Практическое задание 1

Рыбка Елизавета, 474

16 октября 2017 г.

1 Формулы для градиента и гессиана функции логистической регрессии

$f(x) = \frac{1}{m}\sum_{i=1}^{m} \ln(1+exp\{-b_i a_i^T x\}) + \frac{\lambda}{2}x^T x = \wr\wr\ \phi(t) = \frac{1}{m}\ln(1+exp\{-t\})\ \wr\wr =$
$\frac{1}{m}\sum_{i=1}^{m} \phi(b_i a_i^T x) + \frac{\lambda}{2}x\cdot x = \frac{1}{m}\langle 1_m, \phi(b_i a_i^T x)\rangle + \frac{\lambda}{2}\|x\|^2 = \frac{1}{m}\langle 1_m, \phi(\langle b, Ax\rangle)\rangle + \frac{\lambda}{2}x\cdot x =$
$\frac{1}{m}1_m\cdot\phi(b\cdot Ax) + \frac{\lambda}{2}x\cdot x$

$\nabla f = \frac{1}{m}\sum_{i=1}^{m} \frac{-exp\{-b_i a_i^T x\} b_i a_i^T}{1+exp\{-b_i a_i^T x\}} + \frac{\lambda}{2}2x = \frac{1}{m}\sum_{i=1}^{m}(\frac{-1}{1+exp\{b_i a_i^T x\}})b_i a_i^T + \lambda x = \wr\wr\ \varphi = \frac{1}{1+exp\{-t\}}\ \wr\wr =$
$\frac{1}{m}\sum_{i=1}^{m} \varphi(-b_i a_i^T x) b_i a_i^T + \lambda x = -\frac{1}{m}\langle 1_m, (\varphi(-b\cdot Ax)\cdot b)A\rangle + \lambda x =$
$= -\frac{1}{m}\langle A, \varphi(-b\cdot Ax)\cdot b\rangle + \lambda x$

$\nabla^2 f = \frac{1}{m}A^T(\frac{e^{\langle b,Ax\rangle}\langle b,A\rangle}{(1+e^{\langle b,Ax\rangle})^2}\cdot b) + \lambda E = \frac{1}{m}A^T diag\{b^T b\frac{1}{1+e^{\langle b,Ax\rangle}}\frac{1}{1+e^{-\langle b,Ax\rangle}}\}A + \lambda E =$
$\frac{1}{m}A^T diag\{b^T b\varphi(-b\cdot Ax)\varphi(b\cdot Ax)\}A + \lambda E$

2 Эксперимент: Траектория градиентного спуска на квадратичной функции

Описание

Исследуется поведение метода в зависимости от числа обусловленности функции, выбора начальной точки и стратегии выбора шага
В эксперименте оптимизируются функции:

$$f_1(x_1, x_2) = \frac{1}{2}(2x_1^2 + x_2^2)$$
$$f_2(x_1, x_2) = \frac{1}{2}(15x_1^2 + x_2^2),$$

которые имеют константы обусловленности 2 и 15 соответственно.

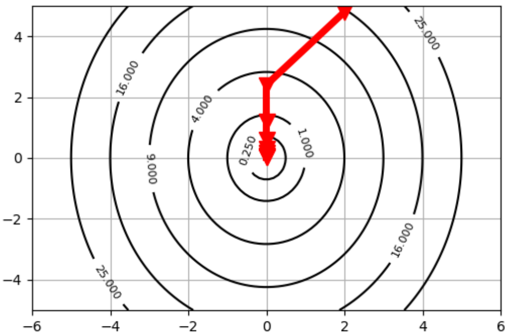
Используется метод градиентного спуска со следующими стратегиями выбора шага и следующими параметрами:

- Константная стратегия, $c_{f_1} = \frac{1}{2}, c_{f_2} = \frac{1}{15}$[1]
- Правило Армихо, $c_1 = 10^{-4}, \alpha_0 = 1.0$
- Правило Вульфа, $c_1 = 10^{-4}, c_2 = 0.9, \alpha_0 = 1.0$

