Ханмурзин Алим Ринатович. МОАД

Лабораторная работа 1: Методы градиентного спуска и метод Ньютона.

<u>Задание 1</u>. Получить формулы для градиента и гессиана функции логистической регрессии

<u>Градиент</u>:

$$f(x) = \frac{1}{m}\sum_{i=1}^{m}\ln(1+\exp(-b_i\langle a_i, x\rangle)) + \frac{\lambda}{2}\|x\|_2^2 \to min$$

$$[Df(x)](h) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{1+\exp(-b_i\langle a_i, x\rangle)}(-b_i(\langle[Da_i](h), x\rangle + \langle a_i, [Dx](h)\rangle)\exp(-b_i\langle a_i, x\rangle) +$$

$$+\frac{\lambda}{2}(\langle[Dx](h), x\rangle + \langle x, [Dx](h)\rangle) = \frac{-1}{m}\sum_{i=1}^{m}\frac{1}{1+\exp(b_i\langle a_i, x\rangle)}b_i(\langle a_i, h\rangle) +$$

$$+\lambda\langle x, h\rangle = \frac{-1}{m}\sum_{i=1}^{m}\frac{b_i a_i}{1+\exp(b_i\langle a_i, x\rangle)} + \lambda x$$

В матричной форме:

$$\nabla f(x) = [Df(x)](h) = -\frac{1}{m}A^T b \odot \frac{1}{(1+\exp(b\odot Ax))} + \lambda x$$

<u>Гессиан</u>:

$$[D\nabla f(x)](h) = [-\frac{1}{m}\sum_{i=1}^{m}(1+\exp(b_i\langle a_i, x\rangle))^{-1}b_i a_i + \lambda x](h)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\frac{\exp(b_i\langle a_i, x\rangle)\, b_i^2 a_i^2}{(1+\exp(b_i\langle a_i, x\rangle))^2} + \lambda$$

$$= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{1+\exp(b_i\langle a_i, x\rangle)}\left(1 - \frac{1}{1+\exp(b_i\langle a_i, x\rangle)}\right)b_i^2 a_i^2 + \lambda$$

В матричной форме:

$$\nabla^2 f(x) = [D\nabla f(x)](h) = \frac{1}{m}A^T\frac{1}{1+\exp(b\odot Ax)}\left(1 - \frac{1}{1+\exp(b\odot Ax)}\right)A + \lambda I_n$$

3.1 Эксперимент: Траектория градиентного спуска на квадратичной функции

Для эксперимента были подобраны 2 матрицы:
1 матрица с числом обусловленности = 2.27:
$$\begin{pmatrix} 1 & 0.3 \\ 0.3 & 2 \end{pmatrix}$$

2 матрица с числом обусловленности = 11.36:
$$\begin{pmatrix} 0.15 & 0.15 \\ 0.15 & 1.5 \end{pmatrix}$$

3 матрица с числом обусловленности = 38.06:
$$\begin{pmatrix} 0.1 & 0.3 \\ 0.3 & 2 \end{pmatrix}$$

Рис.1. Результаты с использованием 1 матрицы

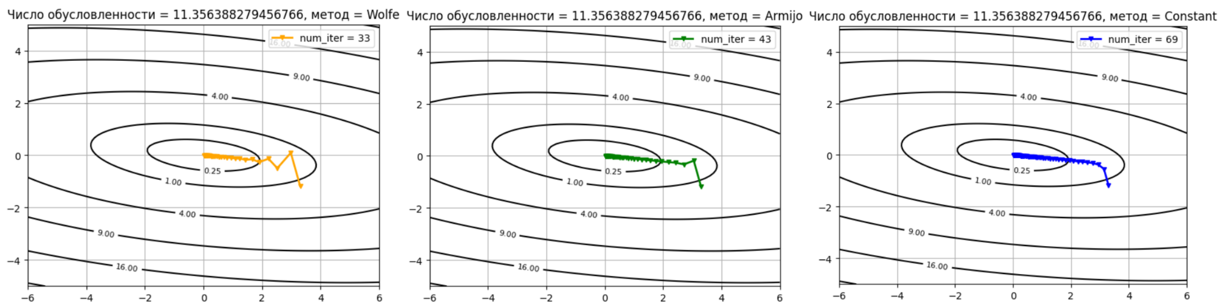

Рис.2. Результаты с использованием 2 матрицы

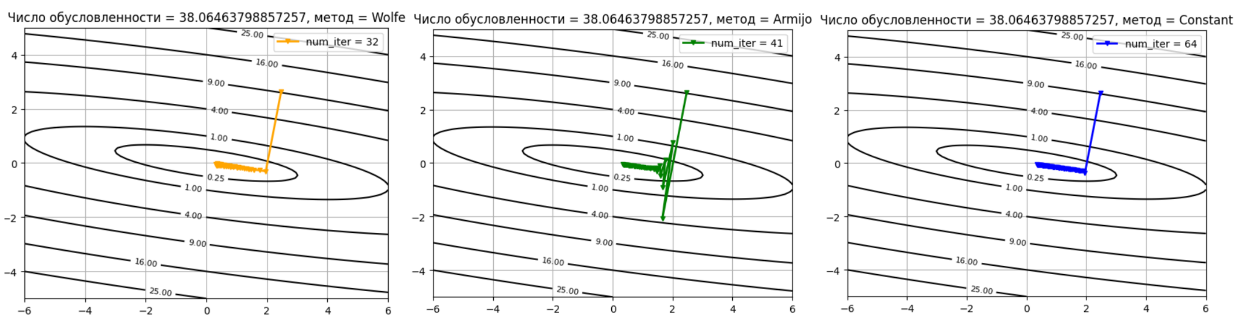

Рис.3. Результаты с использованием 3 матрицы

Вывод: Метод с константным шагом прост в реализации, но требует большего числа итераций, методы Армихо и Вульфа позволяют более эффективно выбирать шаг, основываясь на изменении функции и градиента в каждой точке. Методы Армихо и

Вульфа показывают хорошие результаты, но методу Вульфа требуется меньше число итераций.
Чем выше число обусловленности матрицы, тем сложнее процесс оптимизации. Градиентный метод совершает много маленьких шагов вдоль оси. Высокое число обусловленности приводит к медленной сходимости, а метод с константным шагом может не сойтись к минимуму.
Неправильный выбор начальной точки может привести к зацикливанию метода в локальном минимуме или слишком долгому времени сходимости к оптимальному решению.

