

基于L1正则化和相关熵的极限学习机及应用研究

数学2102班：周子豪 指导教师：王快妮 论文类型：毕业论文

摘要：本文应用理论分析、构建模型、实验对比分析多种研究方法，探讨了传统极限学习机存在的冗余节点、噪声敏感及高维小样本过拟合问题，并针对这些问题，提出融合 L1 正则化和相关熵损失函数的改进极限学习机模型。首先，分析了 L1 正则化的稀疏性原理和高斯核相关熵的抗噪机制；其次，构建了融合 L1 正则化与相关熵的改进极限学习机模型，并设计迭代重加权算法求解闭式解；最后，在基准数据集和真实台湾省房地产价格预估数据集中验证模型性能，得出新模型在均方根误差和稀疏性上优于传统方法的结论，并提出模型在动态场景中参数自适应调节的对策建议。

关键词：L1正则化；相关熵；极限学习机；稀疏性；鲁棒性

1 研究背景

随着计算机科学、神经科学及大数据技术的发展，人工智能研究蓬勃兴起。机器学习作为其核心分支，通过数据收集、模型训练、评估和推理实现预测与决策，在图像识别、语音处理等领域应用广泛。传统机器学习需人工设计特征，深度学习虽能自动提取特征但训练复杂、计算成本高。

新加坡南洋理工大学 Huang 教授提出的极限学习机(Extreme Learning Machine, ELM)是改进的单层前馈神经网络，其随机初始化隐层权重并解析求解输出权重，训练效率较传统反向传播算法提升数百倍，泛化能力良好，已成功应用于目标识别、图像分类等领域。然而，ELM 存在三大瓶颈：隐层权重随机初始化导致冗余节点，增加存储与计算成本；基于均方误差的损失函数对噪声及异常值敏感；缺乏正则化机制，高维或小样本场景易过拟合。

为此，本课题基于 L1 正则化与相关熵损失函数改进 ELM：利用 L1 正则化筛选特征、简化模型，通过调节参数削减冗余节点；以相关熵替换均方误差损失函数，通过调整参数分配权重抑制异常值；优化 L1 正则化与相关熵参数组合，提升模型鲁棒性。

2 发展现状

极限学习机作为高效单隐层前馈神经网络，在回归与分类任务中优势显著，但存在隐层数量依赖经验设定、随机权重导致冗余节点堆积、高维小样本易过拟合，以及传统损失函数对噪声敏感等问题。由此国内外学者从稀疏性与鲁棒性展开改进：

稀疏性研究中，L1 正则化被引入 ELM 目标函数以消除冗余节点。ADMM 算法

解决非凸收敛难题。混合正则化(如弹性网络)调控复杂度。L0 范数近似实现隐层神经元自动选择。鲁棒性研究通过融入 Huber、Welsch 等 M-估计函数，结合迭代重加权动态调整异常样本权重，提升模型抗噪性。

近年联合稀疏与鲁棒性的 ELM 成为前沿，如 L21 范数正则化结合 Huber 损失函数，但参数自适应机制与多目标优化仍待完善。未来需融合 L1 正则化与相关熵设计自适应方案，以适配工业实时数据流场景。

3 相关分析

3.1 L1正则化与相关熵损失函数

L1 正则化(Lasso 正则化)是一种通过向目标函数中添加权重参数的 L1 范数来实现模型稀疏化的技术。其数学形式定义为：

$$\mathfrak{R}(\beta) = \lambda \|\beta\|_1 = \lambda \sum_{j=1}^L |\beta_j| \quad (3.1)$$

其中 $\beta = [\beta_1, \beta_2, \dots, \beta_L]^T$ 为模型的输出权重向量， $\lambda > 0$ 为正则化参数，用于控制稀疏化的程度，L1 正则化图像如图 3.1 所示。

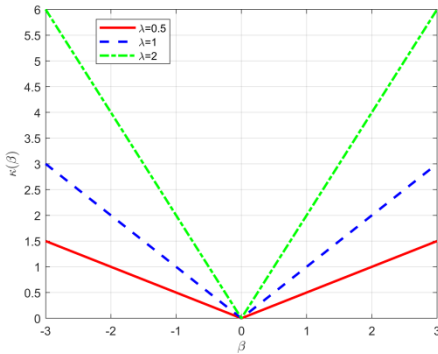


图 3.1 L1 正则化图像

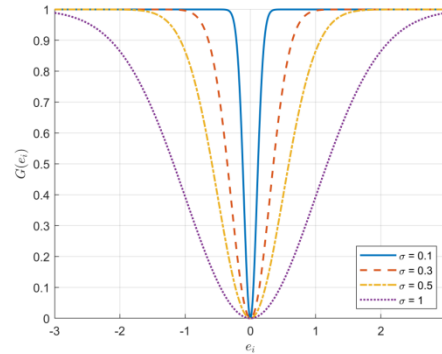


图 3.2 相关熵损失函数图像

从图 3.1 可见， λ 越大斜率越大、惩罚力度越强。L1 正则化的核心作用是稀疏诱导，即通过惩罚权重参数绝对值，使部分权重在优化中趋近于 0，进而自动筛选对模型贡献大的隐层节点，剔除影响小的节点。

在新模型中引入由高斯核所组成的相关熵损失函数：

$$G(e_i) = 1 - \exp\left(-\frac{e_i^2}{2\sigma^2}\right) \quad (3.2)$$

其中 $e_i = y_i - h(x_i)\beta$ 为预测误差，以及相关熵损失函数的图像图 3.2。由图可见，曲线均关于原点对称且满足偶函数性质，表明损失函数对正负误差惩罚等价。核宽参数 σ 决定曲线陡峭程度，该损失函数具有有界性和非凸性：误差趋于无穷时损失趋近 1，异常值不影响优化；非凸特性可降低局部极小值影响，增强全局搜索能力。

