# 第6章 梯度下降算法及变体： 高效求解模型参数

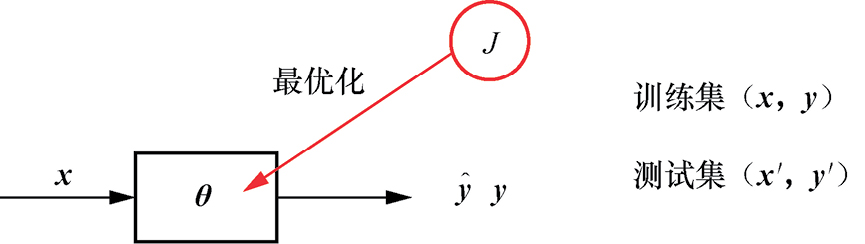


图6-1 深度学习中的最优化示意

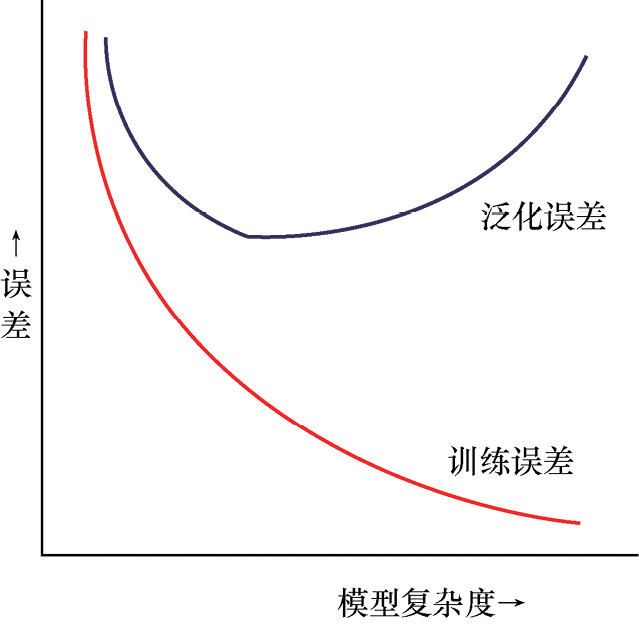


图6-2 模型复杂度与误差间的关系示意

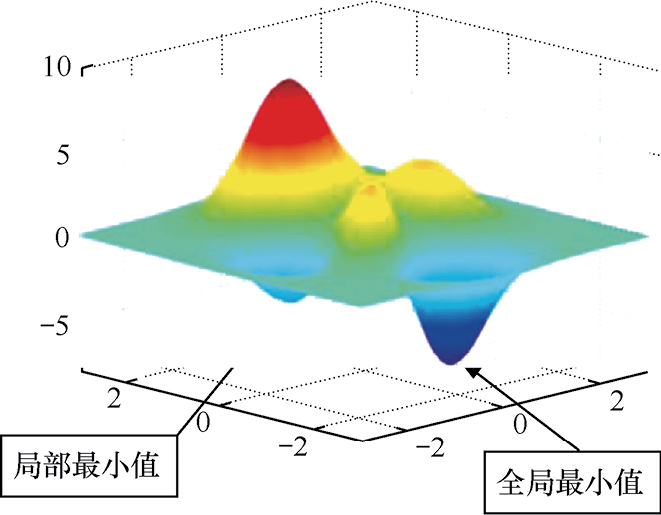


图6-3 局部最小值问题示意

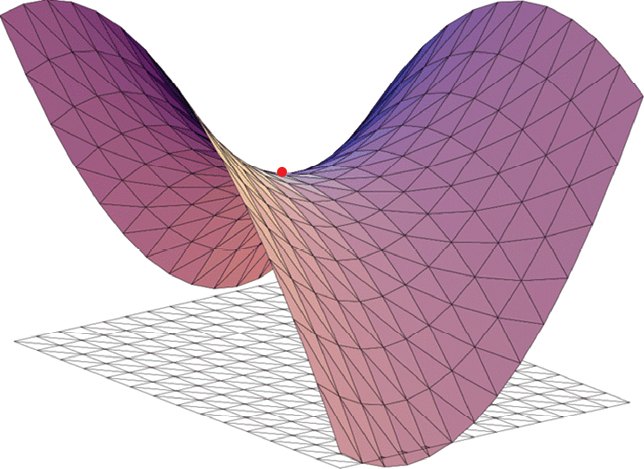