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

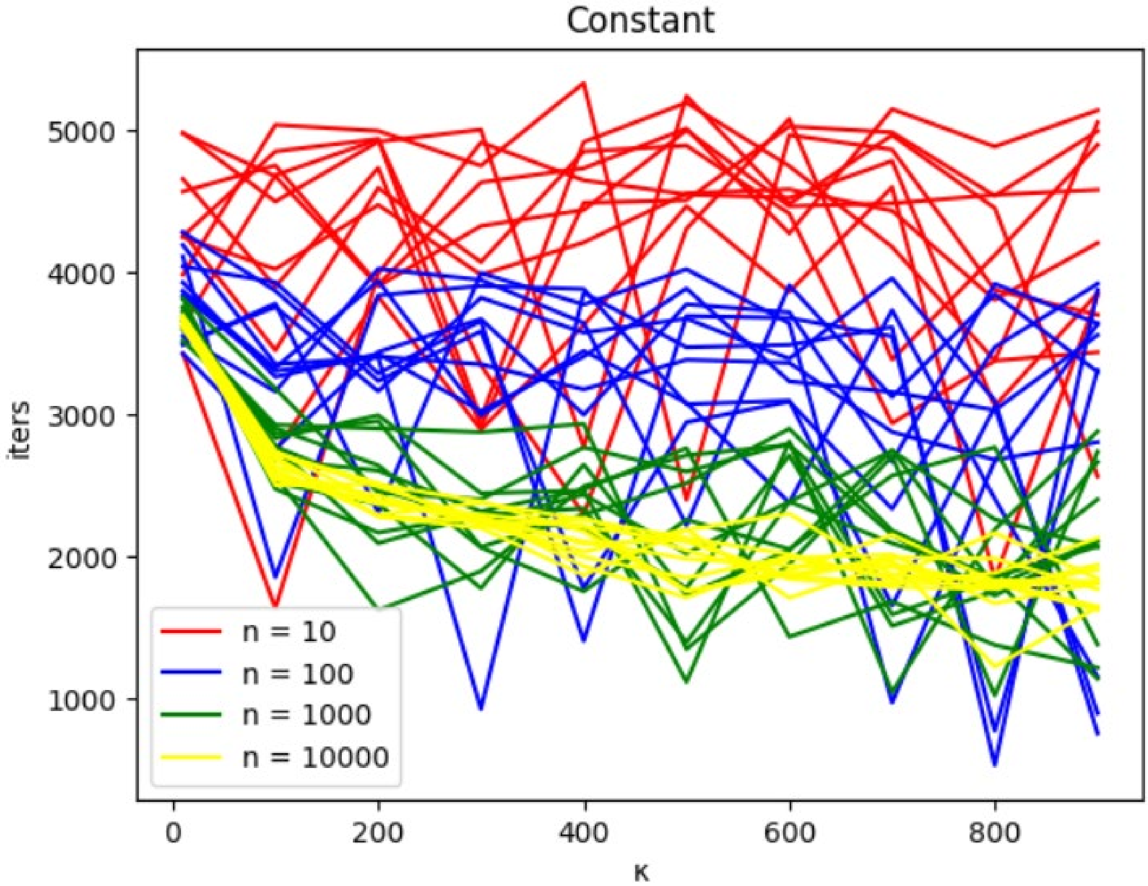

Рис.4. Итерации градиентного спуска зависящие от размерности пространства и числа обусловленности (Константный метод)

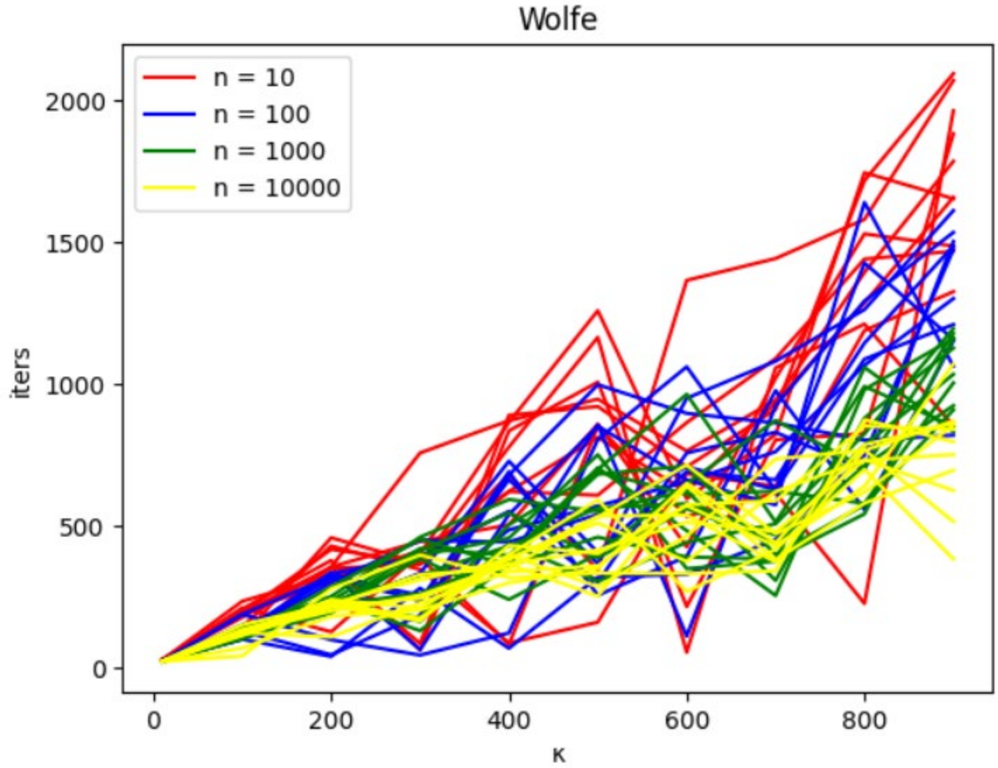

Рис.5. Итерации градиентного спуска зависящие от размерности пространства и числа обусловленности (метод Вульфа)

