Третьякова Д. А.

Лабораторная работа №1

Методы градиентного спуска и метод Ньютона

1. Логистическая регрессия

1.1 Функция логистической регрессии

Для набора данных $X \in R^{n \times d}$, где n — количество образцов, d — количество признаков, и вектора параметров $\theta \in R^{d}$, функция логистической регрессии (или вероятность положительного класса) для всех объектов записывается как:

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \sigma(\mathbf{X}\boldsymbol{\theta})$$

где $\sigma(z)$ — это сигмоида, определенная как:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Функция правдоподобия для логистической регрессии (без явных суммирований):

$$L(\boldsymbol{\theta}) = -\mathbf{y}^{\top} \mathbf{log}\big(\sigma(\mathbf{X}\boldsymbol{\theta})\big) - (\mathbf{1} - \mathbf{y})^{\top} \log\big(1 - \sigma(\mathbf{X}\boldsymbol{\theta})\big)$$

где $\mathbf{y} \in \{0,1\}^{n}$ — вектор меток классов.

1.2 Градиент логистической регрессии

Градиент функции потерь логистической регрессии по параметрам θ (в матрично-векторной форме) выражается как:

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \mathbf{X}^{\top}(\sigma(\mathbf{X}\boldsymbol{\theta}) - \mathbf{y})$$

где $\sigma(\mathbf{X}\boldsymbol{\theta})$ — это поэлементное применение функции сигмоиды ко всем элементам вектора $\mathbf{X}\boldsymbol{\theta}$.

1.3 Гессиан логистической регрессии

Гессиан — это матрица вторых производных функции потерь по параметрам θ. Для логистической регрессии он имеет вид:

$$H(\boldsymbol{\theta}) = \mathbf{X}^{\top\mathbf{S}}\mathbf{X}$$

где $\mathbf{S}$ — это диагональная матрица весов, определенная как:

$$\mathbf{S} = \mathrm{diag}\Big(\sigma(\mathbf{X}\boldsymbol{\theta}) \odot \big(1 - \sigma(\mathbf{X}\boldsymbol{\theta})\big)\Big)$$

Здесь $\odot$ обозначает поэлементное умножение.

1.4 Вывод формул

- Функция логистической регрессии:

$$L(\boldsymbol{\theta}) = -\mathbf{y}^{\top} \mathbf{log}\big(\sigma(\mathbf{X}\boldsymbol{\theta})\big) - (\mathbf{1} - \mathbf{y})^{\top} \log\big(1 - \sigma(\mathbf{X}\boldsymbol{\theta})\big)$$

- Градиент:

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \mathbf{X}^{\top}(\sigma(\mathbf{X}\boldsymbol{\theta}) - \mathbf{y})$$

- Гессиан:

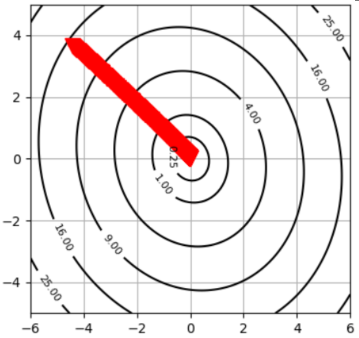
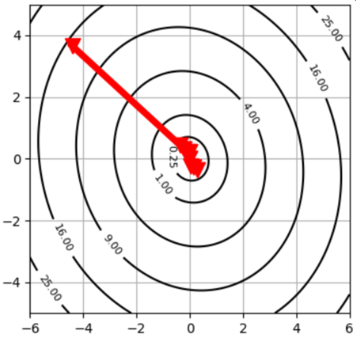
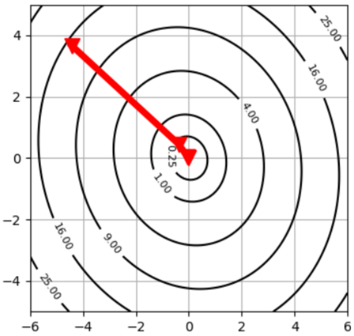
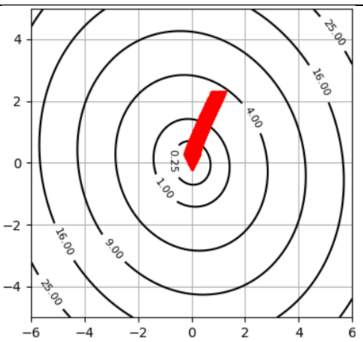
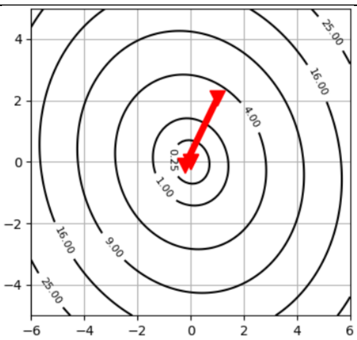
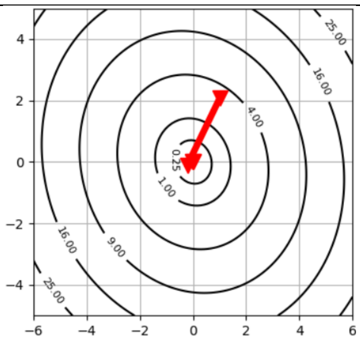
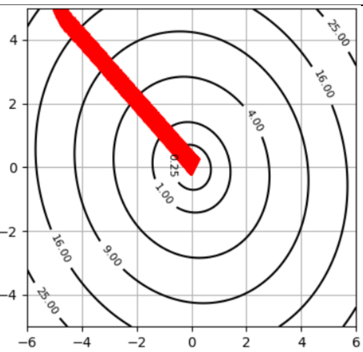
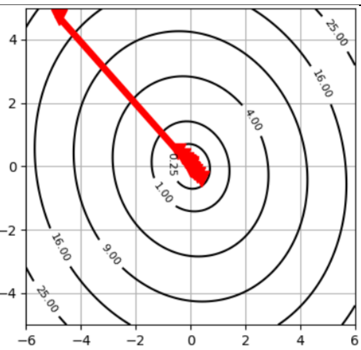
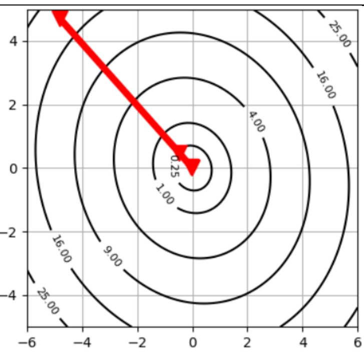
$$H(\boldsymbol{\theta}) = \mathbf{X}^{\top\mathbf{S}}\mathbf{X}), \text{где}(\mathbf{S} = \mathrm{diag}\Big(\sigma(\mathbf{X}\boldsymbol{\theta}) \odot \big(1 - \sigma(\mathbf{X}\boldsymbol{\theta})\big)\Big)$$