图6-4 鞍点问题示意



图6-5 悬崖问题示意

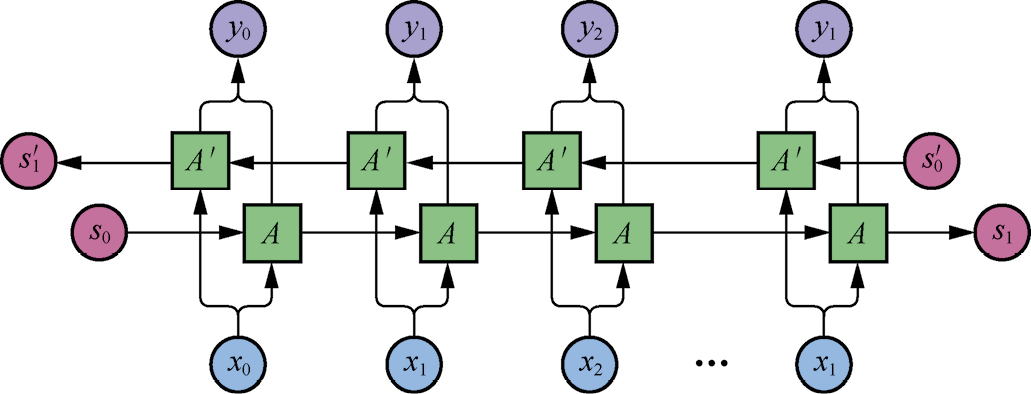


图6-6 长期依赖问题示意

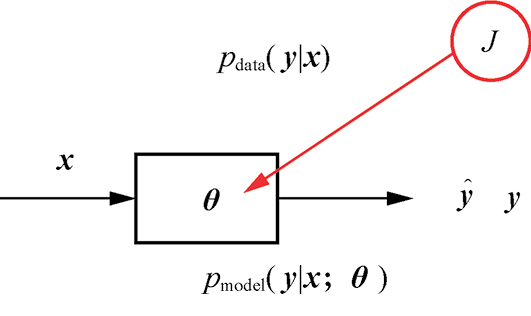


图6-7 从概率分布理解损失函数

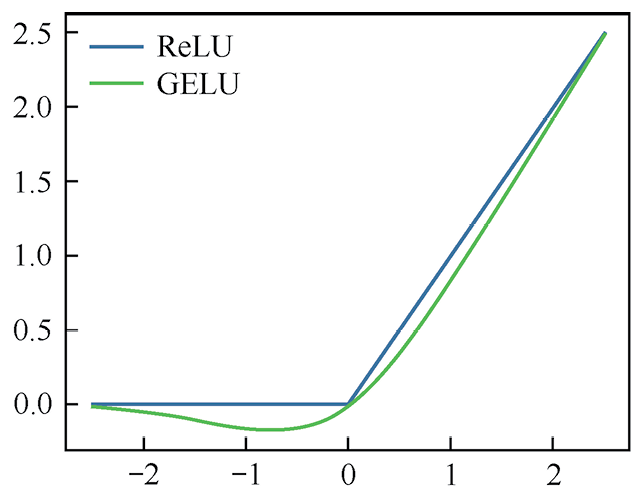


图6-8 改进版ReLU函数示意

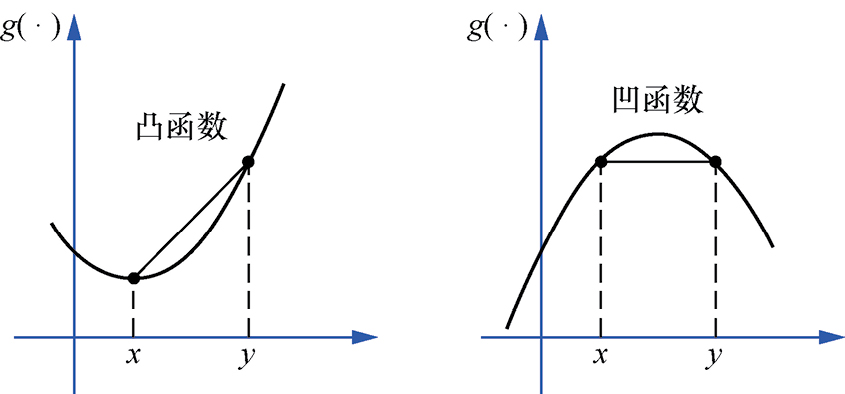


