负梯度方法与 Newton 型方法：上机作业

龚梓阳

日期：2021 年 11 月 18 日

本次作业两道题目任选其一即可，可使用任意编程语言（如，Python，MATLAB，R 等）。请各位同学于 **12 月 3 日前**提交以下内容：

- 纸质版报告：包括对优化问题的推导、题目所要求的讨论结果（以表格或图片形式展示并辅以一段文字说明）以及算法实现中遇到的问题与解决方案等。
- 代码：将**带有注释说明的源代码**打包并命名为"**学号-姓名-上机作业 1**"后发送至助教邮箱：meetziyang@outlook.com。

1. 负梯度方法与 Newton 型方法的数值比较。编写下列程序（至少需实现 **5 种方法**，且必须包含**加粗**方法）：
   - 负梯度方法：**最速下降方法**、最小梯度方法和 BB 方法（BB1&BB2）。
   - Newton 方法：**基本牛顿方法**、阻尼 Newton 方法、混合方法和 LM 方法。
   - 拟 Newton 方法：SR1 方法、DFP 方法和 **BFGS 方法**。

   考虑最优化问题
   $$\min_{\mathbf{x}} f(\mathbf{x}) = \sum_{i=1}^{n/2} \left[100\left(x_{2i-1}^2 - x_{2i}\right)^2 + \left(x_{2i-1} - 1\right)^2\right] \tag{1}$$
   选择不同的问题规模（$n = 2, 10, 50$），对不同方法的有效性进行讨论。其中，
   - 迭代的终止准则为 $f_{k-1} - f_k \leq 10^{-8}$，迭代次数上限为 1000。
   - 讨论内容可以包括收敛结果，算法的迭代次数、函数调用次数、导数调用次数及 CPU 时间等。

2. 由人工神经网络方法解微分方程导出的最优化问题[1]。考虑以下常微分方程问题
   $$\frac{\mathrm{d}y(x)}{\mathrm{d}x} = x - y + 1, \quad y(0) = 1 \tag{2}$$
   其精确解为 $y(x) = x + \mathrm{e}^{-x}$。使用 1-n-1 前馈神经网络 $N(x; \mathbf{p})$ 在 $x \in [0, 1]$ 范围内，求解其实验解，即
   $$y_t(x; \mathbf{p}) = y_0 + xN(x; \mathbf{p}), \quad x \in [0, 1] \tag{3}$$

[1]第 2 题可使用现成软件包，如 Python（SciPy）、MATLAB（Optimization Toolbox）等。

该问题最终可转换为如下优化问题

$$\min_{\mathbf{p}} \sum_{i=1}^{m} \left\{ -x_i + \left[\sum_{j=1}^{n} (1+x_i) \frac{v_j}{1+\mathrm{e}^{\theta_j - w_j x_i}} + x_i \frac{v_j w_j \mathrm{e}^{\theta_j - w_j x_i}}{(1+\mathrm{e}^{\theta_j - w_j x_i})^2} \right] \right\}^2 \tag{4}$$

其中，$\mathbf{p} = (\mathbf{w}, \mathbf{v}, \boldsymbol{\theta})'$ 为该 1-n-1 前馈神经网络的参数，$x_1, x_2, \ldots, x_m \in [0,1]$，为方便起见，令 x_i 取 $[0,1]$ 上均匀分布的点。使用任意方法求解该优化问题，并讨论在不同网络节点数（$n = 2,3,4$）和训练元素个数（$m = 3,5,10$）的情况下，实验解 $y_t(x;\mathbf{p})$ 与精确解 $y(x)$ 的差异。

注. 考虑一阶常微分方程

$$\frac{\mathrm{d}y(x)}{\mathrm{d}x} = f(x,y), \quad x \in [0,1] \tag{5}$$

假设其实验解 $y_t(x)$ 具有如下形式

$$y_t(x;\mathbf{p}) = y_0 + xN(x;\mathbf{p}) \tag{6}$$

$N(x;\mathbf{p})$ 为 1-n-1 前馈神经网络，具有如下形式：

$$N(x;\mathbf{p}) = \sum_{j=1}^{n} v_j \varphi(z_j), \quad z_j = w_j x - \theta_j \tag{7}$$

其中，$\mathbf{p} = (\mathbf{w}, \mathbf{v}, \boldsymbol{\theta})'$，$w_j$ 是从输入节点 x 到隐藏层第 j 个节点的权重，v_j 是隐藏层第 j 个节点到输出节点的权重，θ_j 是隐藏层第 j 个节点的阈值，$\varphi(z) = \frac{1}{1+\mathrm{e}^{-z}}$ 为 Sigmoid 型激活函数。

我们希望对于实验解 $y_t(x)$，它能够尽可能地满足一阶常微分方程 (5)，即

$$\left[\frac{\mathrm{d}y_t(x;\mathbf{p})}{\mathrm{d}x} - f\left(x, y_t(x;\mathbf{p})\right) \right]^2 \to 0, \quad x \in [0,1] \tag{8}$$

因此，将 x 在区间 $[0,1]$ 离散化，取 $0 = x_1, x_2, \ldots, x_m = 1$，得到如下优化目标函数：

$$\min_{\mathbf{p}} \sum_{i=1}^{m} \left[\left. \frac{\mathrm{d}y_t(x;\mathbf{p})}{\mathrm{d}x} \right|_{x=x_i} - f\left(x_i, y_t(x_i;\mathbf{p})\right) \right]^2 \tag{9}$$