2. Эксперимент № 1. Траектория градиентного спуска на квадратичной функции

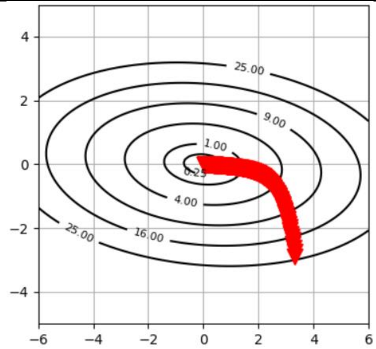
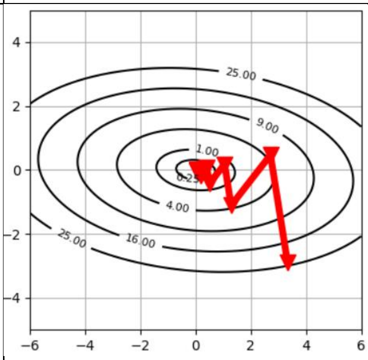
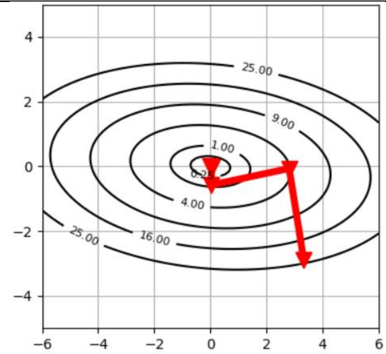
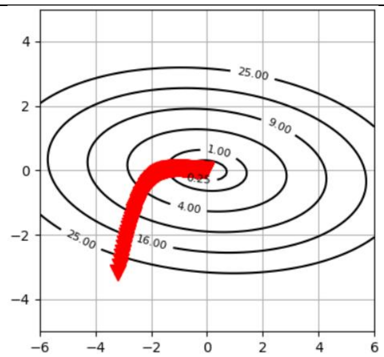
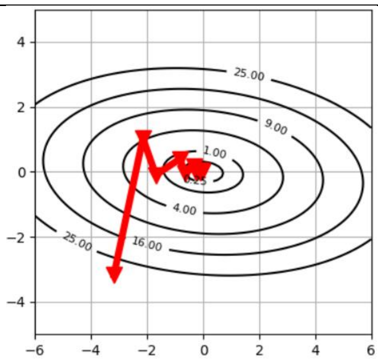
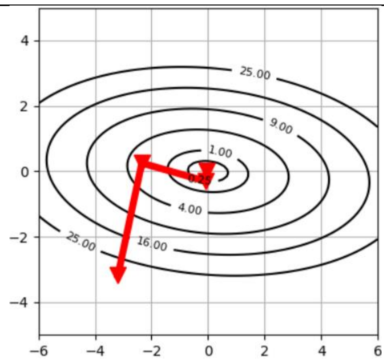
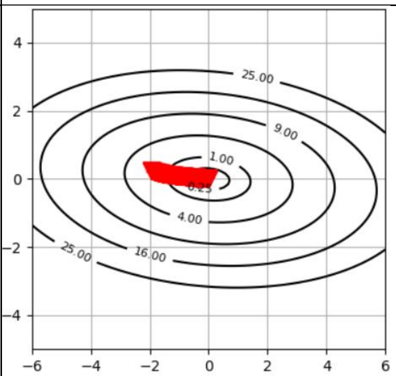
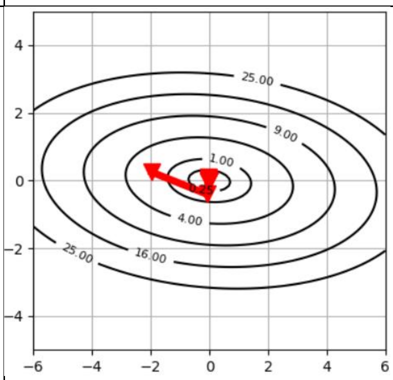
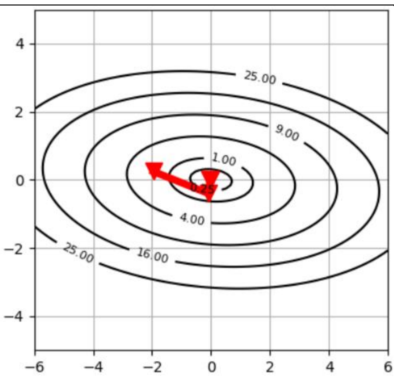
Суть эксперимента заключается в анализе траектории градиентного спуска для квадратичной функции в зависимости от числа обусловленности функции, выбора начальной точки и стратегии выбора шага (константная стратегия, Армихо, Вульф).

Было сгенерировано три числа обусловленности и выбраны три начальные точки.