图6-9 凸函数和凹函数示意

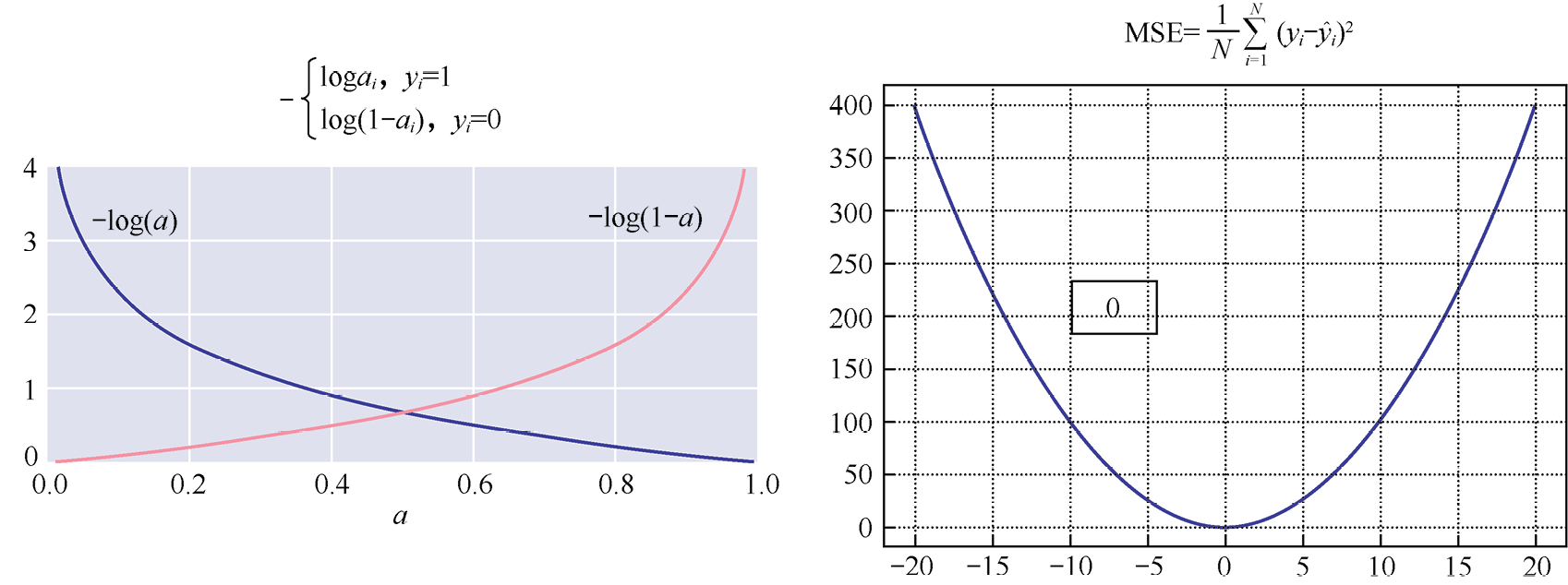


图6-10 对数损失函数和均方误差损失函数示意

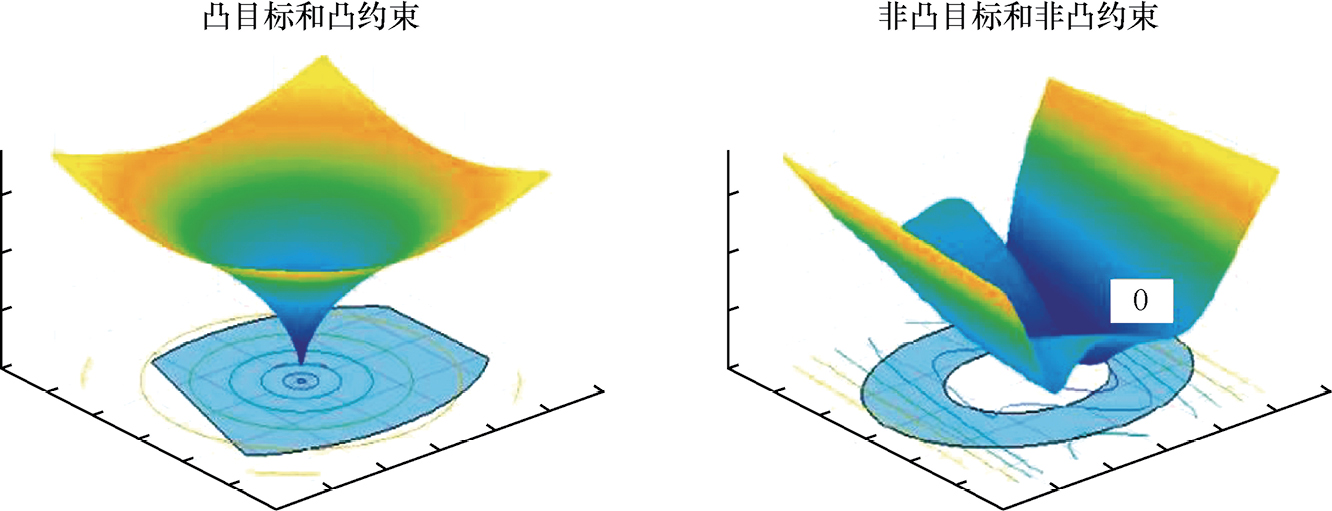


图6-11 凸约束和非凸约束示意



图6-12 凸函数与Jensen不等式示意

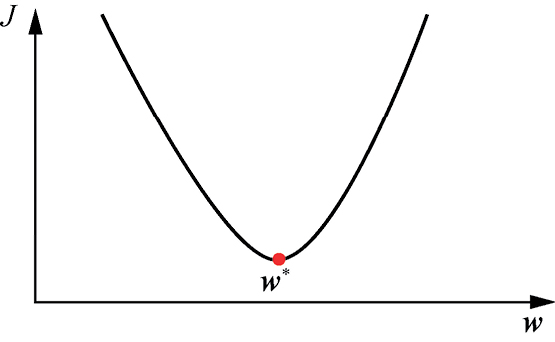


图6-13 一元二次损失函数示意

