作业二

最优化方法

请在规定时间前提交到大夏学堂，超时得分将有折扣。
计算、证明题提交 pdf 电子版，编程题提交 Python 代码、结果及必要的解释。

1 带 ℓ_2 惩罚的部分优化问题

考虑问题

$$\min_{\beta,\,\sigma\geq 0}\ f(\beta)+\frac{\lambda}{2}\sum_{i=1}^{n}g(\beta_i,\sigma_i), \tag{1}$$

其中，f 为定义在 $\mathbb{R}^n$ 上的凸函数，$\lambda\geq 0$，且

$$g(x,y)=\begin{cases}x^2/y+y & \text{if } y>0\\ 0 & \text{if } x=0,\ y=0\\ \infty & \text{else.}\end{cases}$$

a. 证明 g 是凸函数，即上述问题为凸优化问题。（后面我们可根据此进行部分优化，且部分优化后的函数也是凸函数）

b. 证明：

$$\min_{y\geq 0} g(x,y)=2|x|$$

c. 证明(1)中对于 $\sigma\geq 0$ 的优化可得 ℓ_1 惩罚问题

$$\min_{\beta}\ f(\beta)+\lambda\|\beta\|_1.$$

a. Proof. First, the domain of g is convex. When $y>0, g(x,y)=P\left(x^2,(0,1)(x,y)^\top\right)+(0,1)(x,y)^\top$, where P is the perspective transformation. The affine mapping $(0,1)(x,y)^\top$ and x^2 are both convex, so $P\left(x^2,(0,1)(x,y)^\top\right)$ is also convex. And the sum of two convex function is still convex implying $g(x,y)$ is convex, i.e.

$$\min_{y\geq 0} g(x,y)=2|x|$$

b. Proof:

$$\min_{y\geq 0} g(x,y)=2|x|$$

Proof. Given x, the first order partial derivative of $g(x,y)$ is

$$\frac{\partial}{\partial y} g(x, y) = 1 - \frac{x^2}{y^2}$$

Let $\partial g(x,y)/\partial y = 0$ *and we have* $y = |x|$ *since* $y \geq 0$. *In question a., we've known* $g(x,y)$ *is convex, so it attains its minimum at* $y = |x|$ *when* x *is fixed.*

c. Proof (1) 中对于 $\sigma \geq 0$ *的优化可得* ℓ_1 *惩罚问题*

$$\min_{\beta} f(\beta) + \lambda \|\beta\|_1$$

Proof. *We can optimize over* σ *first and then minimize the question over* β. *The result in question b. gives that* $g(\beta_i, \sigma_i)$ *attains the minimum* $2|\beta_i|$ *when* $\sigma_i = |\beta_i|$. *Thus,*

$$\min_{\beta, \sigma \geq 0} f(\beta) + \frac{\lambda}{2} \sum_{i=1}^{n} g(\beta_i, \sigma_i)$$

is equivalent to

$$\min_{\beta} f(\beta) + \frac{\lambda}{2} \sum_{i=1}^{n} |\beta_i| = \min_{\beta} f(\beta) + \lambda \|\beta\|_1.$$

2 Lipschitz 梯度与强凸性

令 f 为二次连续可微的凸函数

a. 证明以下命题等价：

  i. ∇f 为 L-Lipschitz 函数
  （即存在常数 $L > 0$，使得 $\|f(x) - f(y)\|_2 \leq L\|x - y\|_2$, 对任意 x, y.）

  ii. 对任意 x, y，$(\nabla f(x) - \nabla f(y))^\top (x - y) \leq L\|x - y\|_2^2$

  iii. 对任意 x，$\nabla^2 f(x) \preceq LI$

  iv. 对任意 x, y，$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|_2^2$

循环证明 $i \Rightarrow ii, ii \Rightarrow iii, iii \Rightarrow iv, iv \Rightarrow ii, iii \Rightarrow i$

Proof. $i \Rightarrow ii, ii \Rightarrow iii, iii \Rightarrow iv, iv \Rightarrow ii$, $iii \Rightarrow i$.

