ШАД
Методы оптимизации
Практическое задание 1

Андрей Данилов

March 2018

Вывод выражений для градиента и гессиана логистической регрессии

$$f(x) = \frac{1}{m}\sum_{i=0}^{m} \ln\left(1 + exp\left(-b_i(A_i, x)\right)\right) + \frac{r}{2}\|x\|^2$$

$$df(x) = \frac{1}{m}\sum_{i=0}^{m} \frac{-b_i(A_i, dx)\ exp\left(-b_i(A_i, x)\right)}{1 + exp\left(-b_i(A_i, x)\right)} + r(x, dx)$$

$$\nabla f(x) = \frac{1}{m}\sum_{i=0}^{m} \frac{-b_i A_i}{1 + exp\left(b_i(A_i, x)\right)} + rx$$

$$d\nabla f(x) = \frac{1}{m}\sum_{i=0}^{m} \frac{b_i^2 A_i\ exp\left(b_i(A_i, x)\right)(A_i, dx)}{\left(1 + exp\left(b_i(A_i, x)\right)\right)^2} + rdx$$

$$\nabla^2 f(x) = \frac{1}{m}\sum_{i=0}^{m} \frac{b_i^2 A_i A_i^T\ exp\left(b_i(A_i, x)\right)}{\left(1 + exp\left(b_i(A_i, x)\right)\right)^2} + rI_n$$

Перепишем полученные формулы без суммирований.

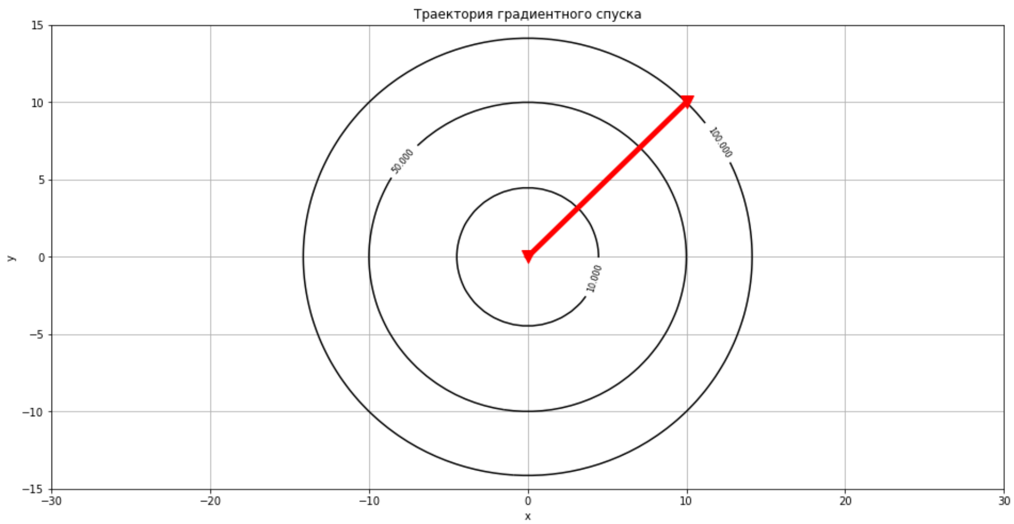
$$f(x) = \frac{1}{m}\big(1_m, \ln\left(1 + exp\left(-b \odot Ax\right)\right)\big) + \frac{r}{2}\|x\|^2$$

$$\nabla f(x) = \frac{A^T}{m}\left(b \odot \frac{-1}{1 + exp\left(b \odot Ax\right)}\right) + rx$$