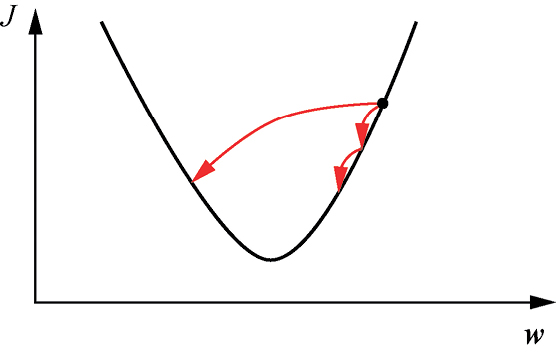


图6-14 学习率效果示意

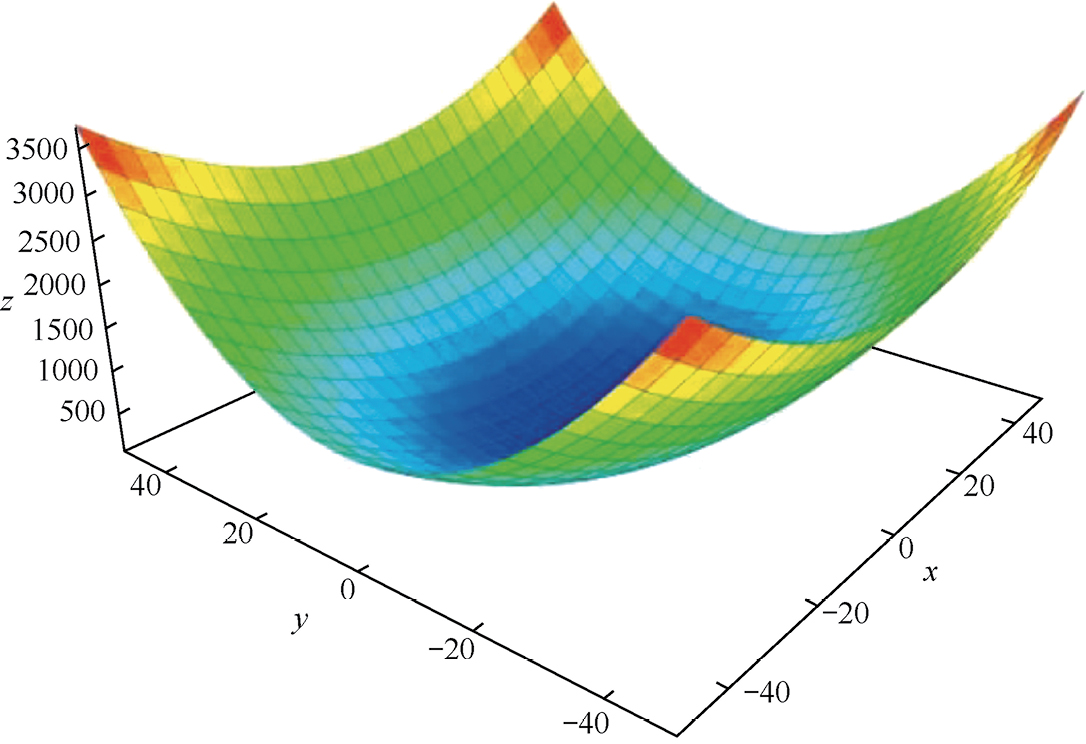


图6-15 二元损失函数示意

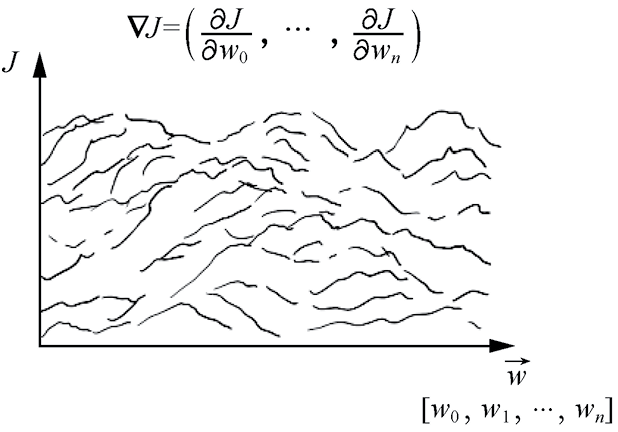


图6-16 参数的偏导数直观示意

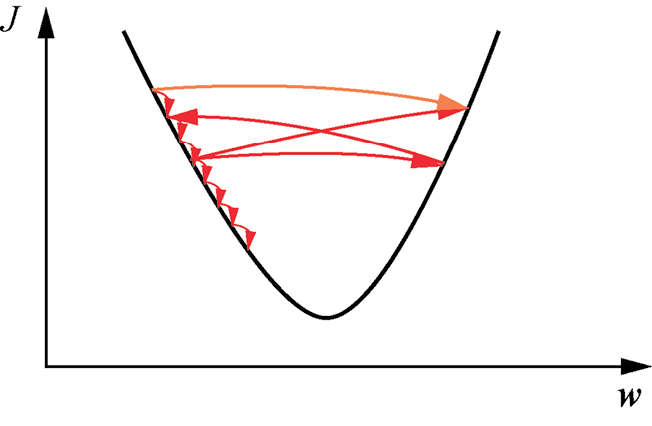


图6-17 学习率效果示意

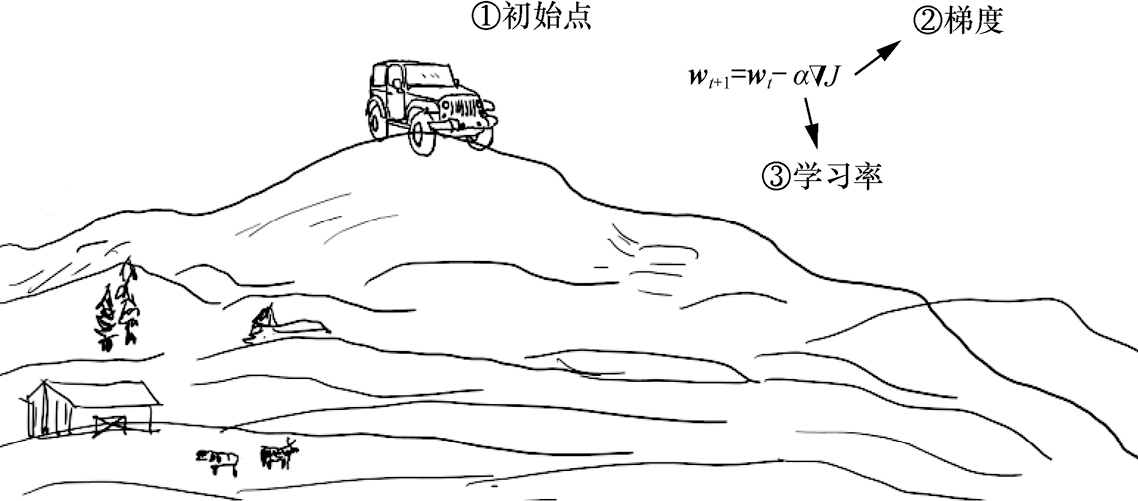


