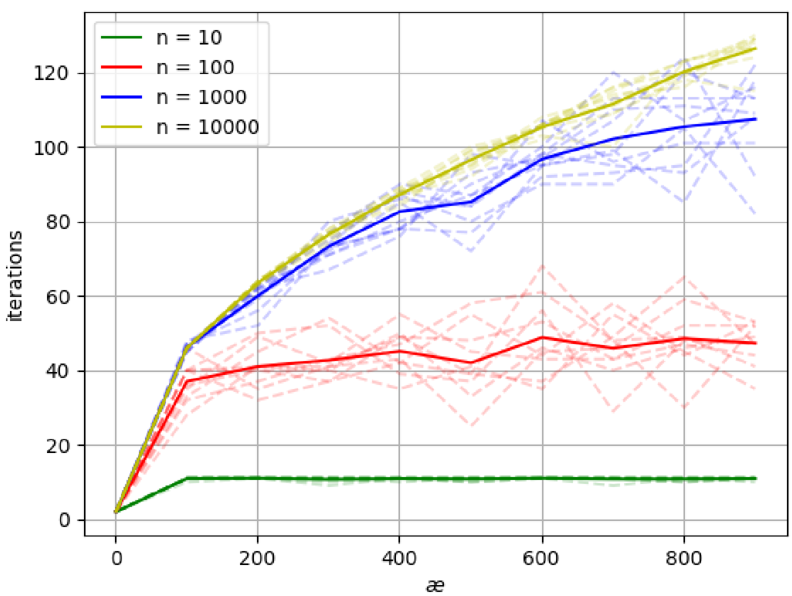
Практическое задание 2: Продвинутые методы безусловной оптимизации.

Калугин Дмитрий

6 апреля 2018 г.

1 Зависимость числа итераций метода сопряженных градиентов от числа обусловленности и размерности пространства



kappa).png

Рис. 1: Зависимость числа итераций метода сопряженных градиентов от числа обусловленности при разных значениях размерности пространства

В случае градиентного спуска, количество итераций расло линейно с ростом числа обусловленности. В случае в методом сопряженных градиентов видно, что количество итераций растет медленнее, число итераций пропорционально корню. Более того, как и следует из алгоритма, количество итераций не превосходит размерность пространства. Это очень четко видно в случае, когда размерность пространства равна 10. В случае, когда размерность пространства 100, видно, что количество итераций так же выходит на асимптоту в окрестности 50. Это может быть связано с тем, что просто алгортм быстро сходится к оптимуму.