2.1 Малое число обусловленности 1.22

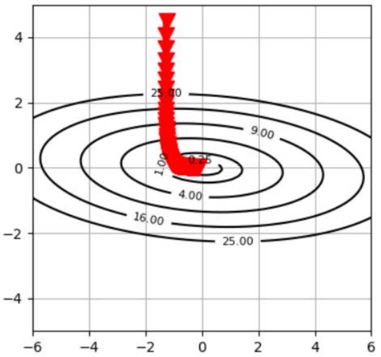
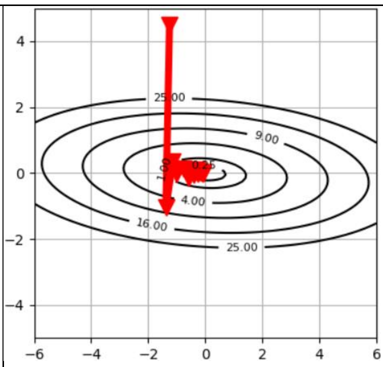
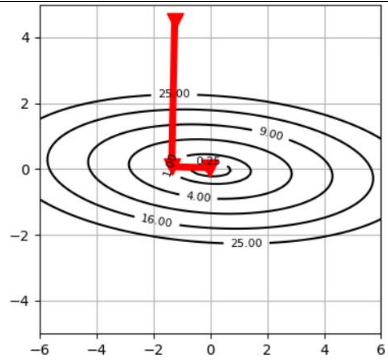
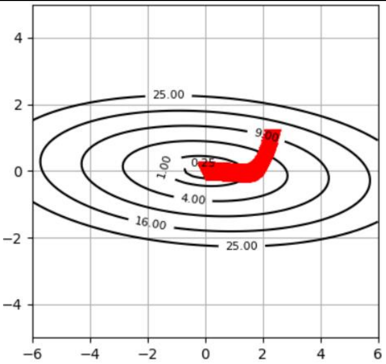
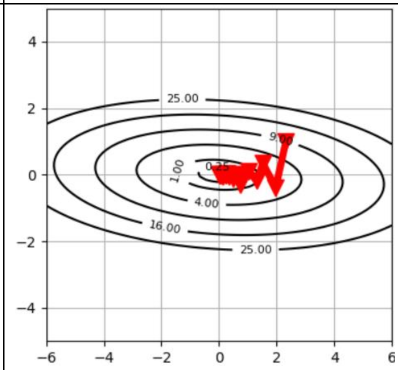
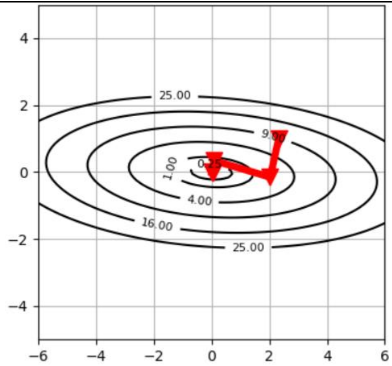
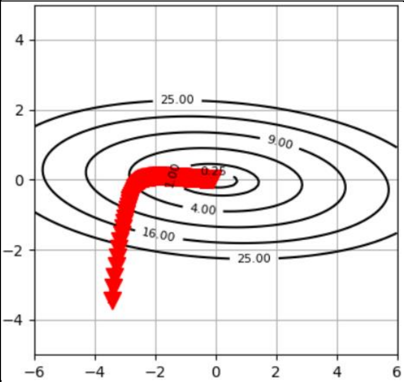
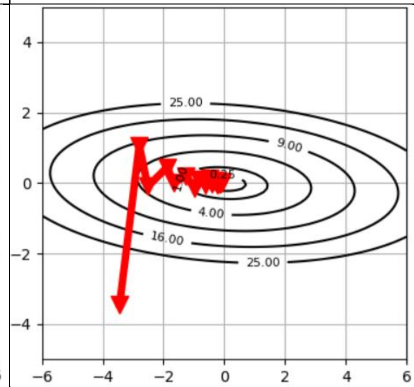
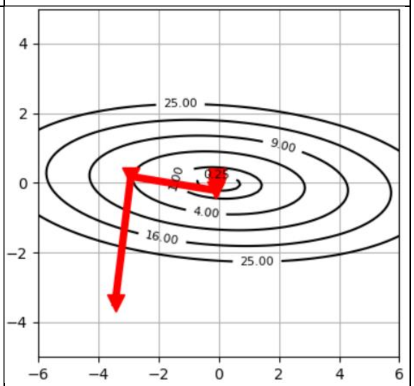
Нач. точка	Стратегия выбора шага		
	Constant	Armijo	Wolfe
0	 Итерации = 638	 Итерации = 10	 Итерации = 4
1	 Итерации = 543	 Итерации = 4	 Итерации = 4
2	 Итерации = 638	 Итерации = 14	 Итерации = 4

