing, 2013, 102: 31-44
- [13] Chen K, Lv Q, Lu Y, et al. Robust regularized extreme learning machine for regression using iteratively reweighted least squares [J]. Neurocomputing, 2017, 230: 345-358
- [14] Xing H J, Wang X M. Training extreme learning machine via regularized correntropy criterion [J]. Neural Computing and Applications, 2013, 23(7/8): 1977-1986
- [15] Ren Z, Yang L M. Robust extreme learning machines with different loss functions [J]. Neural Processing Letters, 2019, 49(3): 1543-1565
- [16] Zhong P. Training robust support vector regression with smooth non-convex loss function [J]. Optimization Methods and Software, 2012, 27(6): 1039-1058

Robust ELM model with truncated 1-norm loss function

WANG Kuaini^{1,2} CAO Jinde²

1 College of Science, Xi'an Shiyou University, Xi'an 710065

2 School of Mathematics, Southeast University, Nanjing 211189

Abstract Sensitivity to noises and outliers and inferior robustness are the primary problems associated with extreme learning machine(ELM). Based on the 1-norm loss function, a truncated 1-norm loss function is proposed to suppress the effects of noises and outliers. A robust ELM model with truncated 1-norm loss function is established. The corresponding optimization problem is solved by iterative re-weighted algorithm. Four simulated data sets and nine real-world data sets are used to verify the validity of the proposed model. The numerical results show that the generalization performance of robust ELM in noisy environment is superior to that of the compared methods and has superior robustness, especially in the case of a substantial proportion of outliers.

Key words neural network; extreme learning machine; robustness; truncated loss function; outliers