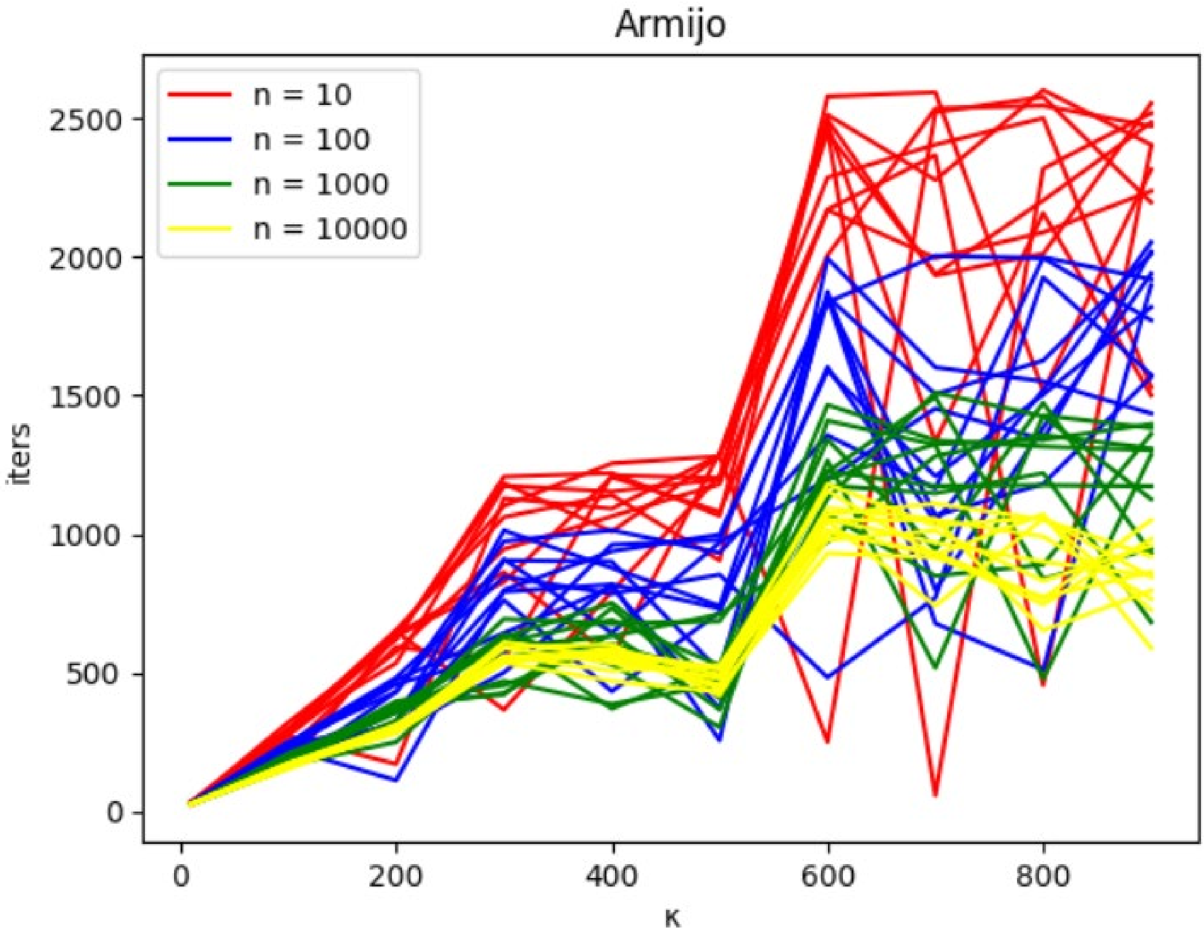

Рис.6. Итерации градиентного спуска зависящие от размерности пространства и числа обусловленности (метод Армихо)

Вывод: Для методов Армихо и Вульфа, чем выше размерность задачи, тем меньше количество итераций. С ростом числа обусловленности растет количество итераций для сходимости. Для константного метода число обусловленности не влияет на количество итераций, но чем больше размерность задачи, тем будет меньше число итераций.