2.2 Среднее число обусловленности 5.14

Нач. точка	Стратегия выбора шага		
	Constant	Armijo	Wolfe
0	 Итерации = 448	 Итерации = 12	 Итерации = 5
1	 Итерации = 405	 Итерации = 11	 Итерации = 4
2	 Итерации = 584	 Итерации = 9	 Итерации = 4

2.3 Большое число обусловленности 10.31

Нач. точка	Стратегия выбора шага		
	Constant	Armijo	Wolfe

0	 Итерации = 241	 Итерации = 15	 Итерации = 4
1	 Итерации = 422	 Итерации = 23	 Итерации = 5
2	 Итерации = 340	 Итерации = 19	 Итерации = 6

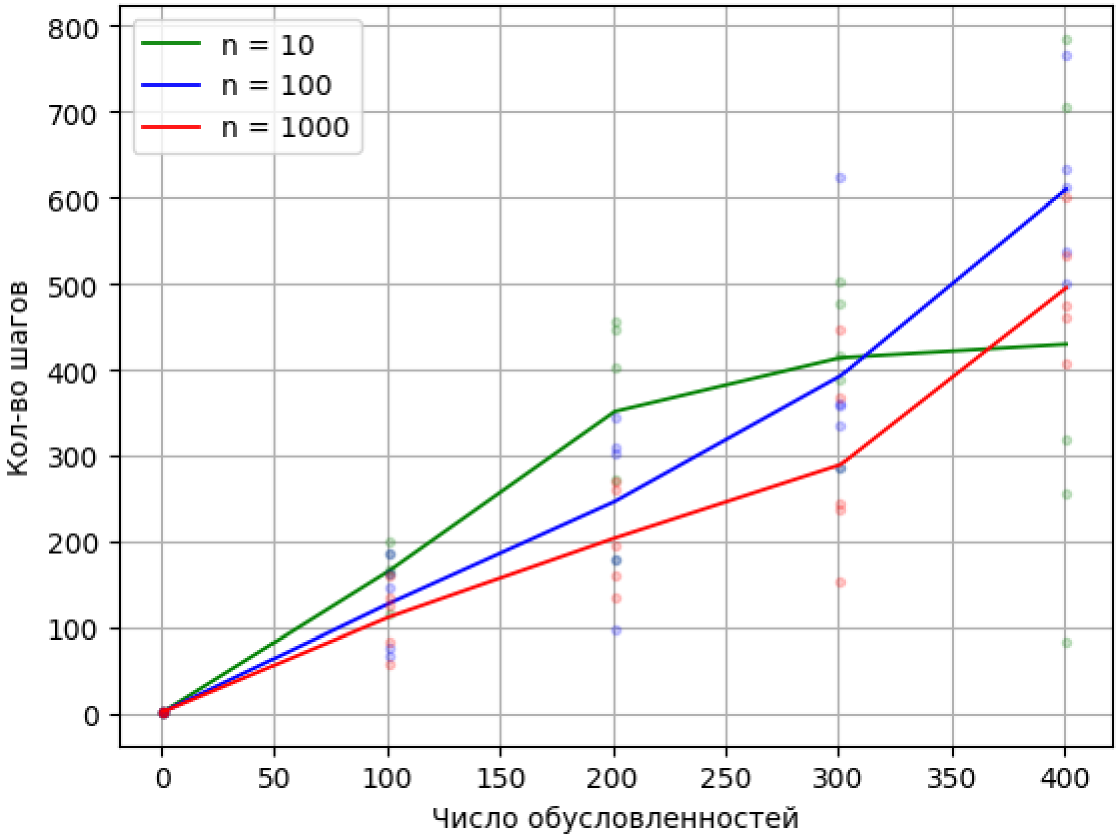
Чем больше число обусловленности матрицы, тем быстрее работает константный метод, однако, по сравнению с другими двумя – гораздо дольше. Для методов Армихо и Вульф рост числа обусловленности работает наоборот – с ростом числа обусловленности растет и количество итераций для этих методов.