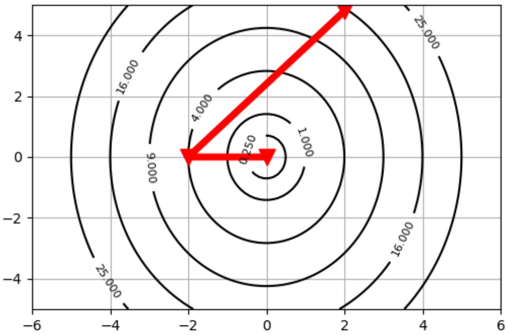
Результаты

На Рис.1 представлены получившиеся траектории градиентого спуска для f_1(слева), f_2 (справа). В подписях к ним указаны метод выбора шага и количество потребовавшихся операций.

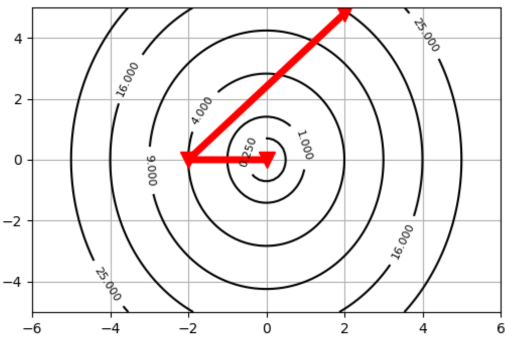
[1]При $c = \frac{1}{L}$ можно гарантировать сходимость константантного градиентного шага, где L — константа липшица



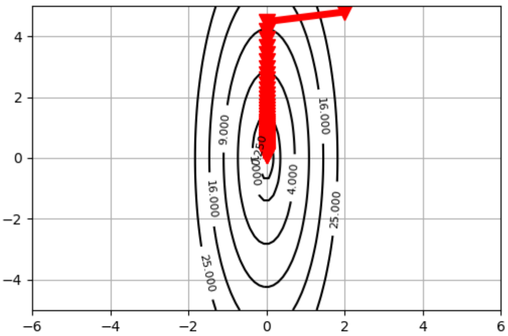
a)Константный, steps=8



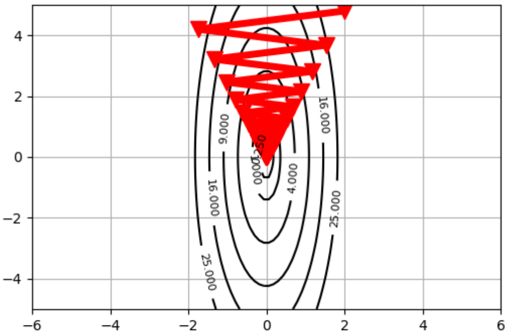
b)Армихо, steps=2



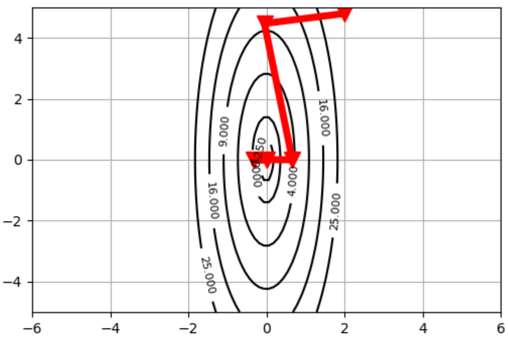
c)Вульф, steps=2



d)Константный, steps=57



e)Армихо, steps=44



f)Вульф, steps=4

Рис. 1: Траектории градиентного спуска для f_1(a) —c)), f_2 (d) —f))

Выводы

- Для оптимизации функции с большим числом обусловленности требуется большее число итераций.
- Градиентный спуск сходится быстрее всего при выборе шага по методу Вульфа. На втором месте - Армихо, самый медленный - константный.

3 Эксперимент: Зависимость числа итераций градиентного спуска от числа обусловленности и размерности пространства

Описание

Исследуется зависимость числа итераций, необходимых градиентному спуску для сходимости, от числа обусловленности k оптимизируемой функции и размерности пространства nоптимизируемых переменных.

