# 摘 要

深度学习神经网络包含多个非线性隐藏层，这些结构使得该神经网络具备学习复杂模型的能力。然而由于训练数据缺失，使得模型参数的学习过分依赖于训练数据中的噪声，从而产生过拟合现象。解决过拟合问题的方法主要有：L1、L2正则化方法和最近提出的DROPOUT方法。

本文讨论了L1、L2正则化方法，以及超参数对两者正则化效果的影响。L1正则化通过稀疏化权重、L2正则化通过缩小权重，从而减小了模型拟合各种函数的能力来减弱过拟合现象。DROPOUT是近些年提出的一种新的正则化方法，该方法通过“模型的平均”和减小神经元之间的共适应性从而弱化过拟合的现象。

最后，本门通过构建784-1000-500-10的全连接神经网络，进行MNIST手写体识别实验，并分别运用L1、L2正则化和DROPOUT对该网络进行正则化处理。实验表明：虽然L1、L2正则化分别能将训练误差和测试误差的差距缩小到0.8%、0.7%，但是由于其相对DROPOUT正则化效果较为振荡，二者的综合效果不如DROPOUT。

**关键字：过拟合；正则化；正则化；DROPOUT；MNIST**

# Abstract

Deep neural networks contain multiple non-linear hidden layers and this makes them very expressive models that can learn very complicated relationships between their inputs and outputs. With limited data, however, many of these complicated relationships will be the result of sampling noise. This leads to overfitting and many methods have been developed for reducing it such as L1, L2 regularization methods and the recently proposed dropout method.

This paper discusses the L1 and L2 regularization methods and the hyper-parameters effects. L1 regularization method penalizes large values of the weights and therefore yields sparse weights. L2 regularization method reduces the weights and therefore reducing overfitting. The newly proposed Dropout method regularizes a model by averaging the predictions of all possible settings of the parameters and this reduces the co-adaptation between the neurons and therefore reduces the overfitting.

Finally, the MNIST handwriting recognition experiment was carried out by constructing a 784-1000-500-10 fully connected neural network, L1, L2 and dropout method were used to regularize the network. Experiments show that L1 and L2 can reduce the discrepancy between the training and test errors to 0.7% approximately. The dropout method is more stable than the previous two methods.

**Key Words: Overfitting; L1 Regularization; L2 Regularization; Dropout; MNIST**