Национальный исследовательский университет «Высшая школа экономики»
Санкт-Петербургская школа физико-математических и компьютерных наук

Отчет по Лабораторной работе 3: Метод барьеров

Выполнил: Кудреватых П.С.

Санкт-Петербург 2024

Содержание

1 Метод барьеров в применении к задаче LASSO. Теория

В этом задании мы будем оптимизировать уже подготовленную гладкую условную задачу вида:

$$min_{x,u \in R^n}\{\frac{1}{2}||Ax-b||^2 + \lambda\langle 1_n, u\rangle : -u \leq x \leq u\}$$

где $1_n := (1,...,1) \in R^n$ и неравенства $-u \leq x \leq u$ обозначают обычные поэлементные неравенства между соответствующими компонентами.

Для этой функции введем вспомогательную функцию $f_t(x,u)$, которую будем минимизировать в методе барьеров. По определению эта функция будет:

$$f_t(x,u) = t * f(x,u) - \sum_i (log(u_i + x_i) + log(u_i - x_i))$$

которая переписывается в

$$f_t(x,u) = t * \frac{1}{2}||Ax-b||^2 + t * \lambda\langle 1_n, u\rangle - \sum_i (log(u_i + x_i) + log(u_i - x_i))$$

Такая функция будет проходить безусловную оптимизацию и ее решение будет совпадать с решением исходной задачей условной оптимизации! По заданию мы будем оптимизировать ее методом Ньютона, следовательно нам нужно выписать систему линейных уравнений, решением которой станет направление метода d_k. Вспомним, что по определению d_k задается как

$$\nabla^2 f(x_k) d_k = -\nabla f(x_k)$$

Следовательно нам нужны градиенты и гессианы по двум переменным, а так же смешанная:

$$\nabla_x f(x,u) = t * A^T * (Ax-b) - (\frac{1}{u_i + x_i} - \frac{1}{u_i - x_i}, ..., \frac{1}{u_n + x_n} - \frac{1}{u_n - x_n})^T$$

$$\nabla^2_{xx} f(x,u) = t * A^T A + diag(\frac{1}{(u_i + x_i)^2} + \frac{1}{(u_i - x_i)^2}, ..., \frac{1}{(u_n + x_n)^2} + \frac{1}{(u_n - x_n)^2})$$

$$\nabla_u f(x,u) = t * \lambda * 1_n^T - (\frac{1}{u_i + x_i} + \frac{1}{u_i - x_i}, ..., \frac{1}{u_n + x_n} + \frac{1}{u_n - x_n})^T$$

$$\nabla^2_{uu}f(x,u) = diag(\frac{1}{(u_i+x_i)^2}+\frac{1}{(u_i-x_i)^2},...,\frac{1}{(u_n+x_n)^2}+\frac{1}{(u_n-x_n)^2})$$
$$\nabla^2_{ux}f(x,u) = diag(\frac{1}{(u_i+x_i)^2}-\frac{1}{(u_i-x_i)^2},...,\frac{1}{(u_n+x_n)^2}-\frac{1}{(u_n-x_n)^2})$$

В итоге мы получим систему в довольно блочном виде, а именно:

$$\begin{pmatrix} \nabla^2_{xx}f(x,u) & \nabla^2_{ux}f(x,u) \\ \nabla^2_{ux}f(x,u) & \nabla^2_{uu}f(x,u) \end{pmatrix} \cdot \begin{pmatrix} \vec{d_x} \\ \vec{d_u} \end{pmatrix} = \begin{pmatrix} \nabla_x f(x,u) \\ \nabla_u f(x,u) \end{pmatrix}$$

Обозначим конкатенацию векторов направления за d. Теперь вспомним, что у нас есть два цикла - внешний по увеличивающемуся параметру t с условием остановки на гарантированную точность по зазору двойственности, а второй - внутренний цикл с методом Ньютона с классическим условием остановки на квадрат нормы градиента. В методе Ньютона мы и будем искать вектора d, чтобы оптимизировать как веса, так и условия на них. После выполнения критерия остановки внутреннего цикла мы проверяем зазор во внешнем цикле и продолжаем цикл дальше, увеличивая параметр t в γ раз, если не выполнено условие остановки.