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

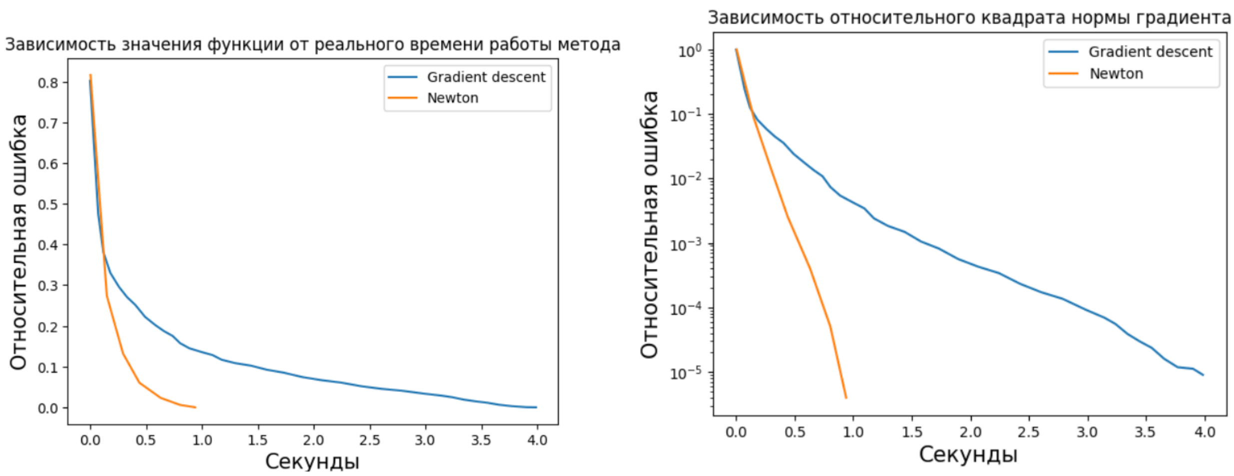

Рис.7. Датасет w8a.txt

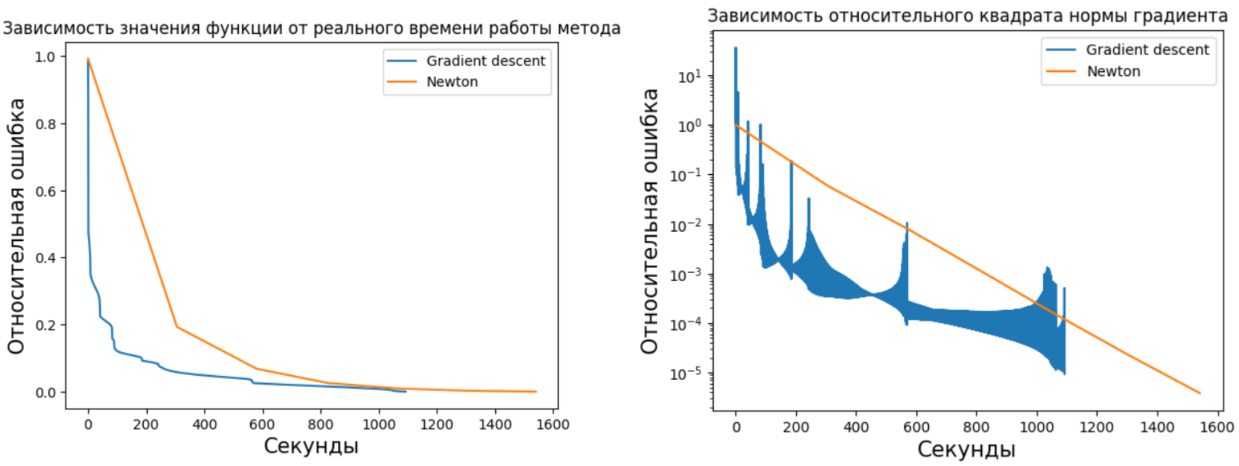

Рис.8. Датасет gisette

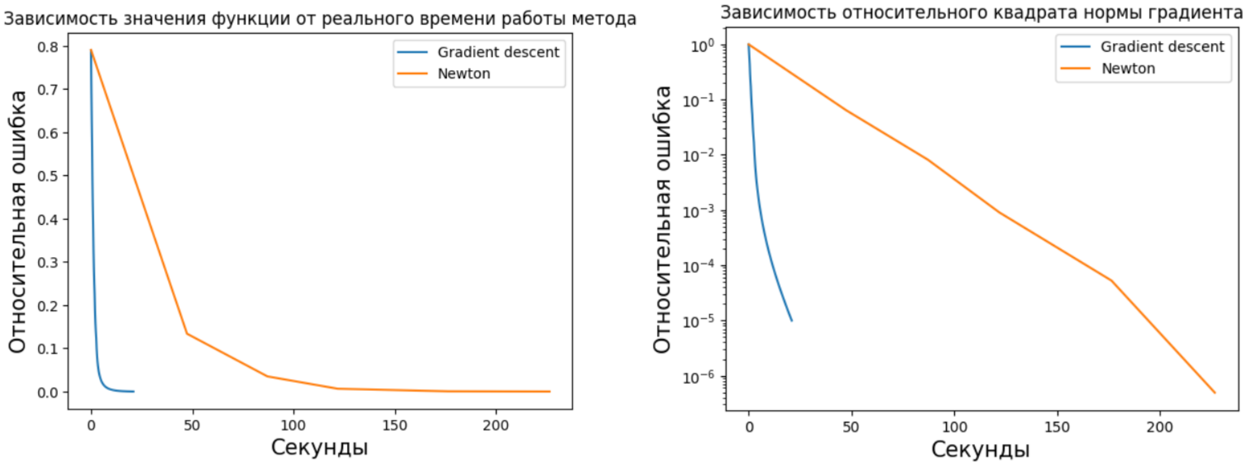

Рис.9. Датасет real-sim

Стоимость итераций для градиентного метода:
Память: нужно хранить значение градиента, следовательно, $O(n)$.

Время: требуется произведение матрицы на число, временная сложность умножения матрицы размером (m × n) на вектор размером (n × 1) считается $O(mn)$, поскольку каждый элемент результирующего вектора размером m × 1 вычисляется путем нахождения скалярного произведения строки матрицы на вектор.

Стоимость итераций для метода Ньютона:
Память: нужно хранить гессиан, следовательно требуется $O(n^2)$.

Время: требуется подсчет гессиана, произведение 3 матриц размерностей (n x m), (m x m), (m x n), умножение матрицы размером n × m на матрицу размером m × m имеет временную сложность $O(m^2n)$. Умножение полученной матрицы размером n × m на матрицу размером m × n имеет временную сложность $O(nm^2)$ $\implies$ временная сложность будет равна $O(2m^2n)$. Итог $O(m^2n)$. Решение системы уравнений $\nabla^2 f(x)d = -\nabla f(x)$ занимает $O(n^3)$, в итоге получаем $O(m^2n) + O(n^3)$.

Вывод: методу Ньютона требуется меньше итераций, но из-за того, что он работает за время $O(m^2n) + O(n^3)$, он не может быстро справиться с большими датасетами, но на маленьких датасетах он работает быстрее. Градиентный метод работает эффективнее на задачах с большими данными.

3.4 Эксперимент: Стратегия выбора длины шага в градиентном спуске

Начальные данные для логистической регрессии: матрица (100 x 80)
Начальные данные для квадратичной функции: диагональная матрица с максимальным числом равным числу обусловленности 10, размерность задачи 10.

Логистическая регрессия (квадрат нормы градиента):