3.2 基于L1正则化和相关熵的极限学习机

将相关熵损失函数同 L1 正则化结合得到新模型 L1G-ELM，优化问题可以表达为：

$$\begin{aligned} \min_{\beta} \quad & C \sum_{i=1}^N (1 - \exp(-\frac{e_i^2}{2\sigma^2})) + \|\beta\|_1 \\ \text{s.t.} \quad & y_i - h(x_i)\beta = e_i, i=1,2,\dots,N \end{aligned} \quad (3.3)$$

针对该优化问题构造拉格朗日函数并使用KKT条件进行求解，可得

$$\beta = \begin{cases} \left(H^T W_N H + \frac{\sigma^2}{C} W_L \right)^{-1} H^T W_L Y, & N \geq L \\ W_L^{-1} H^T \left(H W_L^{-1} H^T + \frac{\sigma^2}{C} W_N^{-1} \right)^{-1} Y, & N < L \end{cases} \quad (3.4)$$

其中 $W_N = \text{diag}\{W_N(e_1), W_N(e_2), \dots, W_N(e_N)\}$ ， $W_L = \text{diag}\{W_L(\beta_1), W_L(\beta_2), \dots, W_L(\beta_L)\}$ ，且有

$$W_N(e_i) = \exp(-\frac{e_i^2}{2\sigma^2}) \quad W_L(\beta_j) = \frac{1}{\max\{|\beta_j|, 10^{-6}\}} \quad (3.5)$$

3.3 实验设计

实验中本研究选取 10 个 UCI 基准数据集(如 Adm、ALE、Autompg 等)开展实验。该 10 个数据集涵盖 7 至 60 维特征，样本规模从 40 至 414 不等；而应用数据集采用台湾省房地产价格数据集(REVDS)开展实验，REVDS 数据集包含 414 条年台湾新北市住宅交易记录，含交易日期、房龄、地铁站距离等 6 维特征及房价目标值。所有数据集按统一比例划分训练集与测试集，并通过在输出值区间随机采样添加 0%、10%、20%比例的异常值以模拟噪声场景。实验参数设置如下：L1 正则化参数 C 取值范围为 $\{2^{-19}, 2^{-18}, \dots, 2^{19}, 2^{20}\}$ ；相关熵核宽 σ 的取值范围为 $\{0.001, 0.003, \dots, 0.9, 1.0\}$ ；激活函数选用 sigmoid 函数。迭代过程中最大迭代次数设为 200 次，收敛阈值为 10^{-6} ，每组实验独立运行 10 次取均值以降低随机性误差。

对比模型包括正则化极限学习机(ELM)、加权正则化极限学习机(WELM)、迭代重加权的鲁棒正则化极限学习机(RELM)采用均方根误差与稀疏度作为评价指标。其中均方根误差用于衡量预测精度；稀疏度通过计算输出权重中零元素比例反映模型复杂度，二者共同验证改进模型的泛化能力与鲁棒性。

3.4 结果与分析

在 10 个基准数据集上，L1G-ELM 在 0%、10%、20%噪声场景中分别于 5 个、7 个、9 个数据集表现出最优的均方根误差。例如，Adm 数据集在 20%噪声下，L1G-ELM 的 RMSE 为 0.0661，显著低于 ELM 的 0.1384；稀疏性方面，L1G-ELM 的稀疏

度普遍处于 0.8~0.9 之间，如 Adm 数据集(20%噪声)稀疏度达较 RELM 提升显著。在台湾省房地产数据集 REVDs 上，L1G-ELM 在 0%、10%、20% 噪声下的 RMSE 分别低于对比模型，稀疏度也维持在 0.8~0.9 之间，显著高于 RELM。参数影响上，高维数据集(如 Triazines)最优 L1 正则化参数 C 更小，低维数据集(如 Diabetes)最优 C 更大；相关熵参数 σ 在高维数据中倾向小值(如 Triazines)，低维数据中倾向大值(如 Diabetes)，且噪声增加时 C 和 σ 的最优值均逐渐减小。因此相较于对比模型，L1G-ELM 更适用于含噪高维的房地产数据场景。

4 研究结论及对策建议

4.1 研究结论

在极限学习机模型中，正则化技术与损失函数选取对提升泛化能力至关重要。本文以L1正则化防止模型过拟合，采用相关熵损失函数增强鲁棒性，探究二者协同机制对综合泛化性能的改善。实验表明，该方法通过迭代重权策略调整输出权重，利用相关熵损失函数按预测误差分配节点权重，减弱异常值影响，使L1G-ELM在不同回归数据集上表现出良好泛化稳定性。与传统ELM相比，L1G-ELM通过调节相关熵参数控制模型覆盖范围，小参数聚焦局部关键特征抑制异常值，大参数平衡全局噪声鲁棒性与拟合能力，结合L1正则化系数实现高维含噪空间中稀疏性与预测精度的协同优化，优势显著优于仅含L2正则化的RELM及其他对比模型。

4.2 对策建议

未来研究可聚焦参数自适应调节和动态场景适配。针对当前参数依赖网格搜索、算力消耗大的问题，设计基于数据稀疏性动态变化的反馈系统或噪声强度推测机制，实现 C 与 σ 的自动协同优化，提升模型在动态数据流场景中的适用性。此外，可进一步探索该模型在其他领域的应用，拓展其工程实践价值。