2 Выбор размера истории в методе L-BFGS

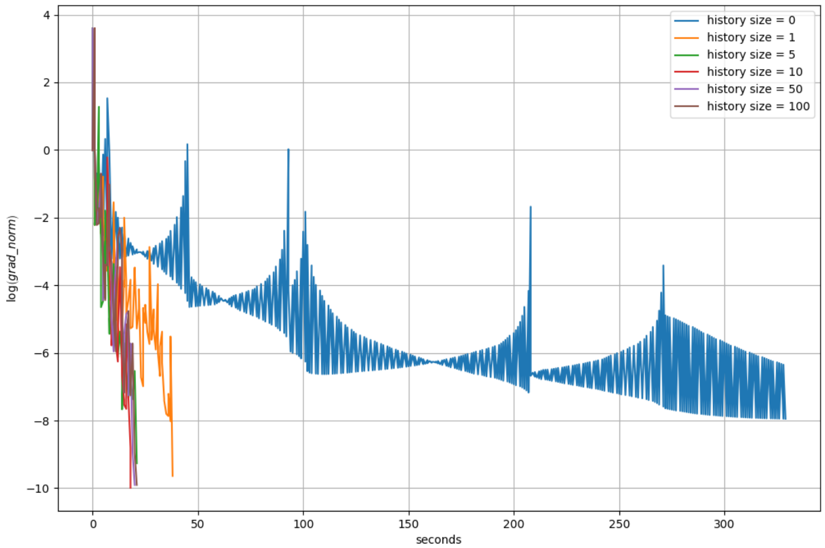


Рис. 2: Относительная норма градиента
от времени в логарифмической шкале

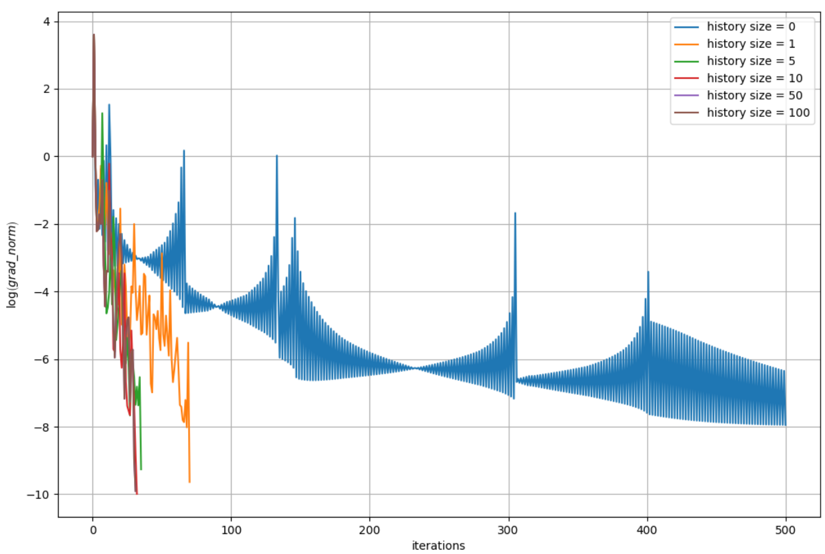


Рис. 3: Относительная норма градиента
от итерации в логарифмической шкале

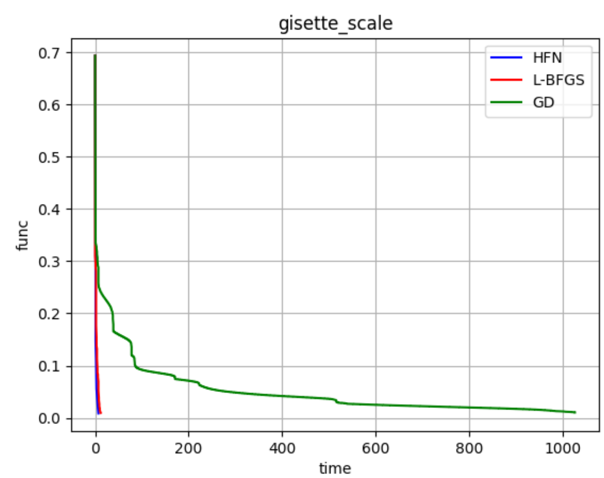
Видно, что очень сильно отстает алгоритм с нулевым размером истории. Это понятно, так как в этом случае алгоритм аналогичен обычному градиентному спуску.

Отдельно можно видеть кривые для историй размера 0, 1, на Рис.3 — 5. Остальные же графики близки настолько, что практически сливаются. Из этого можно сделать вывод, что брать историю большого размера не приводит к выигрышу и оптимальным можно считать размер истории 10. Он и используется в алгоритме по умолчанию.

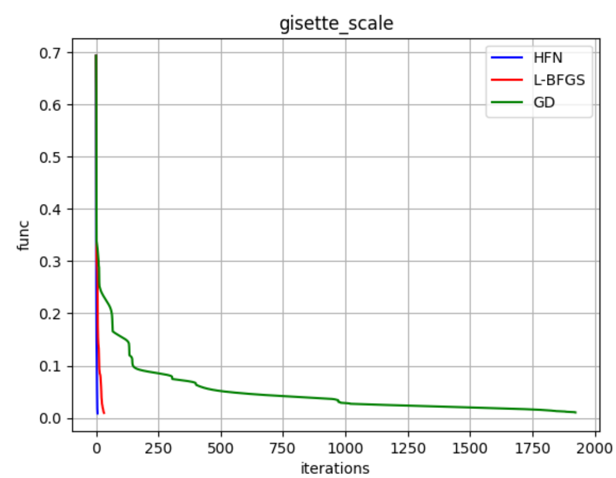
3 Сравнение методов на реальной задаче логистической регрессии

Сравнивались алгоритмы HFN, L-BFGS и GD. Сравнение проводилось на 5 различных датасетах.