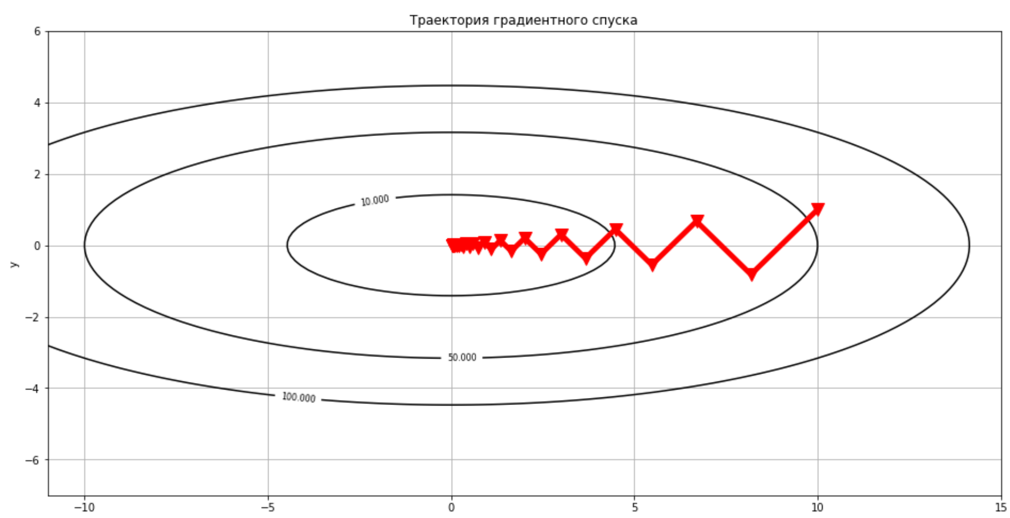
$$\nabla^2 f(x) = \frac{1}{m} A^T\ diag\left(\frac{b \odot b \odot exp\left(b \odot Ax\right)}{\left(1 + exp\left(b \odot Ax\right)\right)^2}\right) A\ + rI_n$$

Эксперимент 1. Траектория градиентного спуска на квадратичной функции.

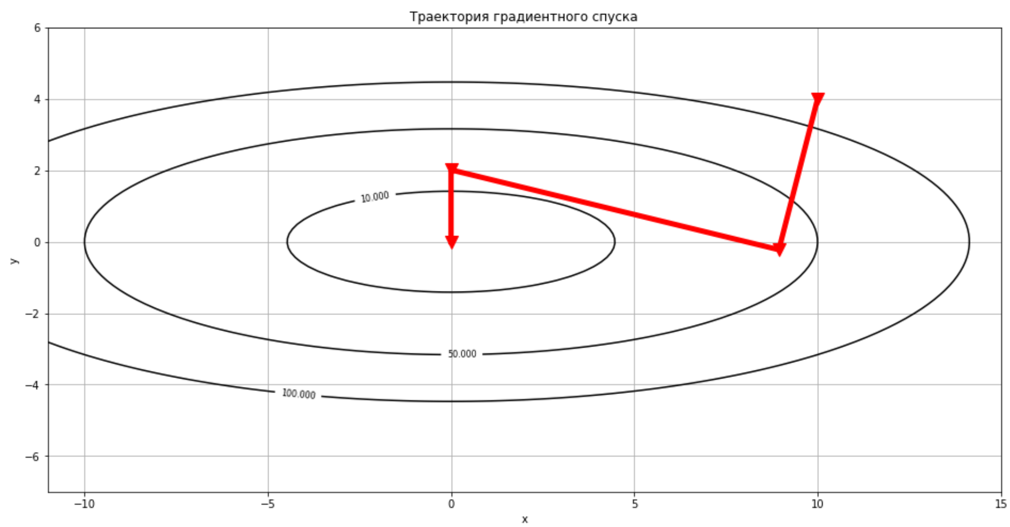
Траектория градиентного спуска для матрицы с числом обусловленности равным единице. Ожидаемо методу потребовался всего один шаг для достижения минимума.