Фиксируемся одно из $n = 10, 100, 1000, 10000$. Берется k по сетке от 10 до 500 с шагом 10. Для каждого k случайным образом генерируется функция и начальная точка (по алгоритму описанному ниже). Проводится оптимизация методом градиентного спуска с шагом Вульфа (т.к. в предыдущем эксперименте он показал наилучшие результаты). Операции повторяются 5 раз. Таким образом, для каждого значения n получается 5 ломанных $T(k, n)$. Также для каждого n строится усреднение указанных кривых.
При фиксированных k, n функция и начальная точка генерируются следующим образом:

- Создается список длины n, на первом месте в котором стоит k, на втором 1, далее идут случайные числа из интервала $[1, k)$. $A = diag(a)$. Таким образом все условия на A выполняются.
- b генерируется как случайный вектор размерности n, с компонентами из интервала $[-2k, 2k]$ (случайно и равномерно).
- x_0 генерируется как случайный вектор размерности n, с компонентами из интервала $[-2k, 2k]$ (случайно и равномерно).

Результаты

На Рис. 2 тонкими линиями представлены графики зависимости числа итераций $T(k, n)$ от числа обусловленности k. Различным размерностям пространства n соответствуют различные цвета (см. легенду). Толстые линии - усредненные графики для каждого n.

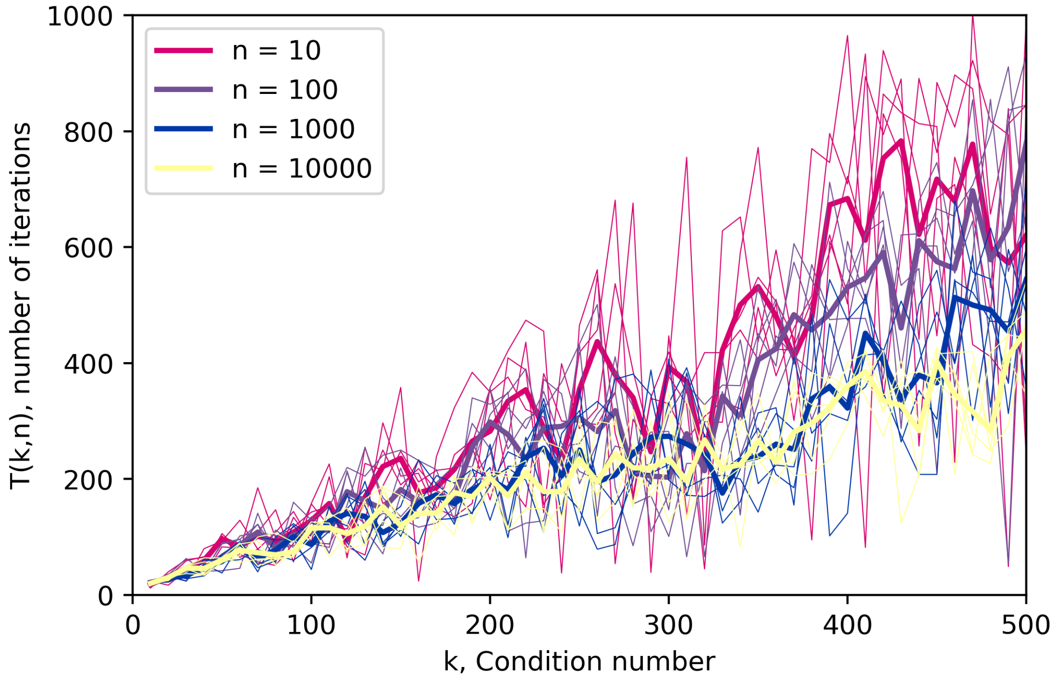


Рис. 2: Зависимость числа итераций градиентного спуска от числа обусловленности и размерности пространства

Выводы

- Складывается впечатление, что размерность пространства не сильно влияет на число операций (хотя в усредненных графиках слабая зависимость на графиках все же видна).
- Число итераций в среднем увеличивается с ростом числа обусловленности. Это согласуется с результатами 1го эксперимента.

4 Эксперимент: Сравнение методов градиентного спуска и Ньютона на реальной задаче логистической регрессии

Провести эксперимент не удалось.