题目 4. Newton-Raphson 迭代法计算隐含波动率

问题背景

Newton-Raphson 迭代法是一种利用一阶导数求根的算法，现被广泛应用于隐含波动率计算。我们以欧式看涨期权为例，应用 Newton-Raphson 计算隐含波动率，并探究迭代起点对结果的影响。当临近行权期时，在 $\sigma \to 0^+$ 附近的导数非常小，Newton-Raphson 在什么时候收敛？通过机器学习模型猜测期权定价，能有更好的表现吗？请对此展开探究。

数据集和资料

数据集 `q_4_call_options.csv`

隐含波动率计算主要使用 Black-Scholes 公式，看涨期权的价格就是公式 1 的零点。

$$
\begin{aligned}
g(\sigma) &= s\Phi(d_1) - ke^{-rt}\Phi(d_2) - c = 0 \\
d_1 &= \frac{\ln\frac{s}{k} + (r + \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}} \\
d_2 &= d_1 - \sigma\sqrt{t} \\
\Phi(x) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{\tau^2}{2}}\mathrm{d}\tau
\end{aligned}
\tag{1}
$$

其中，s 期权对应的现货价格，k 是行权时合约价格，当时的无风险利率 $r = 0.02433$, c 是期权的售价，t 是现在距离行权日期的日数（单位：年）。

$g(\sigma)$ 的导数函数如公式 2 所示，$g''(\sigma) = 0$ 的解是 $\lambda = \sqrt{\frac{2}{t}|\ln\frac{s}{k} + rt|}$, 所以 Newton-Raphson 在 $[0, \lambda]$ 上收敛。

$$
\begin{aligned}
g'(\sigma) &= s\sqrt{\frac{t}{2\pi}}e^{-\frac{d_1^2}{2}} > 0 \\
g''(\sigma) &= s\sqrt{\frac{t}{2\pi}}e^{-\frac{d_1^2}{2}}\frac{d_1 d_2}{\sigma}
\end{aligned}
\tag{2}
$$

研究任务

问题 4-1. 使用矩阵运算，批量求解 λ 然后用 Newton-Raphson 求解 σ, 将结果代回到方程中，求得 $\hat{c}$. 使用 bootstrap 方法估计 $Var(\hat{c})$, $\bar{\hat{c}}$, 阐述这种方法是不是无偏估计。

问题 4-2. 将样本拆分为 80% 的训练集，20% 的测试集，不需要使用交叉验证，使用 bootstrap 随机抽样。训练一个机器学习回归模型，猜测 σ, 将结果代回到方程中，求得 $\hat{c}$. 在测试集上重复问题 4-2，比较机器学习与 Newton-Raphson 的结果。

问题 4-3. 有什么办法能提高模型的无偏性，减小模型的方差？绘制 $\hat{c} - c$ 的残差分布直方图与核密度估计曲线，参考论文 `https://arxiv.org/abs/2007.14313` 评价机器学习与传统迭代法在低频和高频数据上的表现。