图6-18 梯度下降算法直观示意

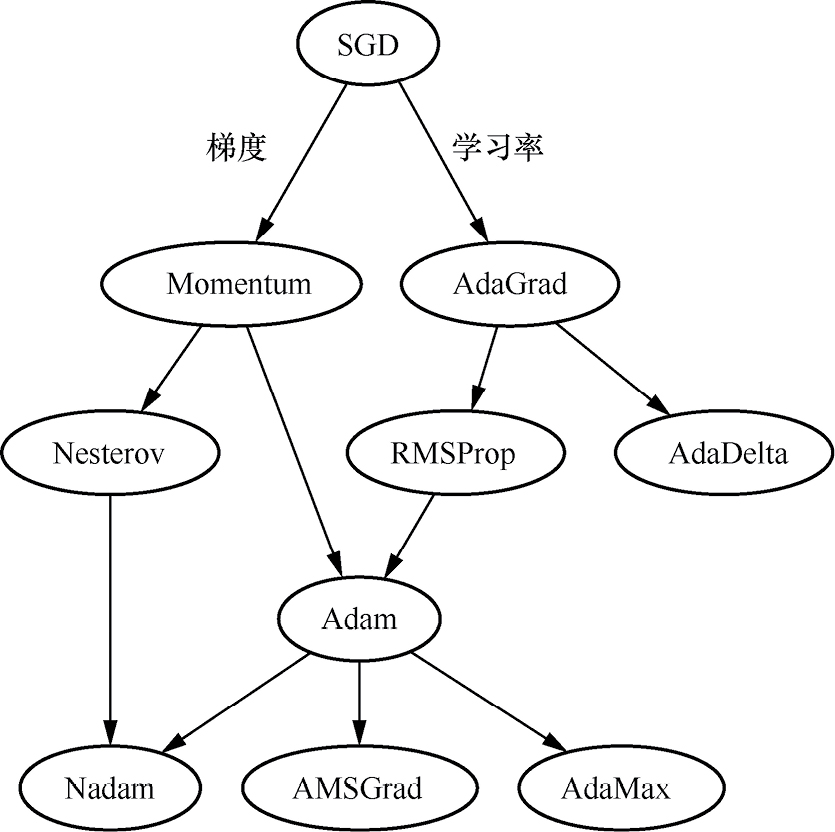


图6-19 不同优化器的相互关系

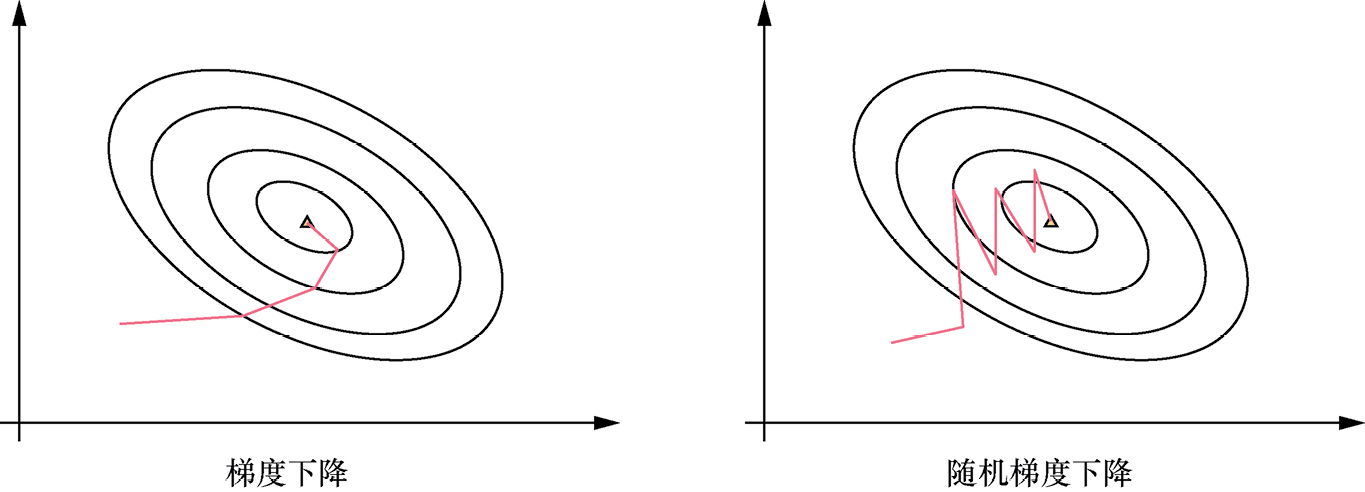


图6-20 随机梯度下降示意

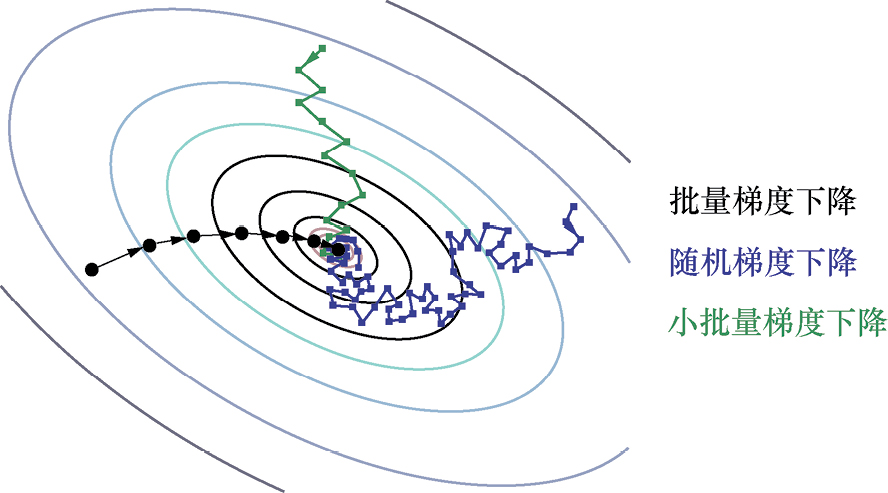


图6-21 小批量梯度下降示意

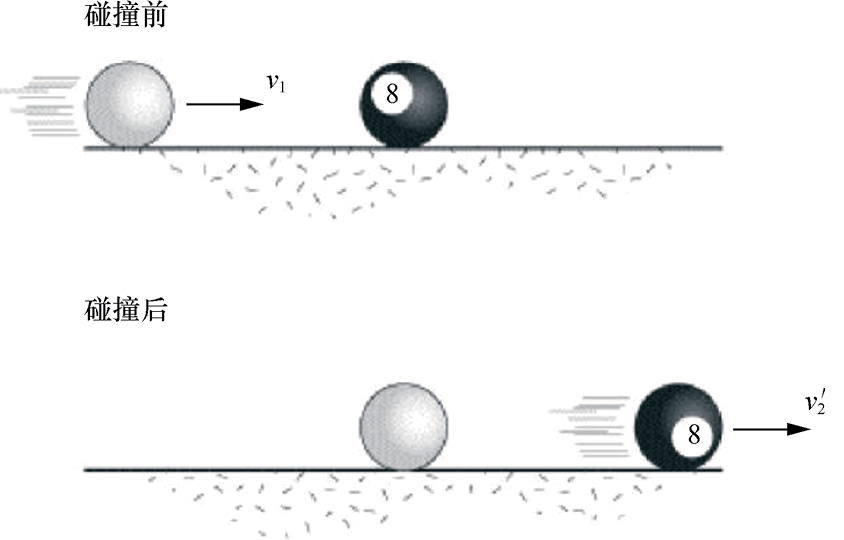


