

Лабораторная работа 1:

Методы градиентного спуска и метод Ньютона.

Сливкина Ксения

23 апреля 2023 г.

1 Реализация прототипов функций

Формула логистической регрессии с L_2 регуляризацией:

$$f(x) = \frac{1}{m} \log(1 + \exp(bA^T x)) + \frac{\text{regcoef}}{2} x^T x$$

Формула градиента:

$$\text{Grad}(x) = \frac{1}{m} (A^T (\sigma(Ax) - b)) + \text{regcoef} * I$$

Формула гессиана:

$$H(x) = \frac{1}{m} A^T D A + \text{regcoef} * I$$

где D - диагональная матрица, и $D = \sigma(bA^T x)(1 - \sigma(bA^T x))$

2 Эксперимент: Траектория градиентного спуска на квадратичной функции.

Чтобы ответить на вопрос: как отличается поведение метода в зависимости от числа обусловленности функции, выбора начальной точки и стратегии выбора шага, были проведены несколько экспериментов.

Рассмотрим функцию вида $f(x) = \frac{1}{2}x^T Ax$, где $A = \begin{bmatrix} 2.0 & 3.0 \\ 3.0 & 5.0 \end{bmatrix}$.
 Посмотрим на поведение метода градиентного спуска в зависимости от выбора начальной точки.

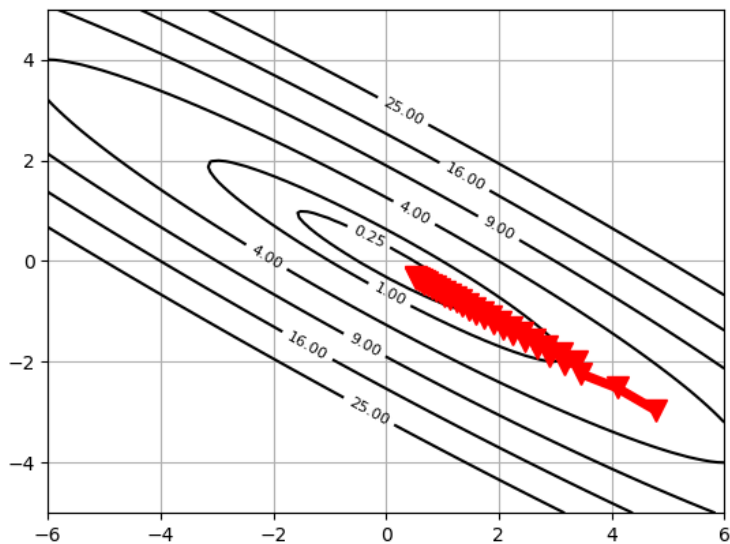


Рис. 1: Начальная точка $[8, 2]$

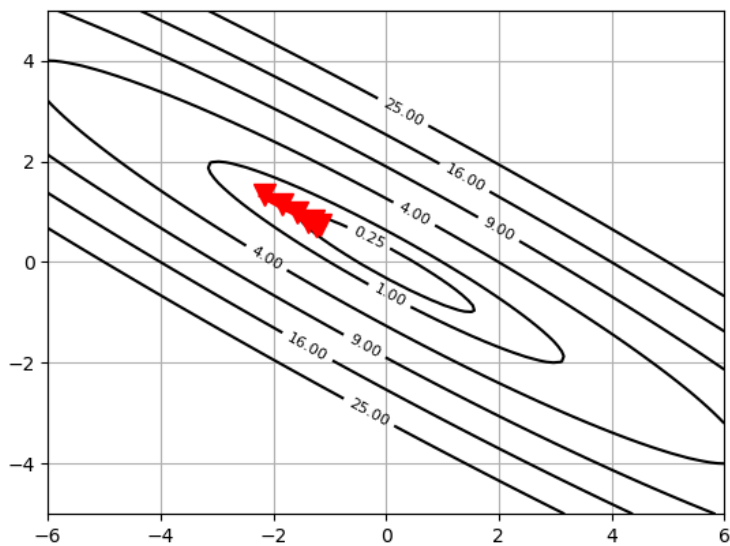


Рис. 2: Начальная точка $[-8, -8]$

Вывод: выбор начальной точки влияет на то, куда сойдется метод. Вместо глобального минимума/максимума он может найти локальный.

Посмотрим на поведение метода градиентного спуска в зависимости от стратегии выбора шага.

