清华大学计算机系科协算协联合暑培 2024
机器学习中的优化方法

张华清 (zhanghq22@mails.tsinghua.edu.cn)

2024 年 8 月 1 日

提示：作业中的部分题目可能还是需要稍想一想。希望大家玩得开心！

Question 1.

梯度下降法应用于光滑非凸函数的收敛速度。

对于可微函数 f，我们称 x 是其的 ϵ-一阶稳定点，如果

$$\|\nabla f\|^2 \le \epsilon$$

考虑梯度下降法：$x_{k+1} = x_k - \eta \nabla f(x_k)$。假设目标函数 f 是 L-光滑的，且有下界。试证明存在一个合适的学习率 η（和 L 相关），使得梯度下降法以 $O(\frac{1}{T})$ 的速率找到一个 ϵ-一阶稳定点。

提示：更具体地，证明

$$\min_{i=0,\cdots,T-1} \|\nabla f(x_i)\|^2 \le \frac{1}{T}\sum_{i=0}^{T-1} \|\nabla f(x_i)\| \le \frac{2L}{T}(f(x_0) - f(x^*))$$

Question 2.

神奇的不等式增加了。

若函数 f 是 L-光滑且为 μ-强凸，试证明对于任意 $x, y \in \mathbb{R}^d$，有

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \ge \frac{\mu L}{\mu + L}\|x - y\|^2 + \frac{1}{\mu + L}\|\nabla f(x) - \nabla f(y)\|^2$$

Question 3.

梯度下降法应用于光滑强凸函数的收敛速度。

考虑梯度下降法：$x_{k+1} = x_k - \eta \nabla f(x_k)$。假设目标函数 f 是 L-smooth 且为 μ-强凸，试证明存在一个合适的学习率 η（和 μ,L 相关），有

$$f(x_T) - f^* \le \frac{L}{2}\exp\left(-\frac{4T}{\kappa + 1}\right)\|x_0 - x^*\|^2$$

其中 $\kappa = \frac{L}{\mu}$，x^* 是 f 的最小值点。

提示：

1. 利用 Question 1. 中的不等式，证明
$$\|x_{t+1} - x^*\| \leq \left(1 - \frac{2\eta L\mu}{L+\mu}\right)\|x_t - x^*\|^2 + \eta\left(-\frac{2}{L+\mu} + \eta\right)\|\nabla f(x_t)\|^2$$
然后选取合适的 η。
2. $(1-x)^t \leq e^{-tx}$。
3. 由光滑性，$f(x) \leq \frac{L}{2}\|x^* - x\|^2$。

Question 4.

Nesterov 加速算法。

下面我们证明 Nesterov 加速算法应用于光滑凸函数的收敛速度为 $O(\frac{1}{T^2})$。我们考虑 Nesterov 加速算法的一个等价形式：

$$\begin{aligned} y_{t+1} &= x_t - \frac{1}{L}\nabla f(x_t) \\ z_{t+1} &= z_t - \eta_t \nabla f(x_t) \\ x_{t+1} &= (1-\tau_{t+1})y_{t+1} + \tau_{t+1} z_{t+1} \end{aligned}$$

我们令 $y_0 = z_0 = x_0, \eta_t = \frac{t+1}{L}, \tau_t = \frac{2}{t+2}$。

1. 证明 $\|z_{t+1} - x^*\|^2 - \|z_t - x^*\|^2 = \eta_t^2\|\nabla f(x_t)\|^2 + 2\eta_t\langle\nabla f(x_t), x^* - z_t\rangle$。
2. 定义势能函数 $\Phi_t = t(t+1)(f(y_t) - f(x^*)) + 2L\|z_t - x^*\|^2$。证明：
$$\Phi_{t+1} - \Phi_t \leq t(t+1)\langle\nabla f(x_t), x_t - y_t\rangle + 2(t+1)\langle\nabla f(x_t), x_t - x^*\rangle + 2(t+1)\langle\nabla f(x_t), x^* - z_t\rangle$$
提示：证明 $f(y_{t+1}) \leq f(x_t) - \frac{1}{2L}\|\nabla f(x_t)\|^2$。利用凸性：$f(x_t) - f(y_t) \leq \langle\nabla f(x_t), x_t - y_t\rangle$。
3. 证明 $\Phi_{t+1} - \Phi_t \leq 0$。
提示：利用 $x_{t+1} = (1-\tau_{t+1})y_{t+1} + \tau_{t+1}z_{t+1}$。
4. 证明 $f(y_t) - f(x^*) \leq 2L\frac{\|x_0 - x^*\|^2}{t(t+1)}$