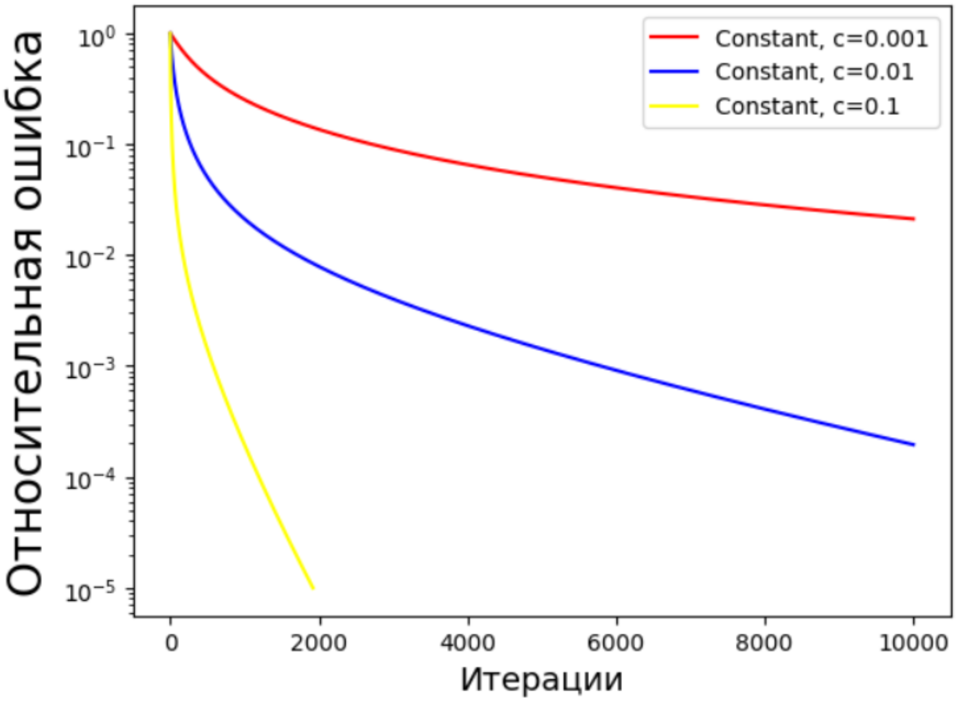

Рис.10. Относительная ошибка для константного метода

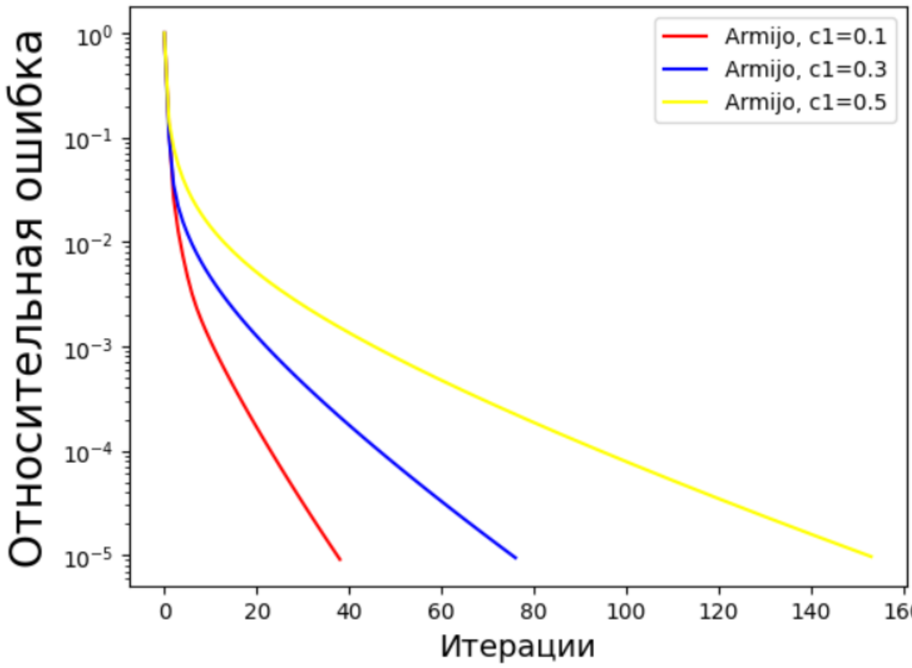

Рис.11. Относительная ошибка для метода Армихо

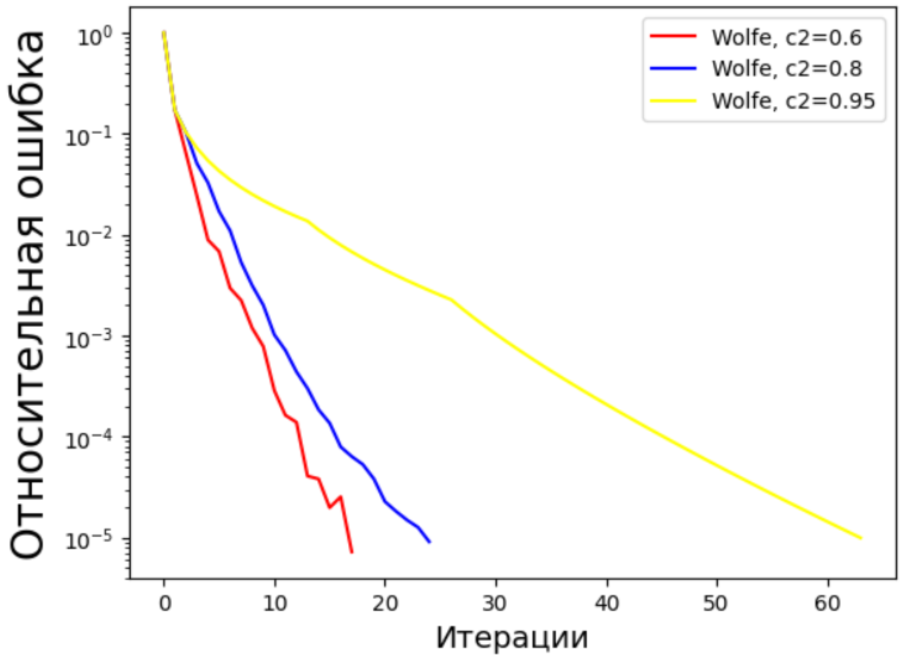

Рис.12. Относительная ошибка для метода Вульфа

Логистическая регрессия (относительная невязка по функции):