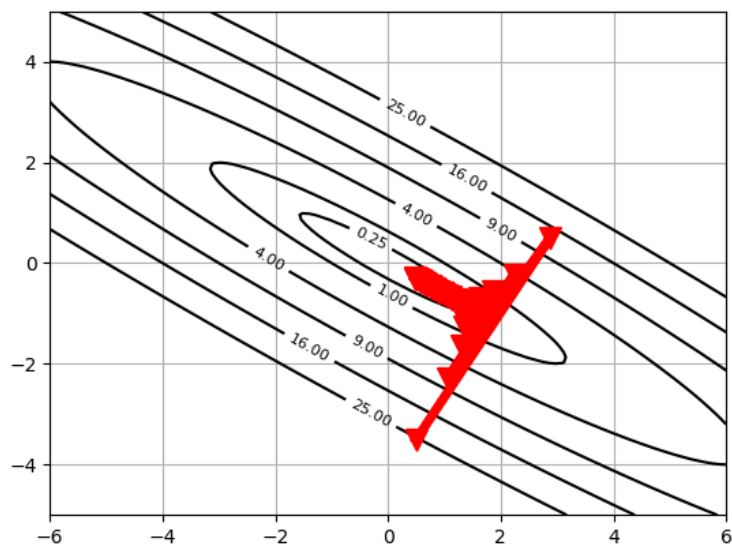


Рис. 3: Стратегии выбора шага - Armijo

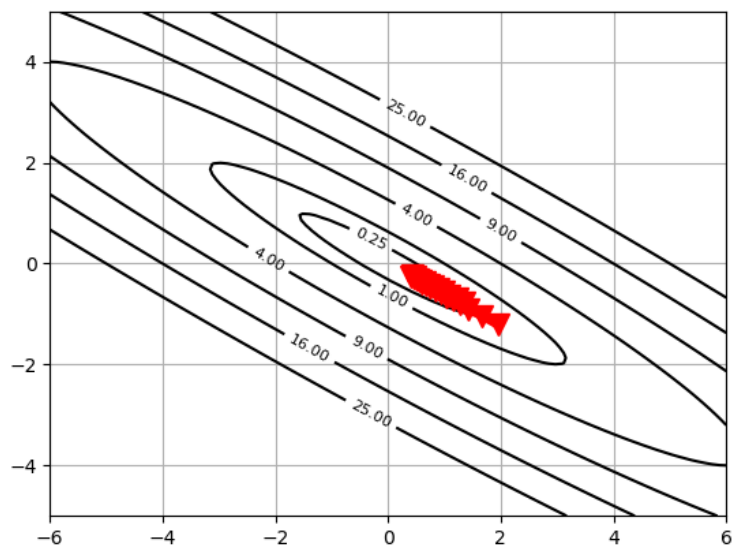


Рис. 4: Стратегии выбора шага - Wolfe

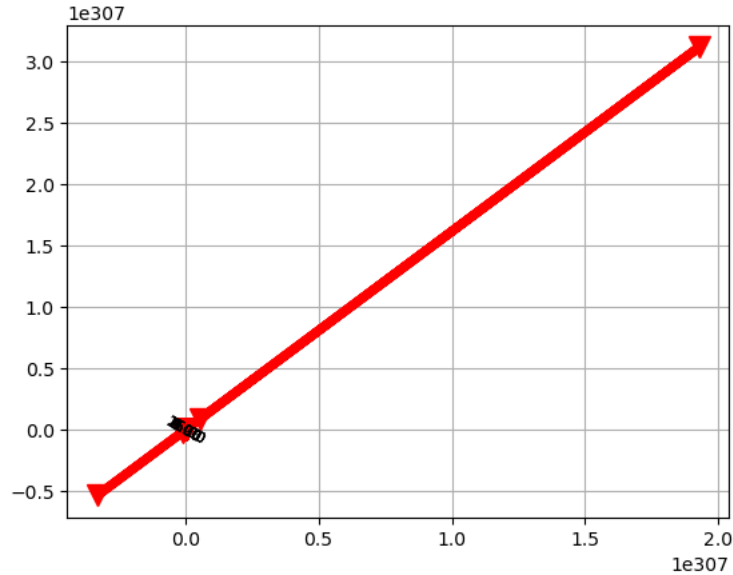


Рис. 5: Стратегии выбора шага - Constant

Вывод: при использовании неподходящей стратегии выбора шага метод может не сойтись, что в данном эксперименте случилось с константным шагом. Можно отметить, что в методе Armijo градиентный спуск в начале делает несколько шагов в разные стороны и только потом начинает двигаться в нужном направлении.

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

3.1 Описание эксперимента.

