第 1 章 第三次作业

作业提交形式为线上提交, 应使用 LaTeX 独立完成作业电子版, 并提交相应程序.

Deadline: 2022 年 5 月 7 日 23:59.

1.1 基础题

1.1.1 凸函数

利用凸函数的二阶条件证明 $f(x) = \log \sum_{k=1}^{n} \exp(x_k)$ 是凸函数.

注: 本问题可以讨论, 查阅资料.

1.1.2 梯度下降法

设 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} + b^T \boldsymbol{x}$ 是正定二次函数, 其中 $Q = diag(1, 2, 10)$, $\;b = (1, 0, 0)^T$ 设初始点为 $\boldsymbol{x}_0 = (1, 1, 1)^T$. 使用最速下降法, 用精确线搜索确定步长.

1. 算法的收敛速率是怎样的?
2. 要使 $f(x_T) - f(x^*) < 10^{-10}$, 大约需要多少次迭代?

1.1.3 强凸与光滑问题判定

设 $\min_x f(x) = \frac{1}{2}\|Ax - y\|^2$, 其中 $A \in \mathbb{R}^{n\times n}$ 是非退化矩阵. 使用梯度下降法求解此问题

$$x_{k+1} = x_k - \alpha \nabla f(x_k),$$

其中 α 为步长.

1. 计算 $f(x)$ 相对于 x 的梯度.
2. 设 $A = diag(-1, 5)$, 证明 $f(x)$ 是强凸函数, 计算参数 μ.
3. 设 $A = diag(-1, 5)$, 证明 $f(x)$ 是光滑函数, 计算光滑函数的参数 L.
4. 设 $A = diag(1, 2)$, 并且步长为 $\alpha_k = 1/4$, 求算法的单步下降幅度 $f(x_{k+1}) - f(x_k)$.
5. 设 $A = diag(1, 2)$, 算法使用固定步长. 问如果要使迭代序列是下降序列, 步长应满足什么条件?
6. 设 $A = diag(1, 2)$, 如果 $f(x_0) - f(x^*) = 1$, 要使 $f(x_k) - f(x^*) < 10^{-3}$, 试估计需要多少次迭代.

1.1.4 共轭梯度法性质

设 $f : \mathbb{R}^3 \to \mathbb{R}$ 连续可微. 使用共轭梯度法优化 $f(x)$, 初始点为 $x^{(0)}$, 迭代序列为 $x^{(k+1)} = x^{(k)} + \alpha_k d_k$, 步长由精确线搜索确定. 如果第一次迭代方向为 $d_0 = (1, -1, 2)^T$, 沿 d_0 精确线搜

索得到 $x^{(1)}$ 并且在 $x^{(1)}$ 处

$$\left.\frac{\partial f(x)}{\partial x_1}\right|_{x^{(1)}} = \left.\frac{\partial f(x)}{\partial x_2}\right|_{x^{(1)}} = -2.$$

1. 求 $x^{(1)}$ 处的搜索方向
2. 如果在 $x^{(2)}$ 处梯度长度为 $\|\nabla f(x^{(2)})\| = 1$. 求 $\nabla f(x^{(2)})$.

1.1.5 共轭梯度法的等价性

设 $f(x) = \frac{1}{2}x^T Qx + b^T x + c$ 是二次函数, Q 是正定矩阵.

1. 给定下降方向 d, 计算精确线搜索确定的步长.
2. 证明, 在上述问题上, 使用精确线搜索确定步长时, 三种共轭梯度法所确定的点列 $\{x_i\}$ 是相同的.

1.1.6 共轭梯度法的子空间性质与收敛性

设 $f(x) = \frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x} + b^T\boldsymbol{x} + c$ 是二次函数, Q 是正定矩阵. 使用共轭梯度法, 证明:

1. 每次迭代的搜索方向 $\mathbf{d}_i$ 都可以写成下面向量组的线性组合
$$\{\mathbf{d}, Q\mathbf{d}, Q^2\mathbf{d}, \cdots, Q^i\mathbf{d}\}.$$
2. 对于第 $i+1$ 次迭代的序列 $\boldsymbol{x}_{i+1}$, 证明, 对于任意由上述向量线性组合的向量 $\boldsymbol{x}'$, 都有 $f(\boldsymbol{x}') \geq f(\boldsymbol{x}_{i+1})$.
3. 使用上述结论证明, 共轭梯度法在二次函数 $f(\boldsymbol{x})$ 上迭代 K 次即可收敛到最优点, 其中 K 是矩阵 Q 的不同的特征值的数目 (number of distinct eigenvalues). (说明, 如果 Q 的特征值为 $1, 3, 3, 5, 5, 7$, 则分离特征值数目为 4 个.)

1.1.7 约束优化问题的 KKT 条件

考虑由 PCA 的约束优化问题: 设 $\boldsymbol{x}_i$ 是一组数据, $\boldsymbol{x}_i$ 的均值为零, $S = \sum_{i=1}^m \boldsymbol{x}_i\boldsymbol{x}_i^T$ 是协方差矩阵, 一般 PCA 问题可以写成:

$$\arg\max_V \quad \mathrm{tr}(V^T SV), \mathrm{s.t.} \quad V^T V = I_k$$

其中 $V \in \mathbb{R}^{n\times k}$, 完成以下问题.

1. 写出其 KKT 条件, 并分析 KKT 点.
2. 如果使用投影梯度法进行求解上述问题, 写出投影公式.

1.2 思考题

思考以下问题, 了解共轭方向与特征向量的关系

1.2.1 共轭方向与特征向量

设 $f(x) = \frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x} + b^T\boldsymbol{x}$ 是正定二次函数, Q 的特征分解为 $Q = U\Lambda U^T$, 特征值满足 $\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$, 其对应的特征向量分别为 $1, \cdots, n$. 令 $\boldsymbol{x}^*$ 表示 f 的最小值点. 取初始点 $\boldsymbol{x}_0$, 并且 $\boldsymbol{x}_0 = \boldsymbol{x}^* + \sum_{i=1}^n \beta_i i$ 满足 $\beta_i \neq 0$,$i = 1, \ldots, n$. 构造迭代算法

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \beta_k k$$

1. 证明: 方向组 $1, \cdots, n$ 是关于 Q 相互共轭的.
2. 对于任意 $k = 1, \ldots, n-1$, 求单步改进幅度 $f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k)$.
3. 如果对于 $k = 1, \ldots, n-1$, 每一次迭代的改进幅度 $\delta_k = |f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k)|$ 都相等, 则步长 β_k 满足什么条件?

1.3 编程题

阅读以下页面内容, 并实现相应算法.

1.3.1 使用 Newton-CG 与 L-BFGS 法求解 LASSO 问题

注: 点击题目会跳转相应网页.

1. Newton-CG 法
   阅读理解上述 MATLAB 程序, 实例程序转写成 Python 程序, 并实现 Newton-CG 算法.
2. 应用牛顿-共轭梯度法解逻辑回归问题.
   阅读上述程序, 并用 Python 实现上述实例.
3. L-BFGS 算法的 MATLAB 实现.
   阅读理解上述 MATLAB 程序, 实例程序转写成 Python 程序, 并实现 L-BFGS 算法.
4. 利用 L-BFGS 方法求解逻辑回归问题.
   阅读上述程序, 并用 Python 实现上述实例.