

Homework #1

截止日期: 3 月 13 日 0:00 之前

问题 #1

考虑一些浮点数值 x_1, x_2, \dots, x_n 。在许多机器学习算法中, 下面的函数通常被称为这些数值的 “log-sum-exp”:

$$\ell(x_1, \dots, x_n) := \ln \left[\sum_{k=1}^n e^{x_k} \right].$$

1. 通常情况下, $p_k := e^{x_k}$ 表示了一个概率 $p_k \in (0, 1]$ 。这种时候, x_k 的可能的取值范围是什么?
2. 假设许多的 x_k 都非常的小 ($x_k \ll 0$)。请解释为什么在这种情况下, 计算上面的 log-sum-exp 可能会产生数值错误。
3. 请证明, 对于任意 $a \in \mathbb{R}$,

$$\ell(x_1, \dots, x_n) = a + \ln \left[\sum_{k=1}^n e^{x_k - a} \right].$$

如果要避免你在 (2) 中提到的问题, 请选择一个合适的 a 的取值来改进计算 $\ell(x_1, \dots, x_n)$ 的稳定性。

问题 # 2

假设 $f(x)$ 和 $p(x)$ 是两个一阶可导的 \mathbb{R} 上的函数。

1. (选做) 我们想计算函数 $f(x)$ 的根。但是因为计算带来的误差, 我们可能实际上算的是另外一个函数 $f(x) + \varepsilon p(x)$ 的根。令 x^* 为函数 f 的

根，它满足 $f(x^*) = 0$ 。如果 $f'(x^*) \neq 0$ ，那么对于足够小的 ε ，存在函数 $x(\varepsilon)$ 使得 $f(x(\varepsilon)) + \varepsilon p(x(\varepsilon)) = 0$ 且 $x(0) = x^*$ 。假设这样的函数 $x(\varepsilon)$ 存在并且是可导的，请证明

$$\left. \frac{dx}{d\varepsilon} \right|_{\varepsilon=0} = -\frac{p(x^*)}{f'(x^*)}.$$

2. 假设 $f(x)$ 是 Wilkinson 多项式，定义为：

$$f(x) := (x-1)(x-2)(x-3)\cdots(x-20).$$

如果展开这个多项式，我们可以得到 $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{20}x^{20}$ ，其中 a_0, \dots, a_{20} 是对应次项的系数。如果我们刚好只有 a_{19} 这一项系数有误差，那么，可以令 $p(X) := x^{19}$ ，并使用 (1) 中的模型来分析最终的误差。根据这里选取的 $f(x), p(x)$ ，请证明对于任意 $j \in \{1, 2, \dots, 20\} =: S$ ，有：

$$\left. \frac{dx}{d\varepsilon} \right|_{\varepsilon=0, x^*=j} = -\prod_{k \in S: k \neq j} \frac{j}{j-k}.$$

3. 在 (2) 中分别取 $x^* = 1$ 和 $x^* = 20$ ，并对比各自的 $\frac{dx}{d\varepsilon}$ ，哪一个对这里的误差更加稳定？

问题 #3

令 $\lambda > 0$ 是一个实数，考虑函数 $f(x) = \frac{\lambda}{2+x}$ 。

1. 计算出函数 f 的所有不动点。
2. 令 $x_0 > 0$ 为函数 f 的不动点，请计算 $f'(x_0)$ 。
3. 假设 $\lambda = 3$ ，请证明从任何 $\mathbb{R}_{\geq 0}$ 中的点出发，关于函数 $f(x) = \frac{3}{2+x}$ 的不动点迭代过程总是会收敛到 x_0 。
4. 假设 $\lambda = 25$ ，请证明从 $x = 0$ 出发，关于函数 $f(x) = \frac{25}{2+x}$ 的不动点迭代过程会收敛到 x_0 。

提示： 你可能需要编写程序模拟不动点迭代过程的前若干步。

5. 令 $\varphi : \mathbb{R}_{>0} \rightarrow \mathbb{R}$ 是一阶可导的双射函数。利用函数 φ ，可以定义另一个函数 $g = \varphi \circ f \circ \varphi^{-1}$ 。注意， $x_0 > 0$ 是 f 的一个不动点，说明 $y_0 = \varphi(x_0)$ 是 g 的一个不动点。
- (a) 证明 $f'(x_0) = g'(y_0)$ 。
 - (b) 证明，从 $\mathbb{R}_{>0}$ 中的任何点出发，关于 f 的不动点迭代过程都会收敛，当且仅当，从 \mathbb{R} 中的任何点出发，关于 g 的不动点迭代过程都会收敛。
 - (c) 证明，从 $\mathbb{R}_{>0}$ 中的任何点出发，关于 f 的不动点迭代过程都会收敛到 x_0 (对于任何 $\lambda > 0$)。

提示: 如果你需要一个具体的 φ 来研究，我们推荐考虑自然对数 \log 。

6. 假设不动点迭代的计算过程中，每一次迭代都会带来一个不可控，任意的误差。设第 k 轮不动点迭代引入的误差为 ε_k ，并且 $\forall k, |\varepsilon_k| < 10^{-6}$ 。试证明：以上带误差的不动点迭代最终也会收敛到 f 不动点的一个邻域。

问题 #4

假设函数 g 为二次可导函数，且 r 为 $g(r) = r$ 的一个不动点，满足 $g'(r) = 0, |g''(r)| < 2$ 。试证明由 g 定义的不动点迭代方程 $x_{k+1} = g(x_k)$ 在 r 的一个邻域上二次收敛。