图6-22 动量示意

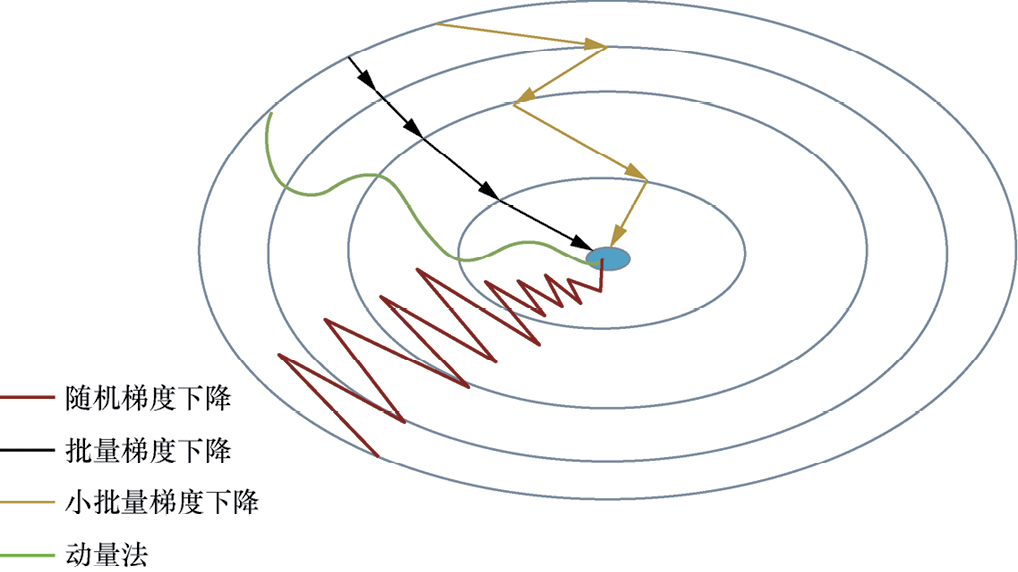


图6-23 4种梯度下降算法的效果对比示意

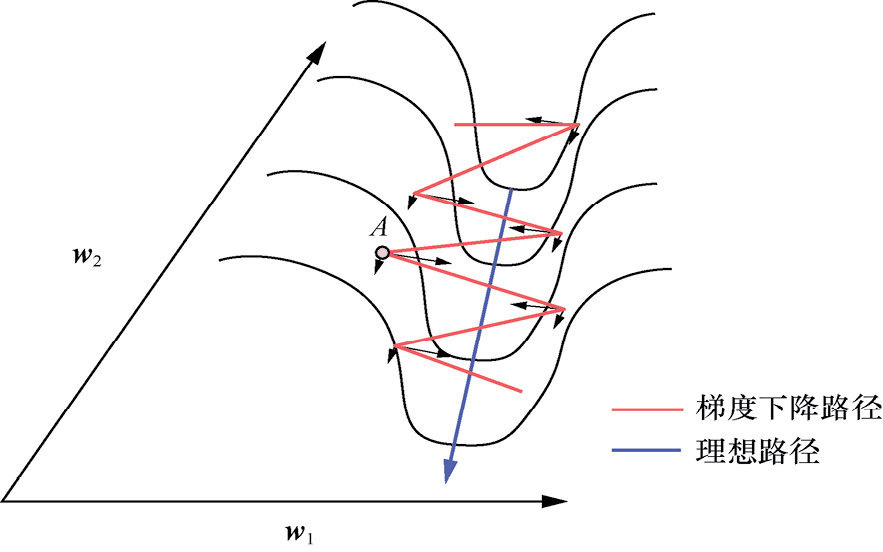


图6-24 病态收敛示意

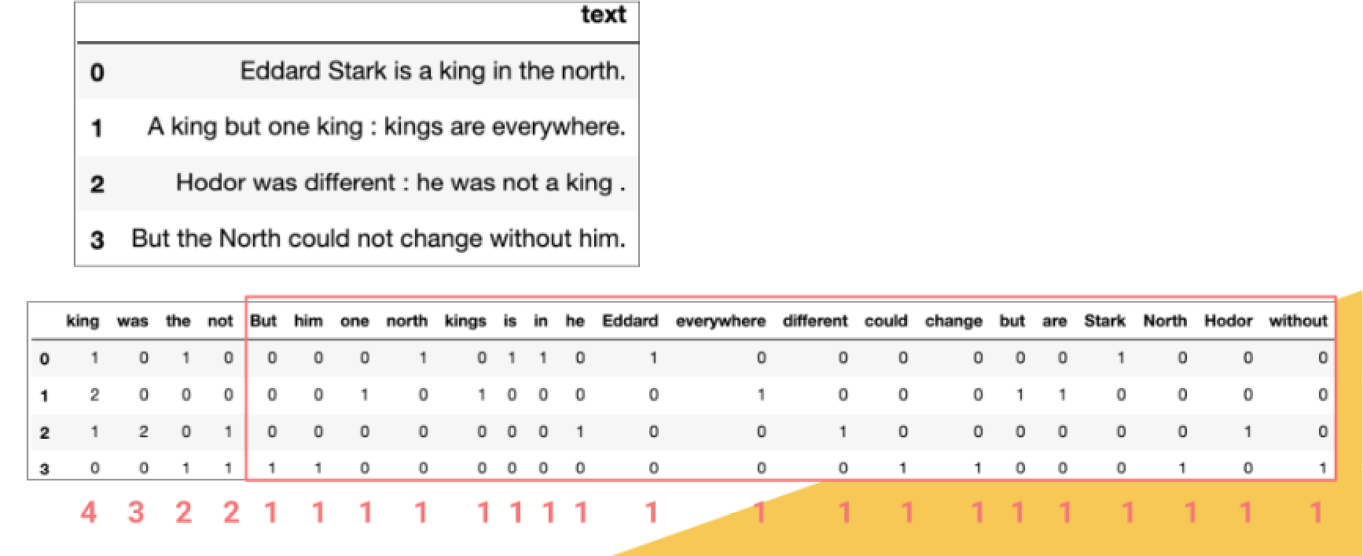


图6-25 文本分类问题示意

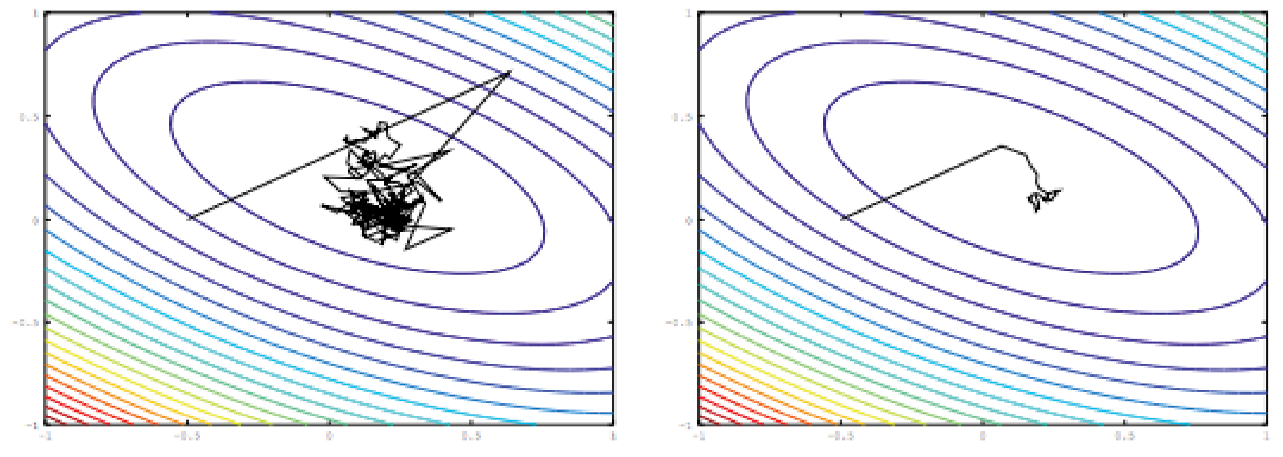