$(i \Rightarrow ii)$. *Since* ∇f *is L-Lipschitz function,*

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2 \textit{ for all } x, y.$$

Multiply both sides by $\|x - y\|_2$ *and we obtain*

$$\|\nabla f(x) - \nabla f(y)\|_2 \|x - y\|_2 \leq L\|x - y\|_2^2.$$

Cauchy-Schwarz inequality guarantees

$$\|\nabla f(x) - \nabla f(y)\|_2 \|x - y\|_2 \geq (\nabla f(x) - \nabla f(y))^\top (x - y).$$

Thus, ii. holds.

(ii ⇒ iii). Let $x = y + t\alpha$*, where* $t \in \mathbb{R}$ *and* $\alpha \in \mathbb{R}^n, \|\alpha\|_2 = 1$*. Apply Taylor's expansion and we have*

$$\nabla f(x) = \nabla f(y) + \nabla^2 f(\xi)(x-y), \xi \text{ between } x \text{ and } y$$

According to result in (ii),

$$\begin{aligned}
&(\nabla f(x) - \nabla f(y))^\top (x-y) = \left(\nabla^2 f(\xi)(x-y)\right)^\top (x-y) \le L\|x-y\|_2^2 \\
&\Rightarrow t^2 \alpha^\top \nabla^2 f(\xi)\alpha \le L\left(t^2\|\alpha\|_1^2\right) \\
&\Rightarrow \alpha^\top \nabla^2 f(\xi)\alpha \le L \quad (\text{ since } \|\alpha\|_2 = 1 \ ).
\end{aligned}$$

Taking $t \to 0$ *gives* $\nabla^2 f(\xi) = \nabla^2 f(x)$*, and let* α *be the eigenvector of the maximal eigenvalue of* $\nabla^2 f(x)$ *gives* $\lambda_{\max}\nabla^2 f(x) \le L$*. Hence,* $\nabla^2 f(x) \preceq LI$*.*

(iii ⇒ iv). By mean value version of Taylor's theorem,

$$f(y) = f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(\xi)(y-x), \text{ where } \xi \text{ between } x \text{ and } y.$$

Since $\nabla^2 f(\xi) \succeq LI$ *confirmed by (iii),*

$$f(y) \le f(x) + \nabla f(x)^\top (y-x) + \frac{L}{2}\|y-x\|_2^2$$

(iv ⇒ ii). From (iv), we have

$$\begin{aligned}
f(y) \le f(x) + \nabla f(x)^\top (y-x) + \frac{L}{2}\|y-x\|_2^2, \\
f(x) \le f(y) + \nabla f(y)^\top (x-y) + \frac{L}{2}\|x-y\|_2^2.
\end{aligned}$$

The sum of the two inequalities gives

$$\begin{aligned}
&f(y) + f(x) \le f(x) + f(y) + (\nabla f(y) - \nabla f(x))^\top (x-y) + L\|x-y\|_2^2 \\
\Rightarrow &(\nabla f(x) - \nabla f(y))^\top (x-y) \le L\|x-y\|_2^2.
\end{aligned}$$

(iii ⇒ i). Apply the mean value version of Taylor's theorem and then

$$\nabla f(x) - \nabla f(y) = \nabla^2 f(\xi)(x-y), \text{ where } \xi \text{ between } x \text{ and } y.$$

Taking the norm of both sides gives

$$\|\nabla f(x) - \nabla f(y)\|_2 = \left\|\nabla^2 f(\xi)(x-y)\right\|_2.$$

Since $\nabla^2 f(x) \preceq LI$ *for* $\forall x, \nabla^2 f(\xi) \preceq LI$*. By Cauchy-Schwarz inequality, we have*

$$\|\nabla f(x) - \nabla f(y)\|_2 = \left\|\nabla^2 f(\xi)(x-y)\right\|_2 \le \left\|\nabla^2 f(\xi)\right\|_2 \|x-y\|_2 \le L\|x-y\|_2.$$