При любых выбранных переменных метод Вульфа справляется быстрее, чем Армихо и константный, количество итераций метода Вульфа почти не меняется или меняется не сильно. Константный метод же наоборот – работает очень долго.

3. Эксперимент № 2. Зависимость числа итераций градиентного спуска от числа обусловленности и размерности пространства

Эксперимент заключается в исследовании зависимости числа итераций градиентного спуска от числа обусловленности $\kappa \geq 1$ оптимизируемой функции и размерности пространства n.

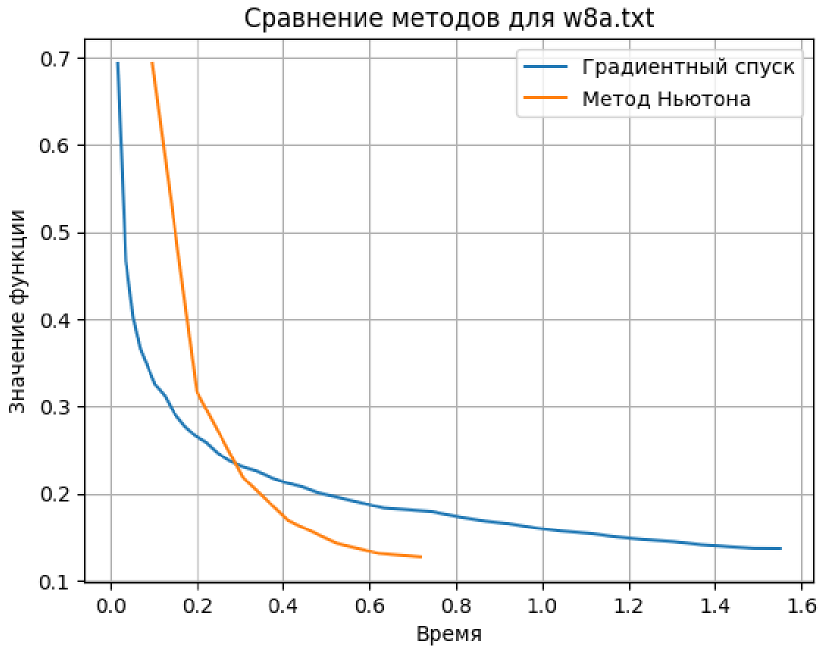
Числа обусловленности от 1 до 1000 с шагом 50.
Размерность пространства 10, 100, 1000.



На графике усредненные значения случайных генераций. С ростом числа обусловленности количество итераций градиентного спуска растет. Для большей размерности пространства градиентный спуск работает быстрее при числе обусловленности менее 350.

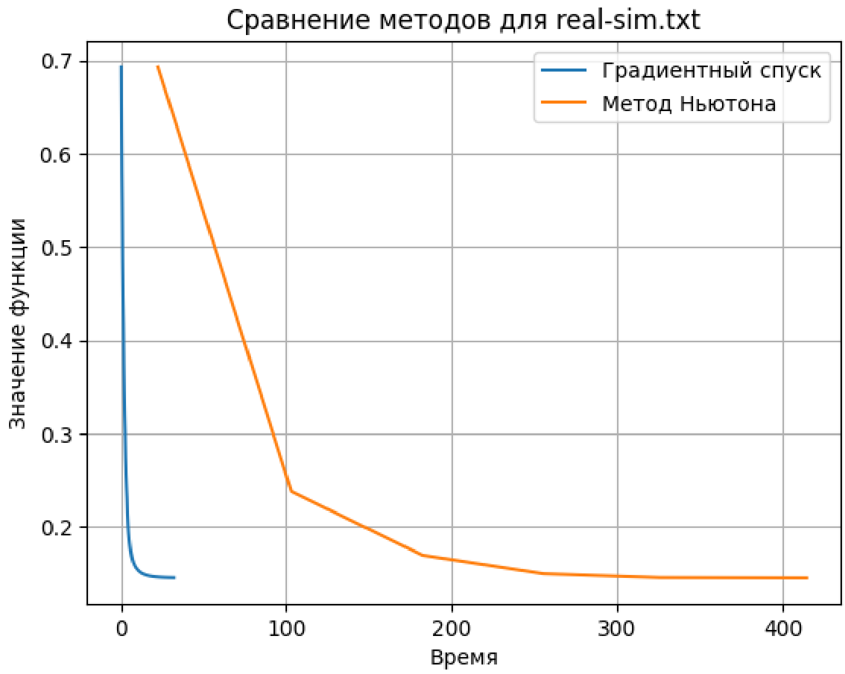
4. Эксперимент № 3. Сравнение методов градиентного спуска и Ньютона на реальной задаче логистической регрессии