В данном эксперименте было необходимо исследовать зависимость числа итераций, необходимое градиентному спуску для сходимости, от двух параметров: 1) числа обусловленности оптимизируемой функции и 2) размерности пространства n оптимизируемых переменных.

Были рассмотрены значения размерности пространства n равные 10, 100, 1000 и 10000. Для каждого фиксированного n перебирались числа обусловленности k от 1 до 500 с шагом 10 и строился график зависимо-

сти $T(k, n)$ от k . Так как каждый эксперимент необходимо было повторить несколько раз, то также генерировался массив различных random seed от 1 до 500 с шагом 10 и при повторении эксперимента выбиралось новое random seed.

В качестве матрицы A берется диагональная матрица $A = \text{Diag}(a)$ (создается через `scipy.sparse.diags`), у которой диагональные элементы сгенерированы случайно в пределах $[1, k]$, причем $\min(a) = 1$, $\max(a) = k$ (1-й и 2-й элементы сгенерированного массива вручную заменены на 1 и k , поэтому даже если данные числа не были сгенерированы случайно, то 1 и k все равно станут соответственно минимумом и максимумом). Вектор b заполняется случайными числами (`np.random.rand`).

3.2 Результаты.

По полученному графику можно сделать следующие выводы: чем больше размерности пространства, тем сильнее разброс в количестве итераций, необходимых для сходимости; также в среднем у квадратичной задачи большего размера значение числа итераций до сходимости немного выше.

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

4.1 Описание эксперимента.

В данном эксперименте было необходимо сравнить методы градиентного спуска и Ньютона на задаче обучения логистической регрессии с реальными данными.

Коэффициент регуляризации взят стандартным образом: $\lambda = \frac{1}{m}$. Параметры обоих методов - по умолчанию, начальная точка $x_0 = 0$.

В качестве матрицы A берется диагональная матрица $A = \text{Diag}(a)$ (создается через `scipy.sparse.diags`), у которой диагональные элементы сгенерированы случайно в пределах $[1, k]$, причем $\min(a) = 1$, $\max(a) = k$ (1-й и 2-й элементы сгенерированного массива вручную заменены на 1 и k , поэтому даже если данные числа не были сгенерированы случайно, то 1