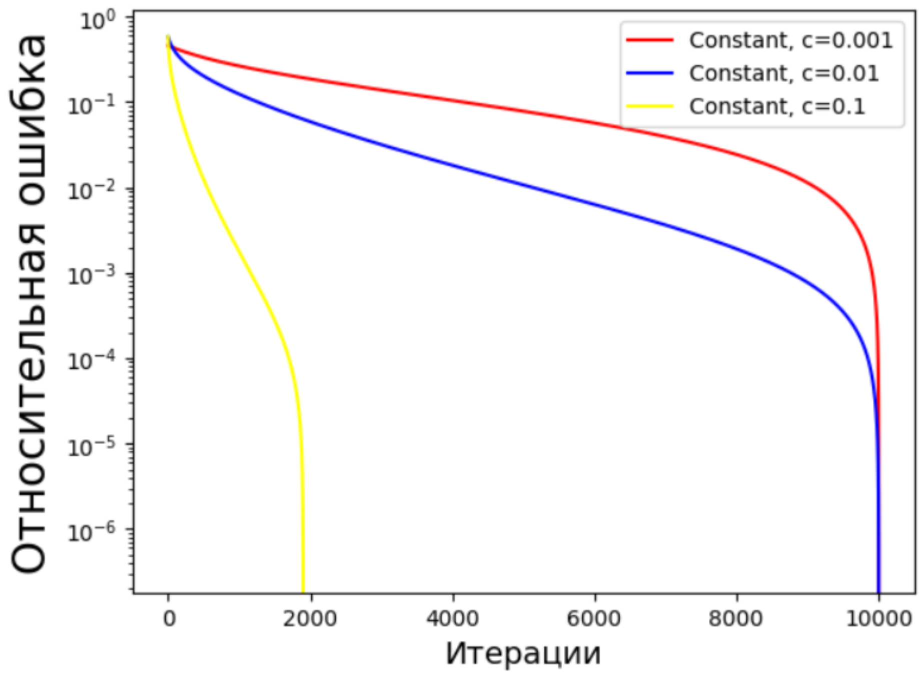

Рис.13. Относительная ошибка для константного метода

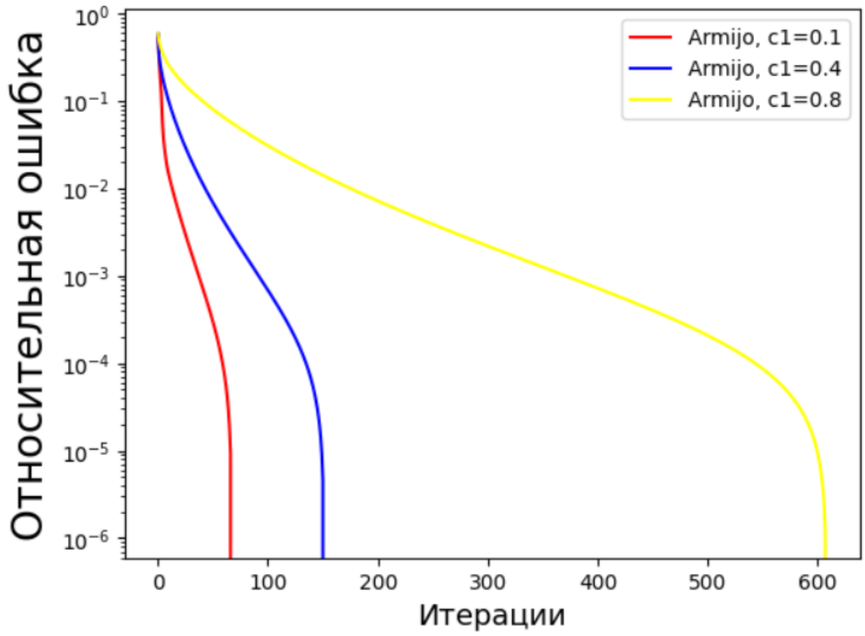

Рис.14. Относительная ошибка для метода Армихо

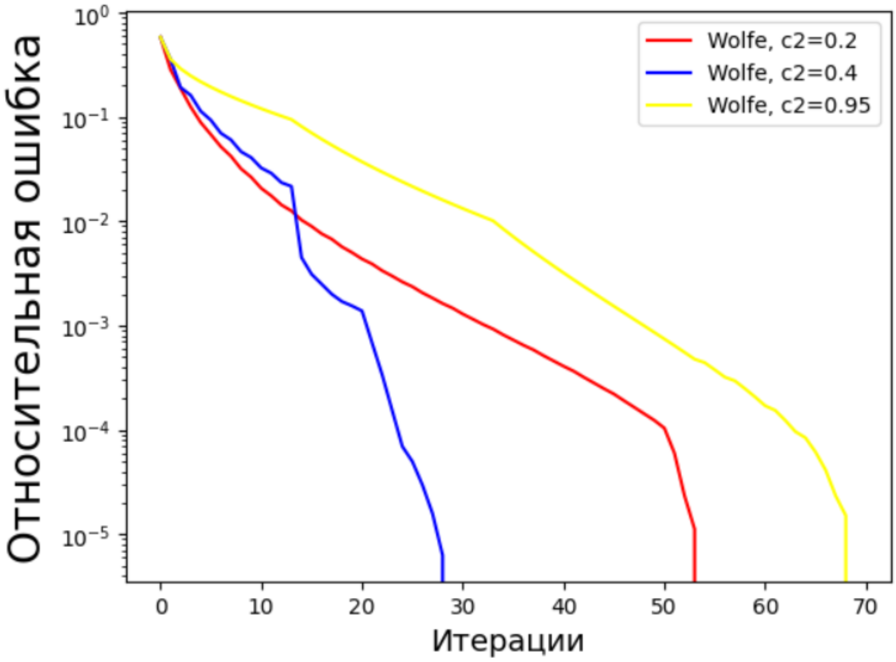

Рис.15. Относительная ошибка для метода Вульфа

Квадратичная функция (квадрат нормы градиента):

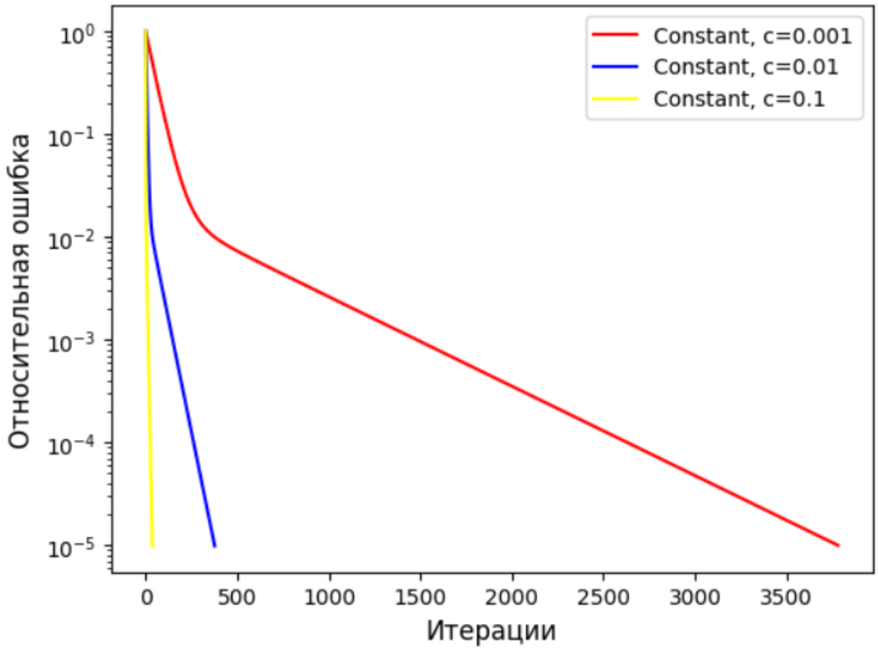

Рис.16. Относительная ошибка для константного метода

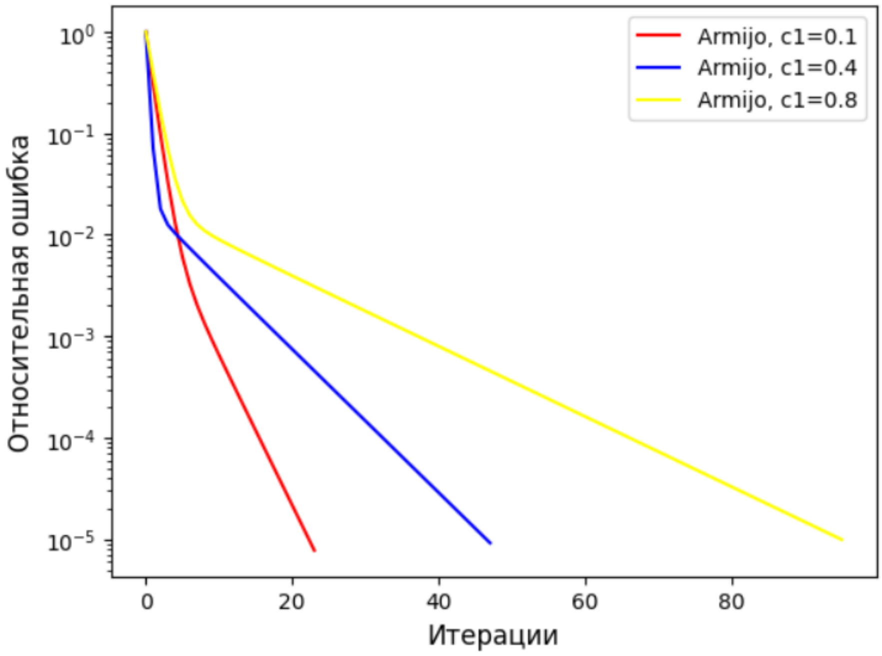

Рис.17. Относительная ошибка для метода Армихо

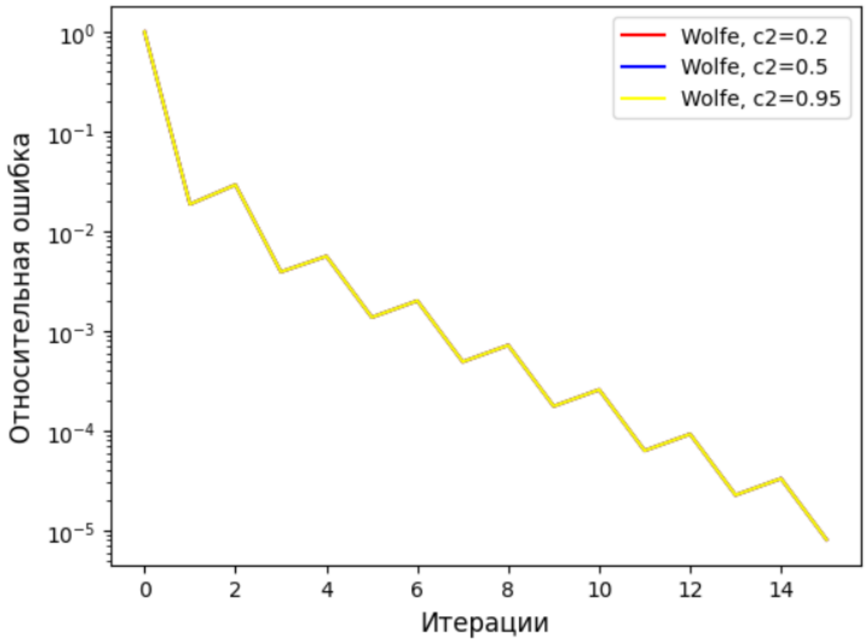

Рис.18. Относительная ошибка для метода Вульфа

Квадратичная функция (относительная невязка по функции):