Размерность датасета w8a: (49749, 300).



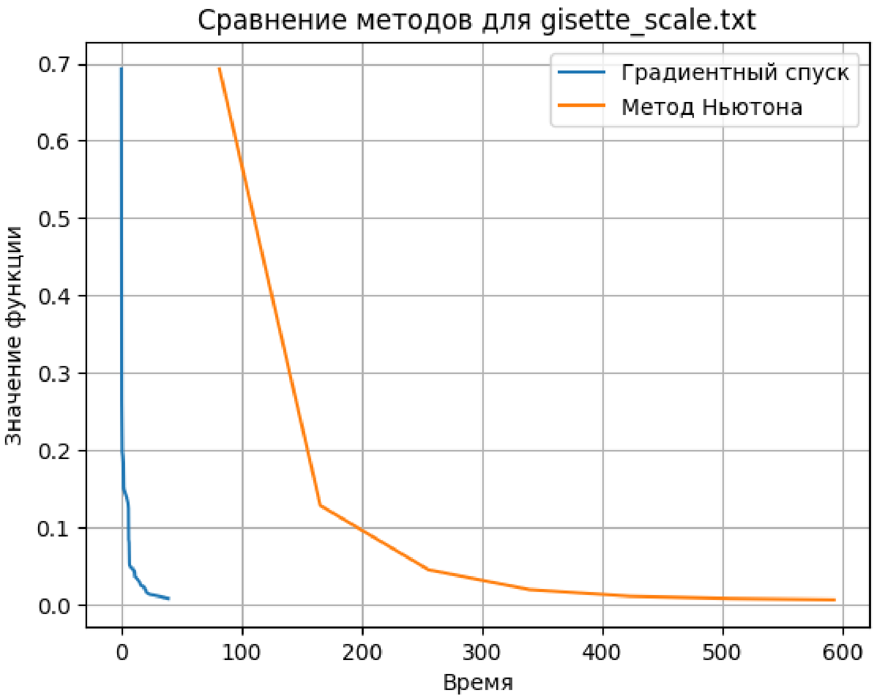
Метод градиентного спуска на начальных этапах быстрее по времени, однако его темп замедляется по мере приближения к оптимальному решению. Метод Ньютона, наоборот, имеет более медленное начало, но за счет более крупных шагов достигает решения быстрее на последних этапах.

Размерность датасета real-sim: (72309, 20958).



Размерность датасета gisette_scale: (6000, 5000).

.