图6-26 SGD与AdaGrad效果对比示意

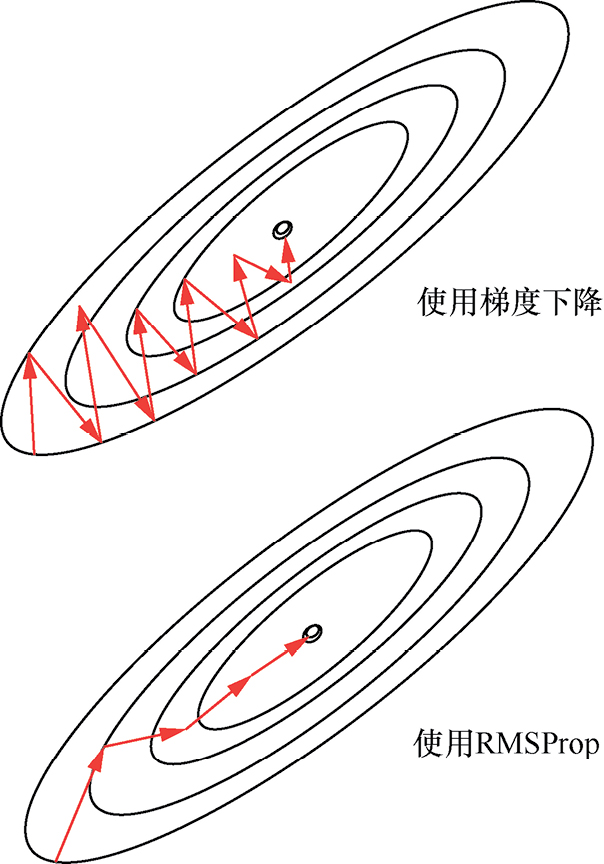


图6-27 RMSProp效果示意

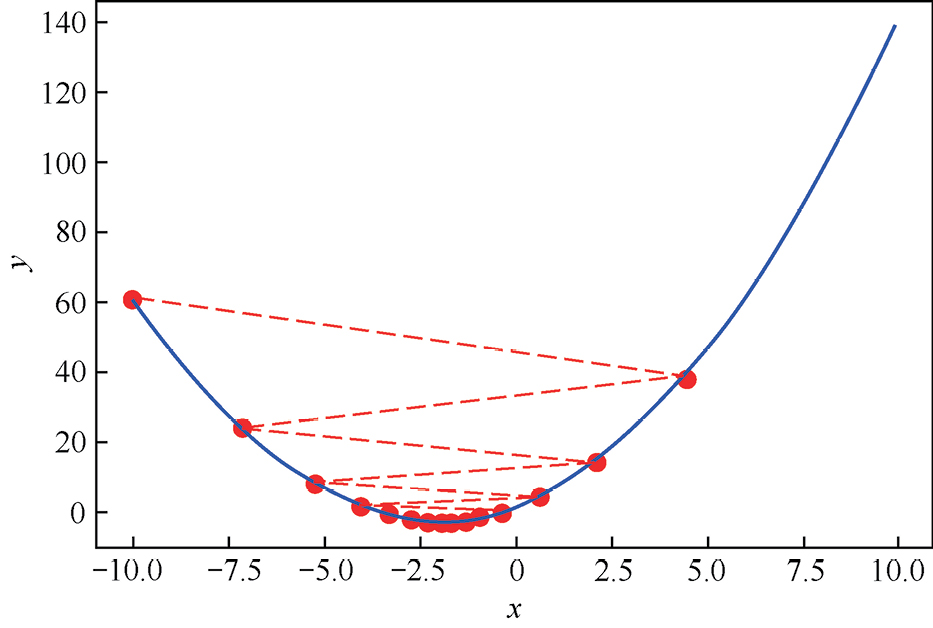


图6-28 梯度下降可视化

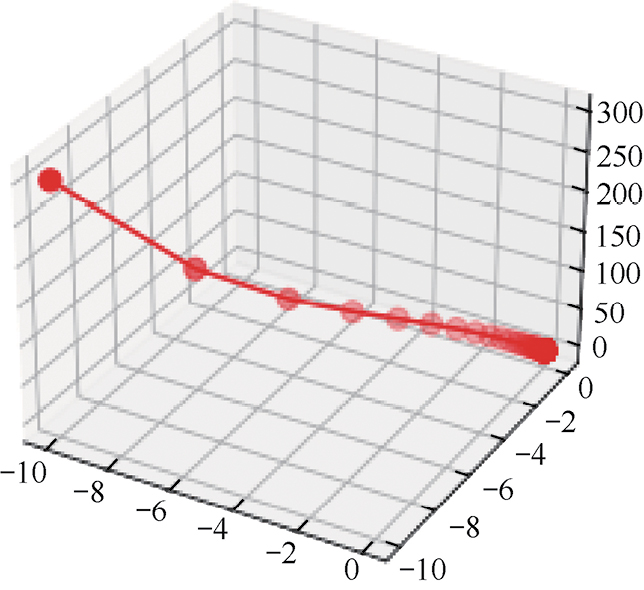


图6-29 梯度下降三维可视化

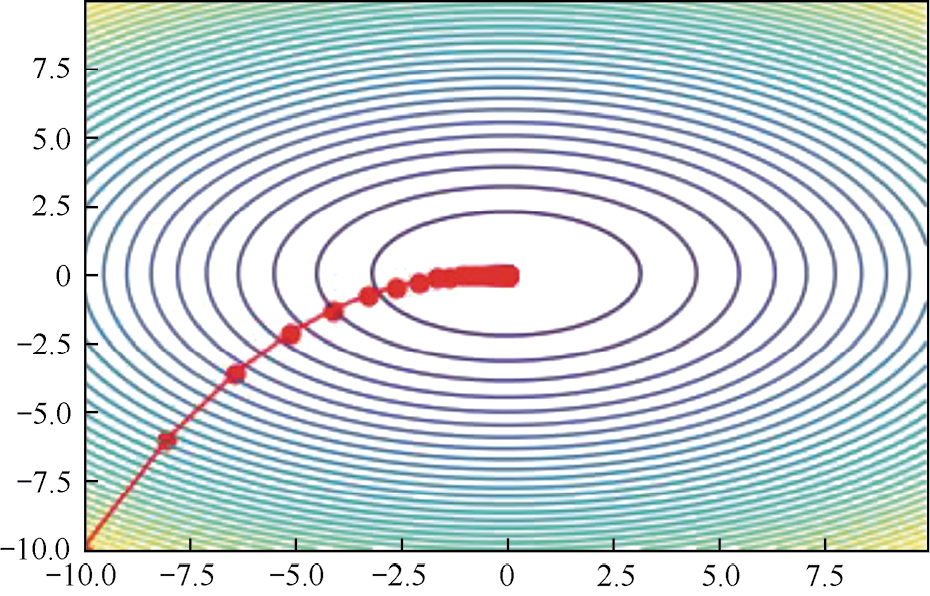


图6-30 梯度下降图像