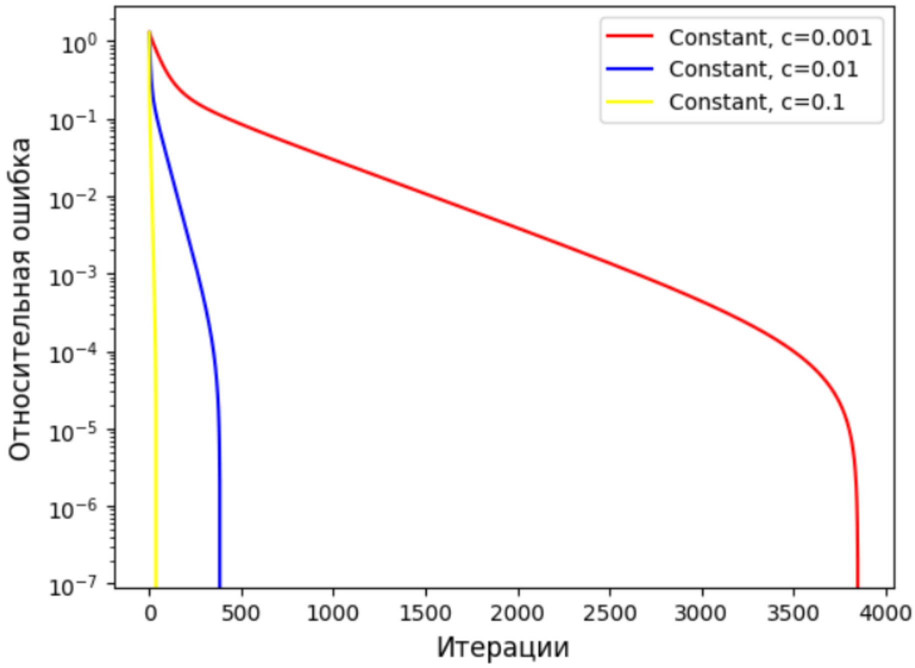

Рис.19. Относительная ошибка для константного метода

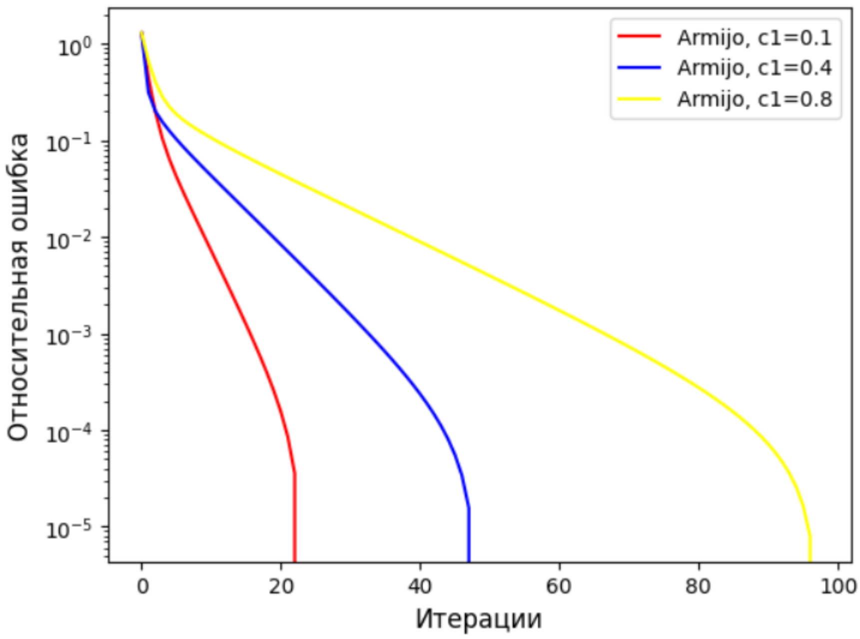

Рис.20. Относительная ошибка для метода Армихо

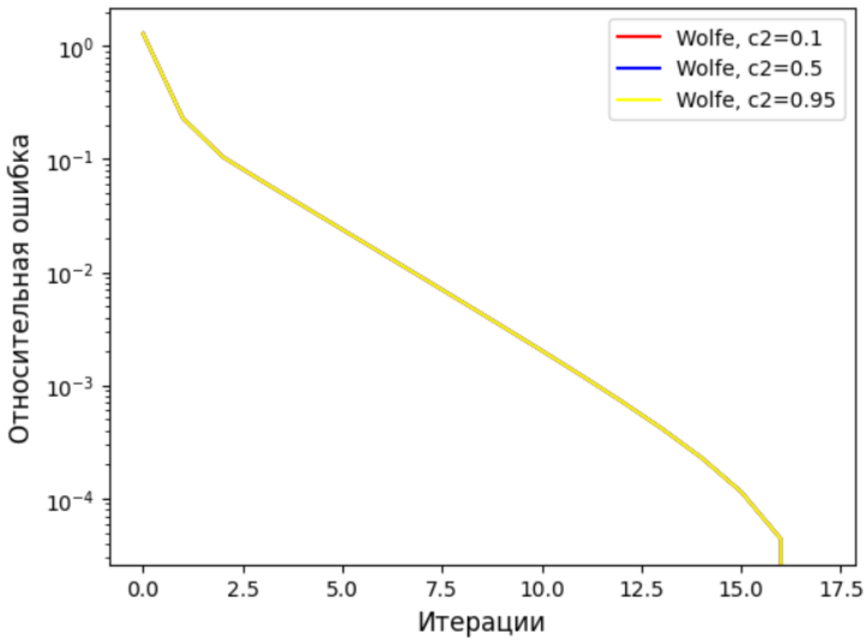

Рис.21. Относительная ошибка для метода Вульфа

Вывод: лучшей стратегией является метод Вульфа, неэффективной же стратегией является константный метод.