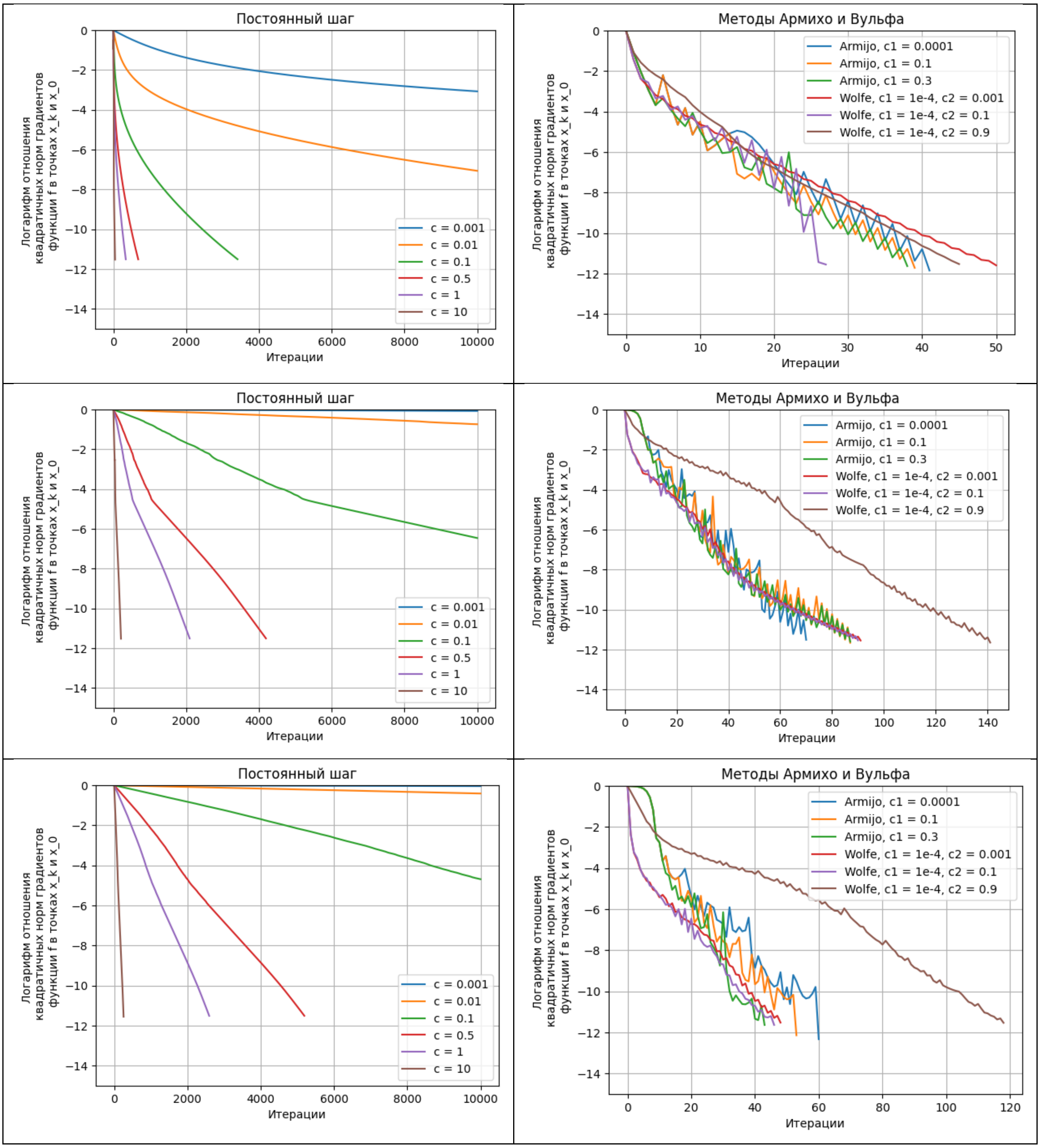


Градиентный спуск лучше работает на задачах с высокой размерностью пространства и плотными данными, где стоимость каждой итерации является критичным фактором.

Метод Ньютона, хотя требует больше ресурсов на каждую итерацию, показывает высокую эффективность в задачах с меньшей размерностью пространства или на разреженных данных.

5. Эксперимент № 4. Стратегия выбора длины шага в градиентном спуске

Исследование зависимости поведения метода от стратегии подбора шага на логистической регрессии.



Из графиков видно, что градиентный спуск сильно зависит от выбора шага: при большом шаге алгоритм сходится быстрее, при маленьком — может не сойтись вовсе. Для методов с постоянным шагом важно точно подобрать его значение, иначе есть риск не сойтись. С шагами по методам Армихо и Вульфа алгоритм стабильно сходился, независимо от начальных условий, и различия в результатах минимальны.

Для задач логистической регрессии предпочтительнее использовать методы Армихо или Вульфа, так как они автоматически подбирают шаг, что снижает риск расхождения.