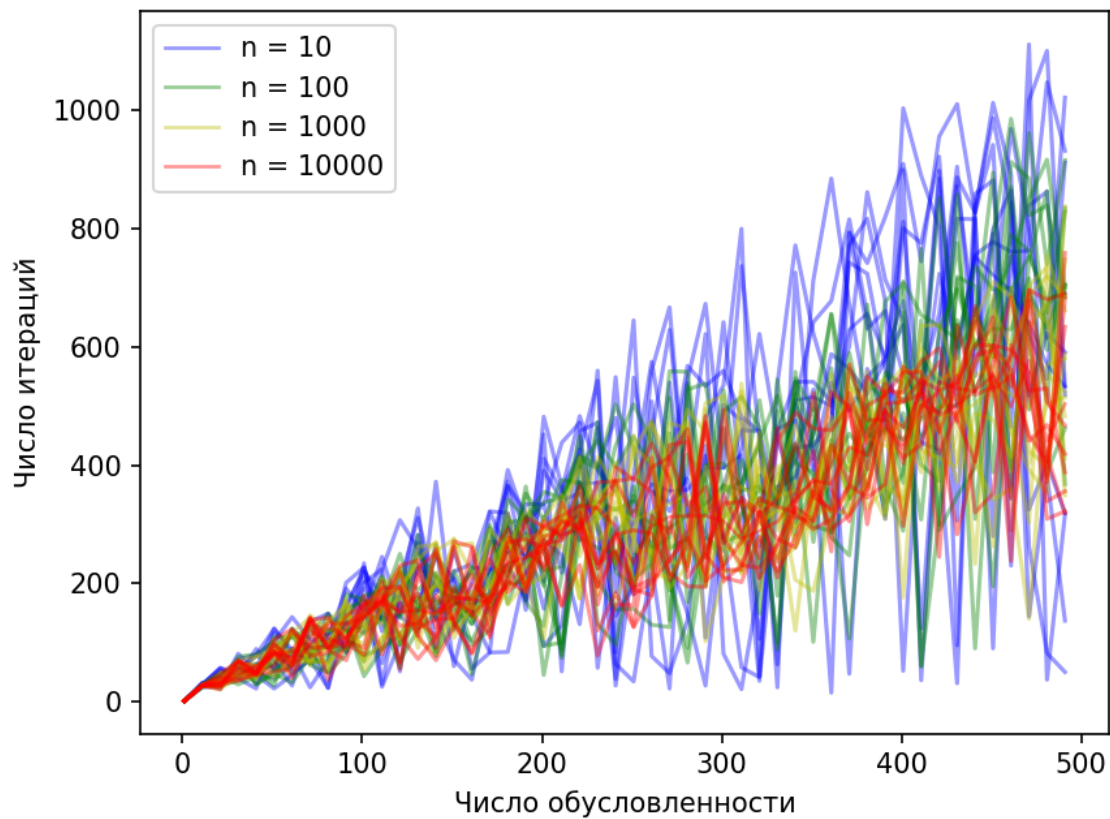


Рис. 6: Зависимость числа итераций от числа обусловленности и размерности пространства

и k все равно станут соответственно минимумом и максимумом). Вектор b заполняется случайными числами (`np.random.rand`).

4.2 Результаты.

Графики для набора данных w8a.

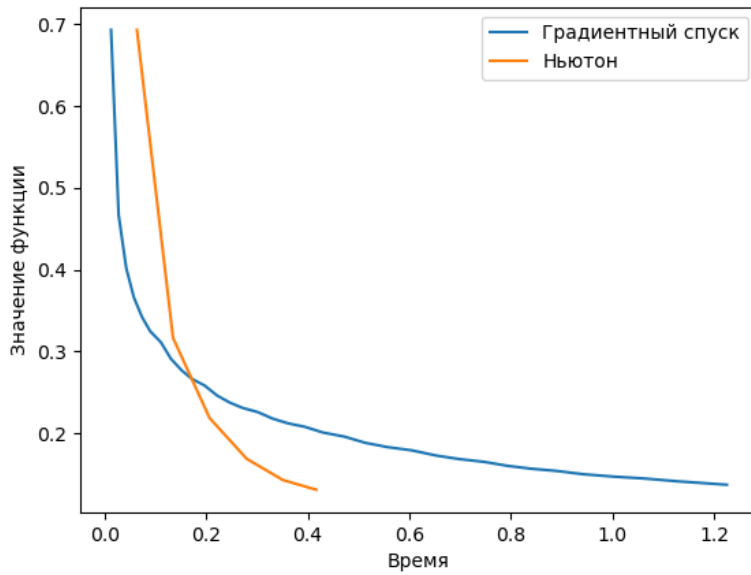


Рис. 7: Значение функции от времени работы

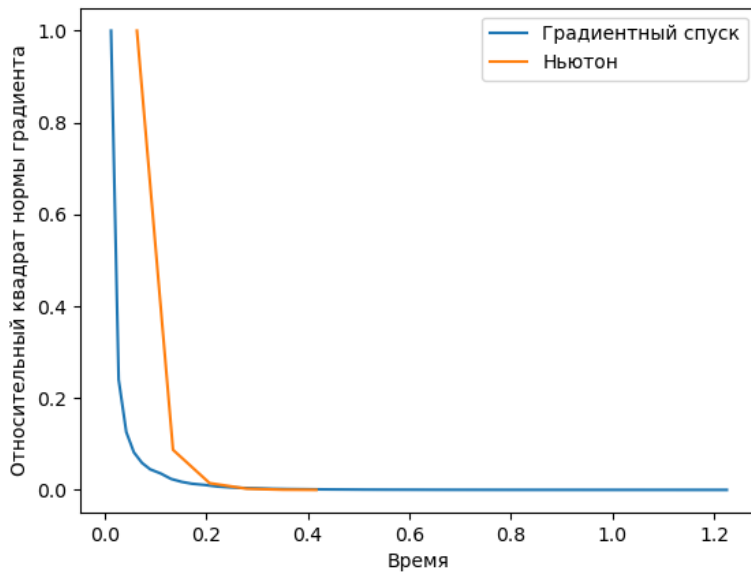


Рис. 8: Относительный квадрат нормы градиента от времени работы

n - размерность выборки, m - размерность пространства.

Оценка стоимости итерации для градиентного спуска: $O(mn)$ на итерацию.

Оценка стоимости итерации для метода Ньютона: $O(mn^2 + n^3)$ на итерацию.

По Рис.7 можно сделать вывод, что градиентный спуск сходится значительно медленнее, чем метод Ньютона. Однако из графика зависимости относительного квадрата нормы градиента от реального времени работы видно, что градиентный спуск хорошо работает, даже если точка x_0 далека от точки сходимости. Таким образом, преимущество градиентного спуска в слабых требованиях к исходным данным.