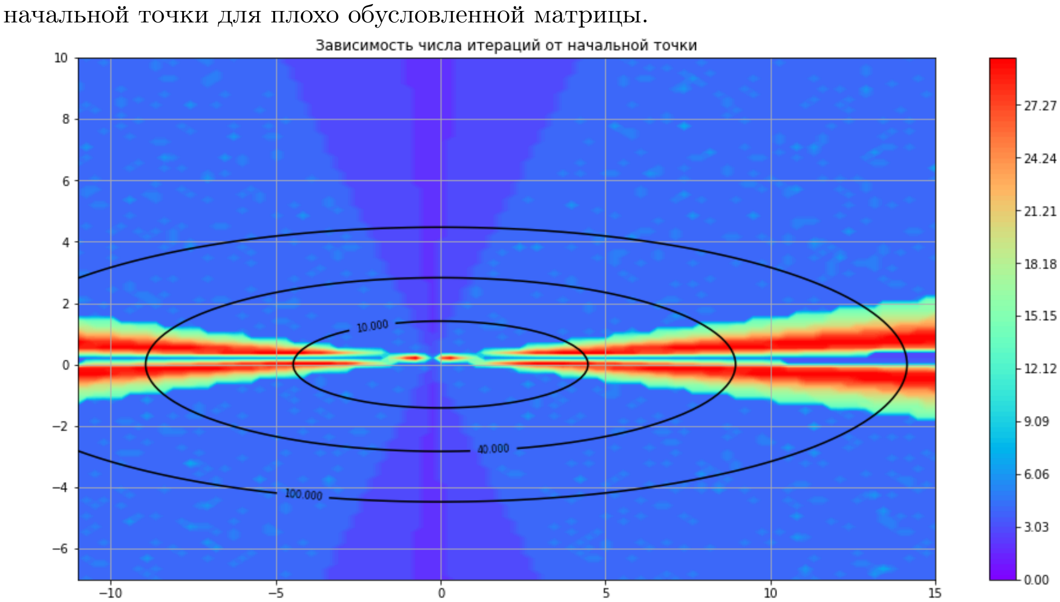
Для плохо обусловленной матрицы траектория становиться зигзагообразной и методу требуется уже порядка 30 итераций для достижения минимума.



Однако для другой начальной точки ситуация кардинально меняется.



Посмотрим, как меняется число итераций алгоритма в зависимости от начальной точки для плохо обусловленной матрицы.



Область медленной работы (красная) расположена близко к оси эллипса.

Таким образом, для хорошо обусловленной матрицы алгоритм всегда работает быстро, однако плохая обусловленность не является необходимым условием медленной работы. Для этого нужна ещё и не удачная начальная точка.

В зависимости от стратегии выбора шага алгоритм имеет следующие особенности:

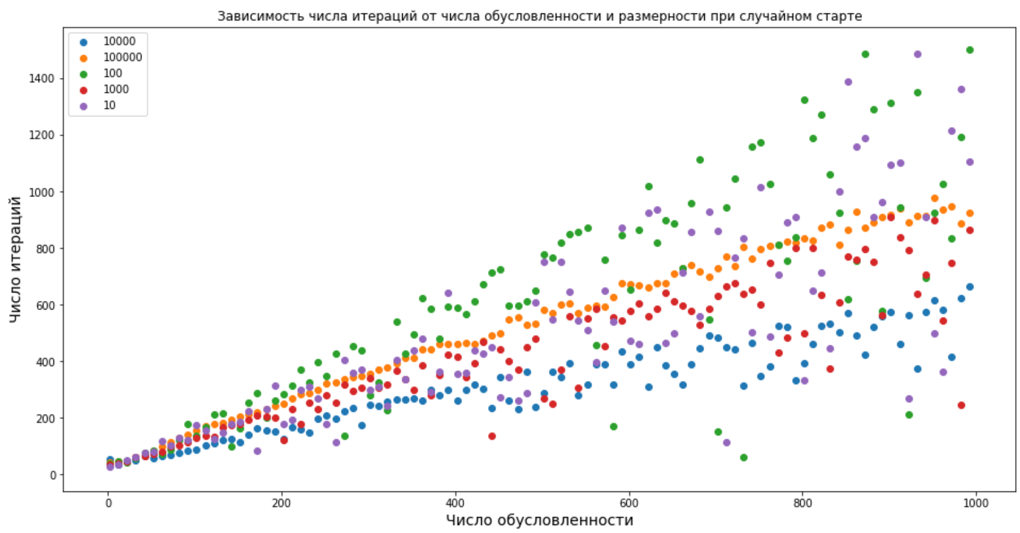
- Константная стратегия.
  Алгоритм может перескакивать через минимум в следствие чего может даже зациклиться или разойтись.

- Армихо.
  Гарантируется оптимизация по функции, то есть в итоге алгоритм сойдётся к минимуму, однако шаг не ограничен снизу, поэтому алгоритм может очень долго это делать.

- Вульф.
  Как Армихо, но шаг ограничен снизу, поэтому метод сходится быстрее.

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

Исследуем число итераций градиентного спуска на квадратичной функции в зависимости от числа обусловленности и размерности пространства.



На графике видны следующие закономерности:

- Количество итераций пропорционально числу обусловленности.

- С увеличением размерности уменьшается количество итераций и его разброс.