3.5 Эксперимент: Стратегия выбора длины шага в методе Ньютона

Начальные данные для логистической регрессии: матрица (100 x 80)
Логистическая регрессия (квадрат нормы градиента):

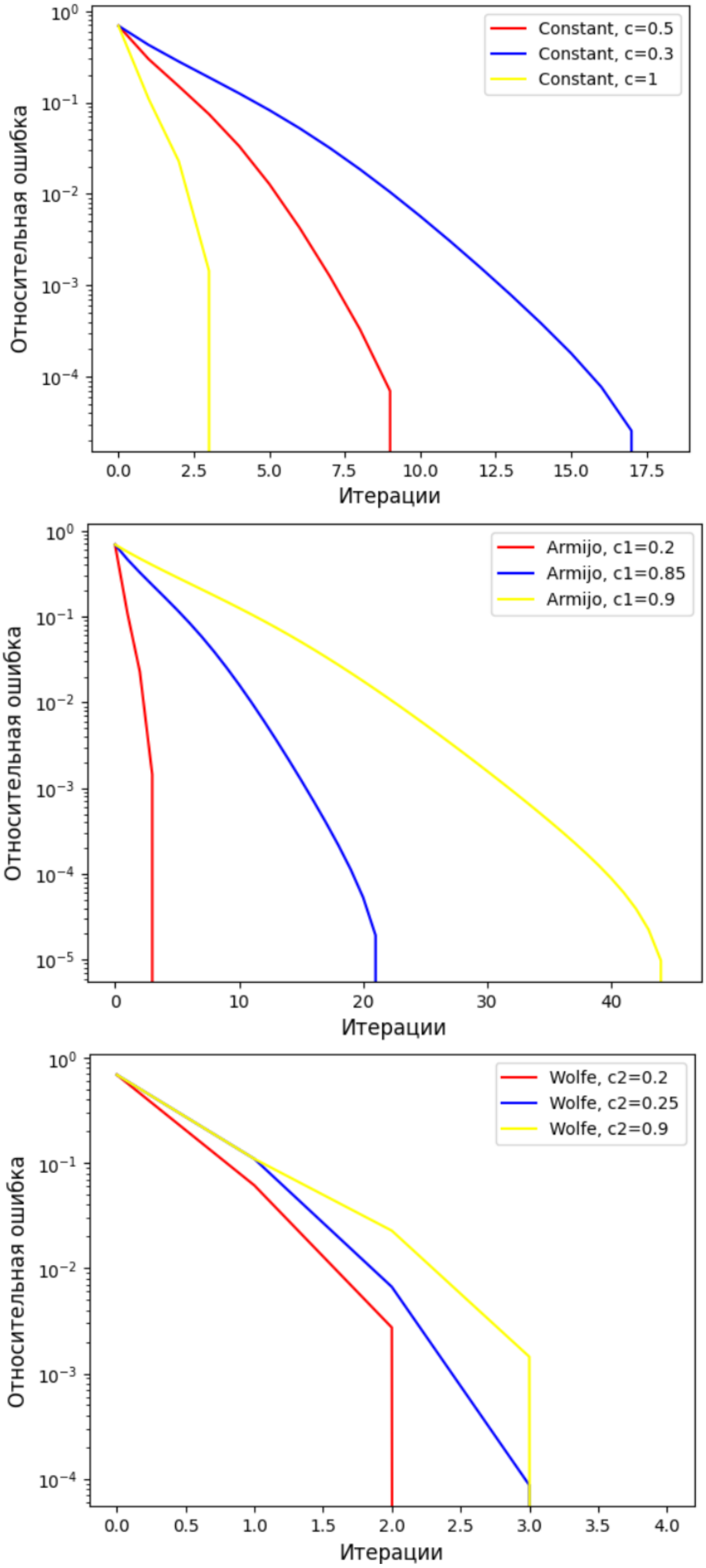

Рис.22. Относительная ошибка квадрат нормы градиента

Логистическая регрессия (относительная невязка по функции):

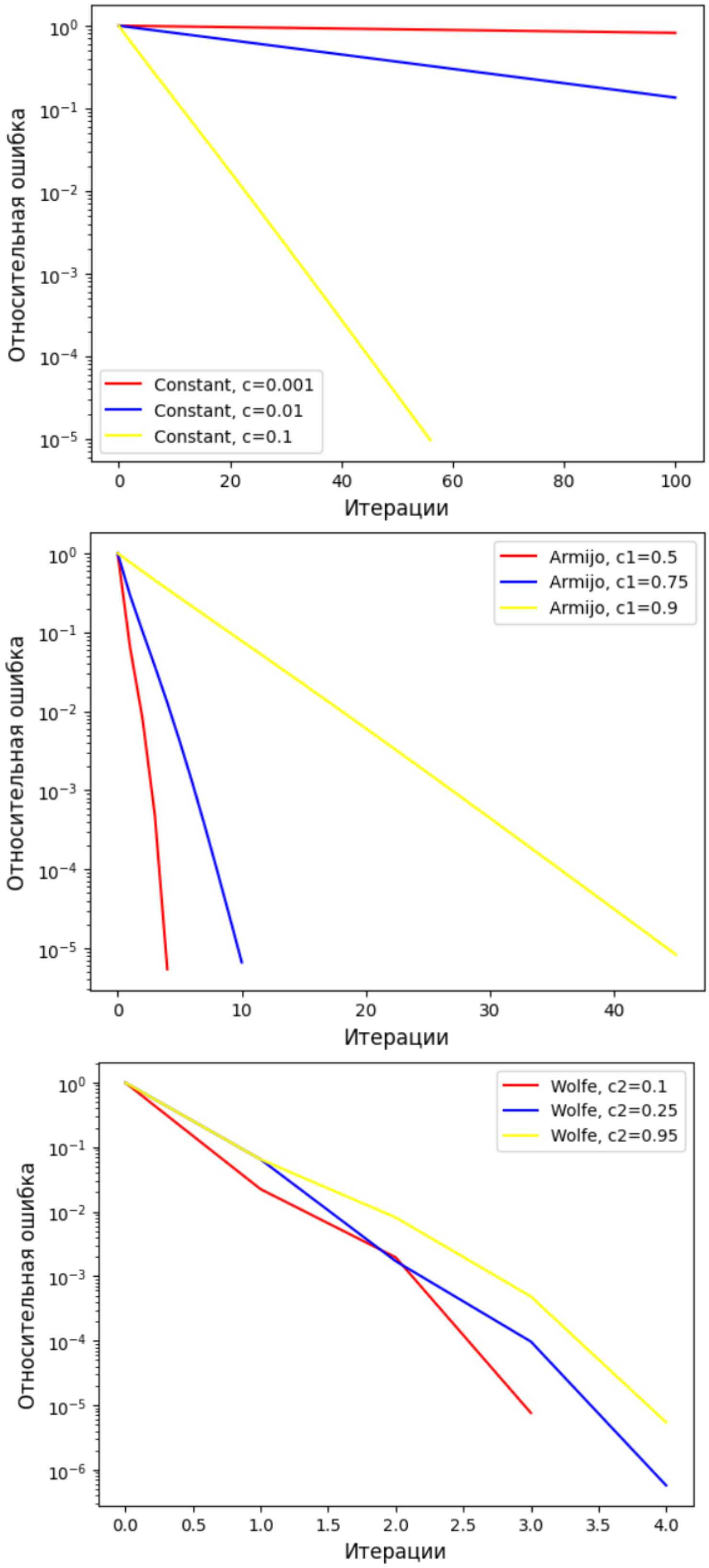

Рис.23. Относительная ошибка невязки по функции

Вывод: как и в случае с градиентным спуском, в методе Ньютона лучшей стратегией является метод Вульфа, неэффективной же остается стратегия константного метода.