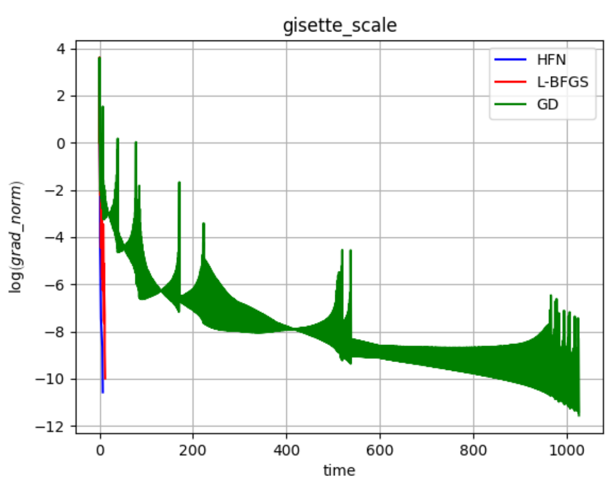
Результаты на датасете *gisette_scale*:



(a) Зависимость функции от времени



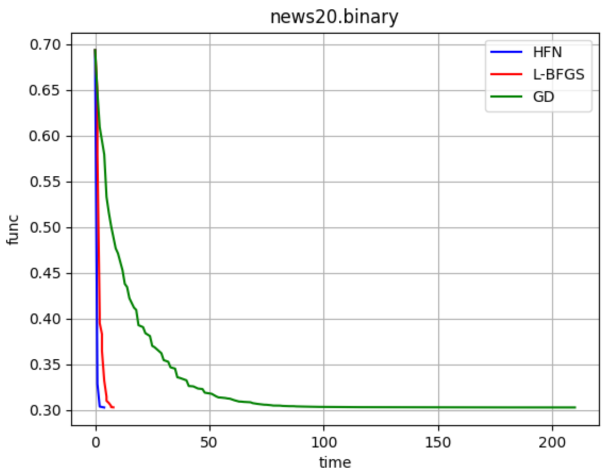
(b) Зависимость функции от итерации



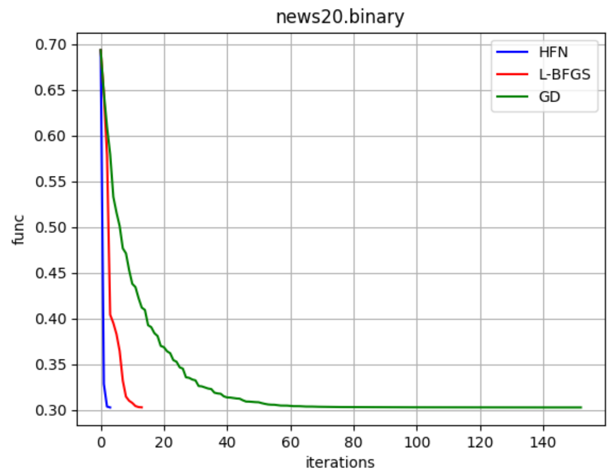
(c) Зависимость нормы градиента от времени

Рис. 4: *gisette_scale*

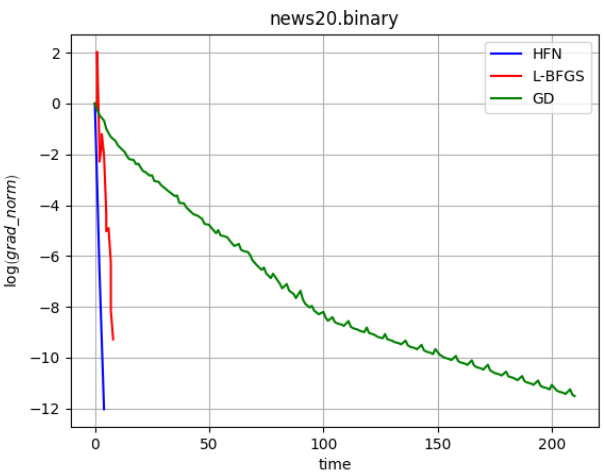
Результаты на датасете $news20.binary$:



(a) Зависимость функции от времени



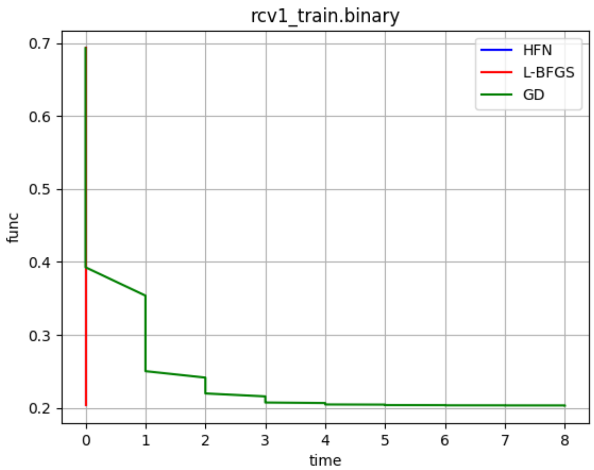
(b) Зависимость функции от итерации



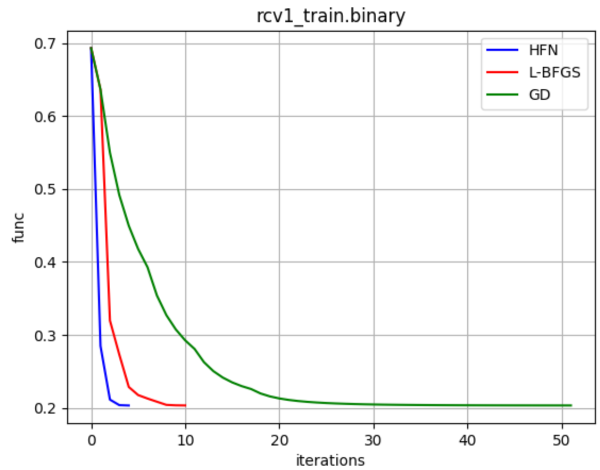
(c) Зависимость нормы градиента от времени

Рис. 5: $news20.binary$

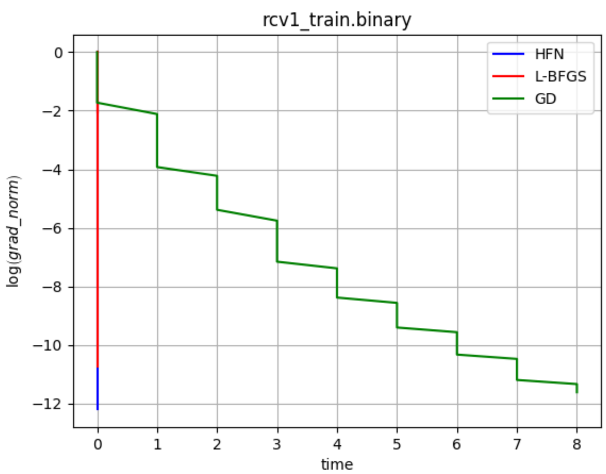
Результаты на датасете $rcv1\_train.binary$:



(a) Зависимость функции от времени



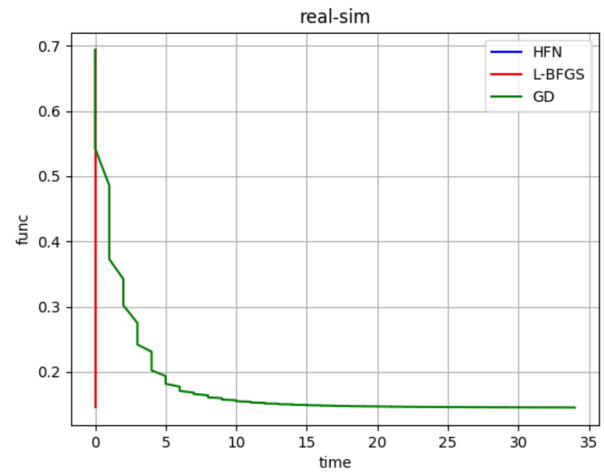
(b) Зависимость функции от итерации



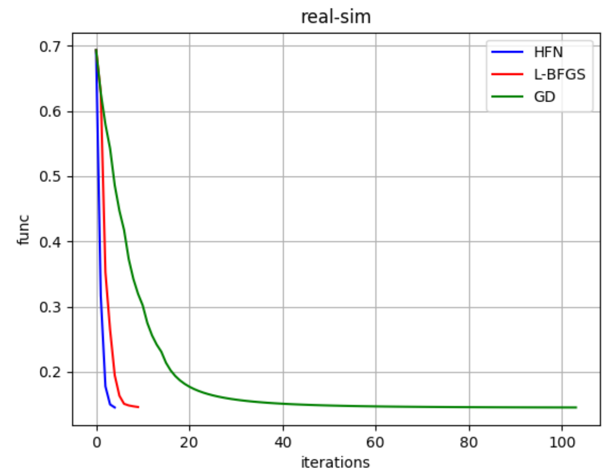
(c) Зависимость нормы градиента от времени

Рис. 6: $rcv1\_train.binary$

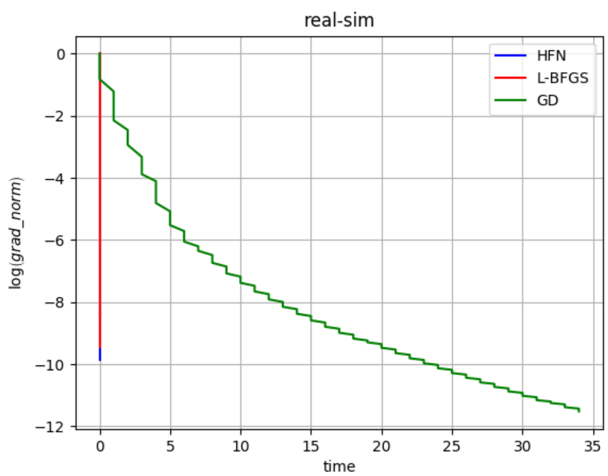
Результаты на датасете $real - sim$:



(a) Зависимость функции от времени



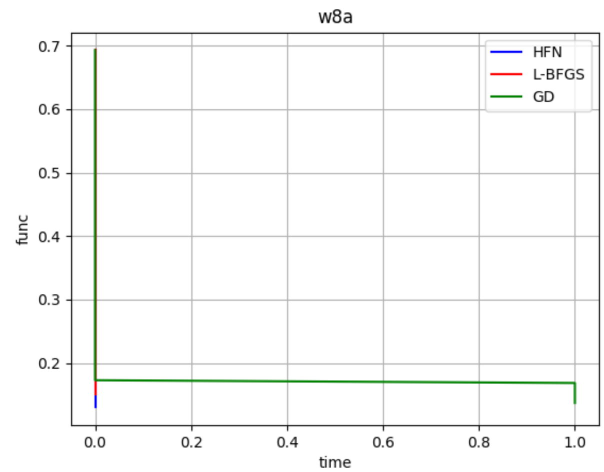
(b) Зависимость функции от итерации



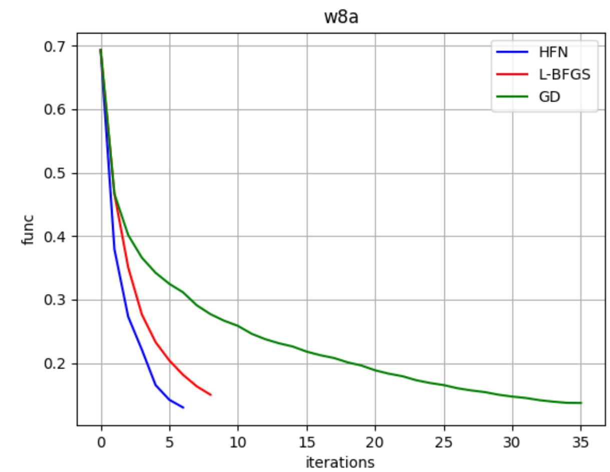
(c) Зависимость нормы градиента от времени

Рис. 7: $real - sim$

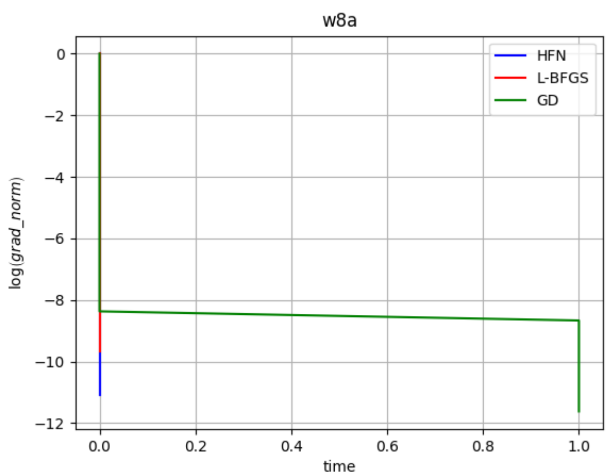
Результаты на датасете $w8a$:



(a) Зависимость функции от времени



(b) Зависимость функции от итерации



(c) Зависимость нормы градиента от времени

Рис. 8: *w8a*

На всех выборках наблюдается одинаковый результат: градиентный спуск очень сильно отстает от методов HFN и L-BFGS. HFN и L-BFGS показывают близкие результаты, но HFN всегда оказывается быстрее. Это связано со следующим. В методе Hessian Free Newton самая сложная операция — умножение гессиана на произвольный вектор. В общем случае это занимает $O(n^3)$ времени. Но в данном случае, для задачи логистической регрессии, зная явную формулу для гессиана, мы смогли сильно уменьшить это время.

На некторых графиках(например, на датасетах $rcv1\_train.binary$, $w8a$) видны странные участки "застоя"зависимостей от времени для градиентного спуска. Скорее всего, это связано с какими-то сторонними процессами в ноутбуке, так как при этом зависимости от номера итерации остаются настолько же гладкими, насколько и с другими датасетами.