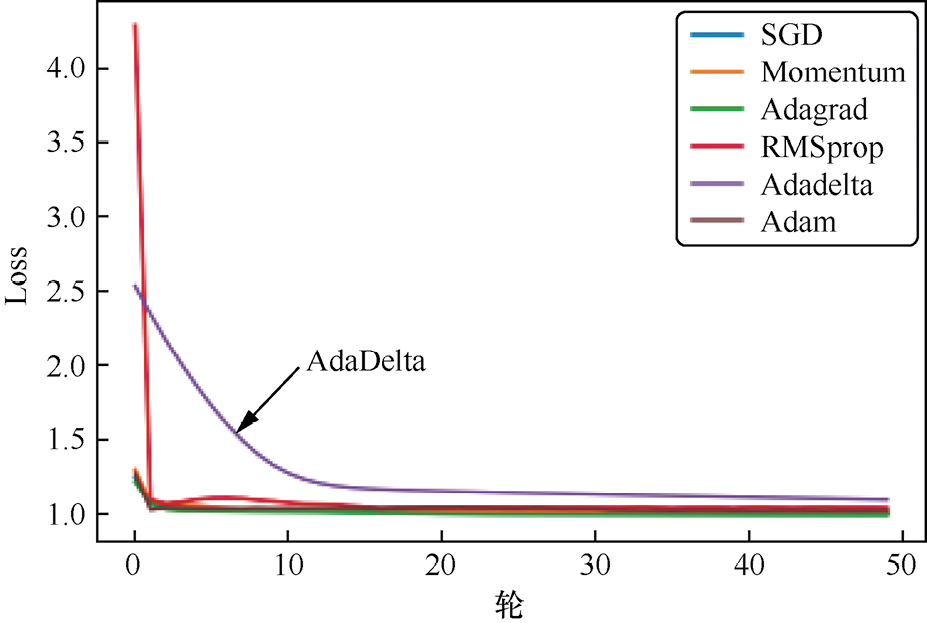


图6-31 训练集6种模型的收敛曲线

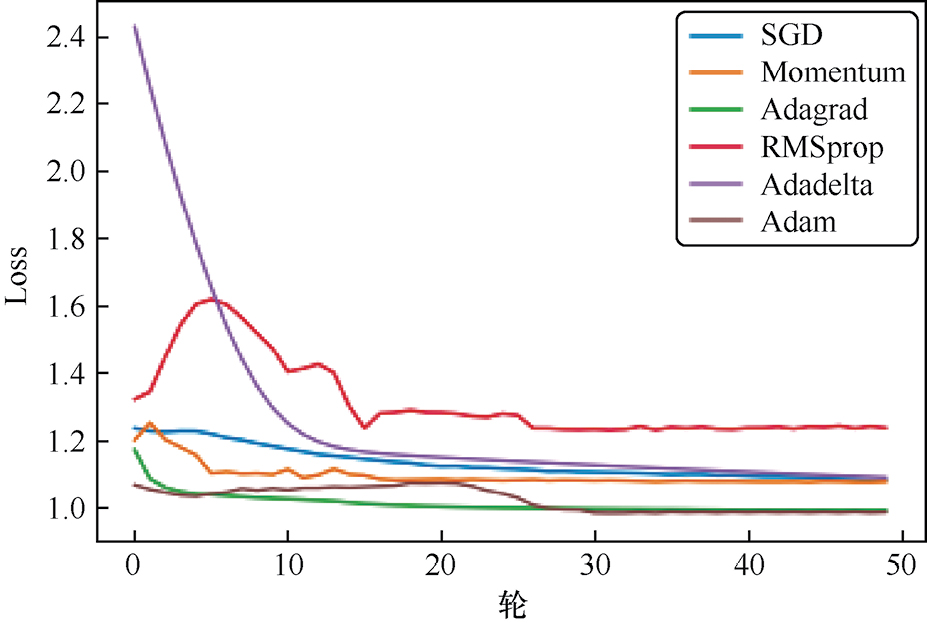


图6-32 测试集上6种模型的收敛曲线

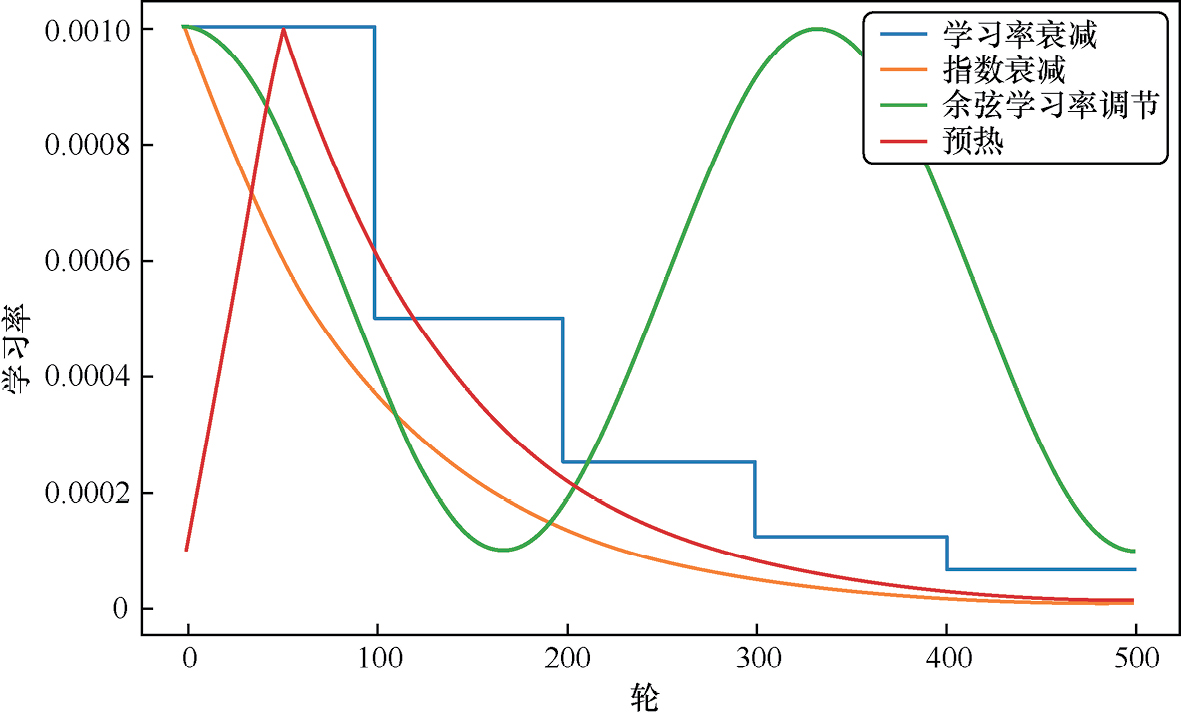


图6-33 常见学习率调节器示意

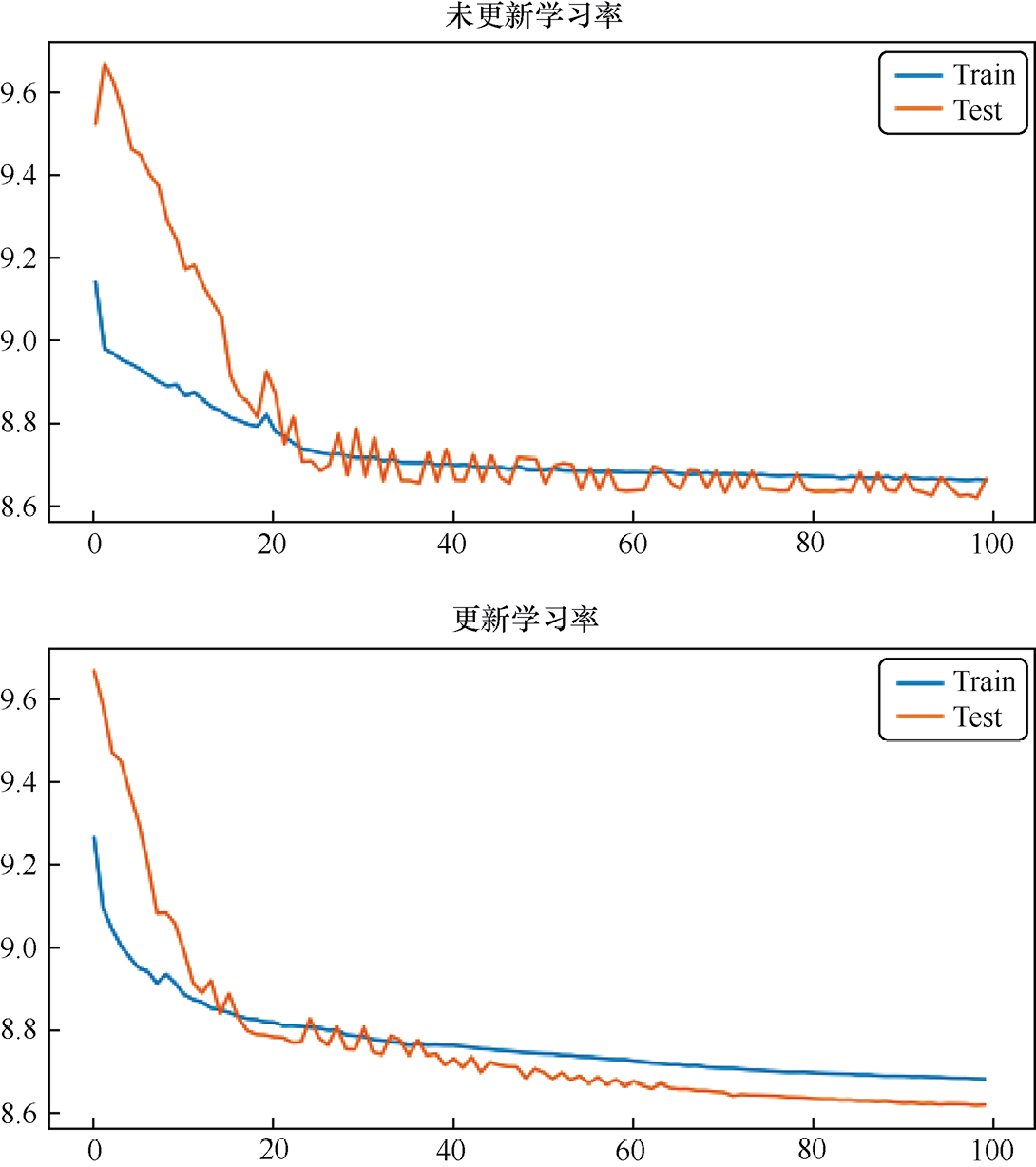


图6-34 训练和测试误差曲线