b. 证明以下命题等价：

i. f 为 m-强凸函数（即 $f(x)-\frac{m}{2}\|x\|_2^2$ 为凸函数）

ii. 对任意 x,y，$(\nabla f(x)-\nabla f(y))^\top(x-y)\geq m\|x-y\|_2^2$

iii. 对任意 x，$\nabla^2 f(x)\succeq mI$

iv. 对任意 x,y，$f(y)\geq f(x)+\nabla f(x)^\top(y-x)+\frac{m}{2}\|y-x\|_2^2$

循环证明 $i\Rightarrow ii, ii\Rightarrow iii, iii\Rightarrow iv, iv\Rightarrow i$

($i\Rightarrow ii$). Since f is m-strongly convex, $g(x)=f(x)-m/2\|x\|_2^2$ is convex for some $m>0$. $\nabla g(x)=\nabla f(x)-mx$. It follows from the monotone gradient condition for convexity of $g(x)$, i.e.

$$\begin{aligned}&(\nabla g(x)-\nabla g(y))^\top(x-y)\geq 0\\ \Rightarrow&(\nabla f(x)-\nabla f(y))^\top(x-y)\geq m\|x-y\|_2^2.\end{aligned}$$

($ii\Rightarrow iii$). Let $x=t\alpha+y$, where $t\in\mathbb{R}$ and $\alpha\in\mathbb{R}^n, \|\alpha\|=1$. Use Taylor's theorem and we have

$$\nabla f(x)-\nabla f(y)=\nabla^2 f(y)(x-y)+o(\|x-y\|)$$

Then multiplying both sides by $x-y$ gives

$$\begin{aligned}&(\nabla f(x)-\nabla f(y))^\top(x-y)=(x-y)^\top\nabla^2 f(y)(x-y)+o\left(\|x-y\|_2^2\right)\geq m\|x-y\|_2^2\\ \Rightarrow&t^2\alpha^\top\nabla^2 f(y)\alpha+o\left(t^2\|\alpha\|_2^2\right)\geq mt^2\|\alpha\|_2^2\\ \Rightarrow&\alpha^\top\nabla^2 f(y)\alpha+o(1)\geq m\end{aligned}$$

Let α be the eigenvector of the minimal eigenvalue of $\nabla^2 f(y)$ and taking $t\to 0$ gives $\lambda_{\min}\nabla^2 f(y)\geq m$. Hence, $\nabla^2 f(x)\succeq mI$. ($iii\Rightarrow iv$). Let $g(x)=f(x)-\frac{m}{2}\|x\|_2^2$. Since $\nabla^2 g(x)=\nabla^2 f(x)-mI\succeq 0, g(x)$ is convex. The convexity of $g(x)$ gives

$$\begin{aligned}&g(y)\geq g(x)+\nabla g(x)^\top(y-x)\\ \Leftrightarrow&f(y)-\frac{m}{2}\|y\|_2^2\geq f(x)-\frac{m}{2}\|x\|_2^2+(\nabla f(x)-mx)^\top(y-x)\\ \Leftrightarrow&f(y)\geq f(x)+\nabla f(x)^\top(y-x)+\frac{m}{2}\|x-y\|_2^2.\end{aligned}$$

($iv\Rightarrow i$). Let $g(x)=f(x)-\frac{m}{2}\|x\|_2^2$. $f(y)\geq f(x)+\nabla f(x)^\top(y-x)+\frac{m}{2}\|x-y\|_2^2$ implies that $g(y)\geq g(x)+\nabla g(x)^\top(y-x)$, i.e. $g(x)$ is convex. Thus, $f(x)$ is m-strongly convex.

3 实践：使用 Pytorch 编写带回溯线搜索的梯度下降算法

- 根据所给模板，编写带回溯线搜索的梯度下降函数，并用该函数解决以下非约束光滑优化问题：
  - 逻辑回归问题（已给定模拟数据）

$$\min_{\beta\in\mathbb{R}^p}\ \frac{1}{n}\sum_{i=1}^n\log\bigl(1+\exp(-y_i\beta^\top x_i)\bigr).\tag{2}$$

更正逻辑回归目标函数（本小题不扣分）

$$\min_{\beta \in \mathbb{R}^p} \ -\frac{1}{n}\sum_{i=1}^{n}\left[y_i\beta^\top x_i - \log\bigl(1+\exp(\beta^\top x_i)\bigr)\right]. \tag{3}$$

见文件：homework2_ans_ref.ipynb