Так как у нас есть условия на оптимизацию, нужно сделать так, чтобы шаг в методе Ньютона не привел к тому, что мы выйдем за границу и точка, в которой мы находимся, перестанет быть внутренней. Это решается выбором размера шага, алгоритм для выбора которого описан в методичке. Но сделаю оговорку, что если не найдется такого вектора q_i, что его скалярное произведение с Ньютоновским направлением d будет положительным, то это означает следующее: какой шаг мы бы не взяли, точка всё равно останется внутренней, давая карт-бланш на выбор шага.

Поговорим о том, какой взять начальную точку в методе барьеров. Не трудно догадаться, что k-ой итерации начальным приближением должно служить k-1 значение вектора z. А вот для нулевой итерации требование лишь одно - чтобы начальная точка была внутренней, к примеру, $\vec{x} = \vec{0}, \vec{u} = \vec{1}$.

2 Эксперимент

В этом разделе будет оптимизироваться оракул LASSO, о котором было подробно рассказано в теоретической части. Единственное что стоит упомянуть это стратегию линейного поиска. Я использую метод Армихо с бектрекингом и коэффициентом $c_1 = 1e - 4$.

В этом эксперименте я решил взять следующие начальные условия:

- $A_{[10000,1000]} \sim N_{0,100}$
- $b_{[10000,1]} \sim N_{0,1}$
- $u_0 = 1_n$
- $x_0 = 0_n$
- $\lambda = 1e - 3$

Все остальные параметры метода взяты по умолчанию. Если в конкретном эксперименте что-то изменяется, то это отражено на графике или текстом!

2.1 Чувствительность метода к выбору параметра γ

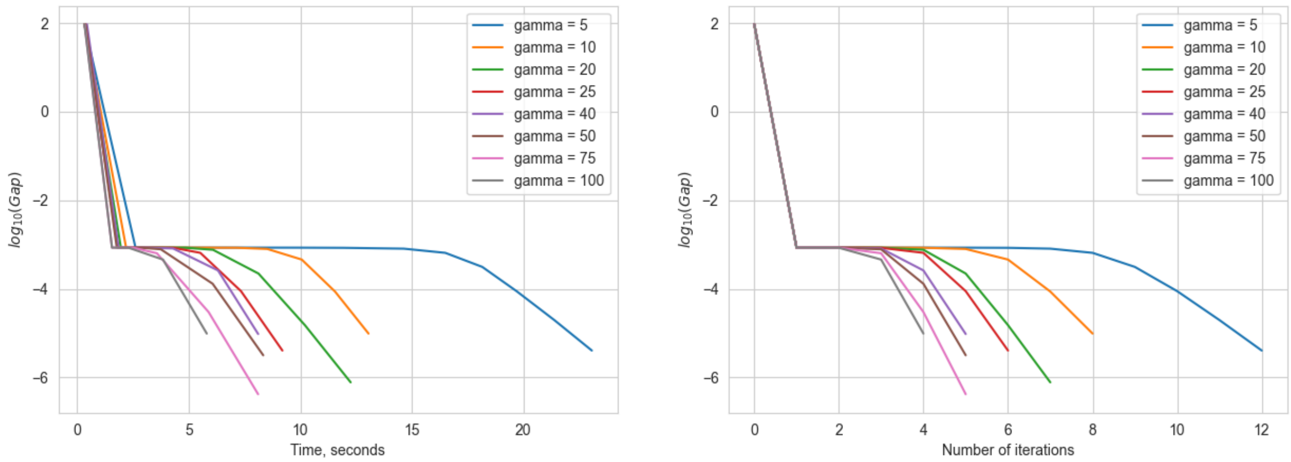

Рис. 1: Зависимость гарантированной точности по зазору от реального времени выполнения и количества итераций

Видно, что параметр увеличения шага по t сильно влияет на скорость сходимости метода, чем меньше его взять - тем дольше метод будет сходиться к точке оптимума. Оптимальными значениями для меня показались значения в диапазоне от 25 до 100.

2.2 Чувствительность метода к выбору параметра ϵ_{inner}

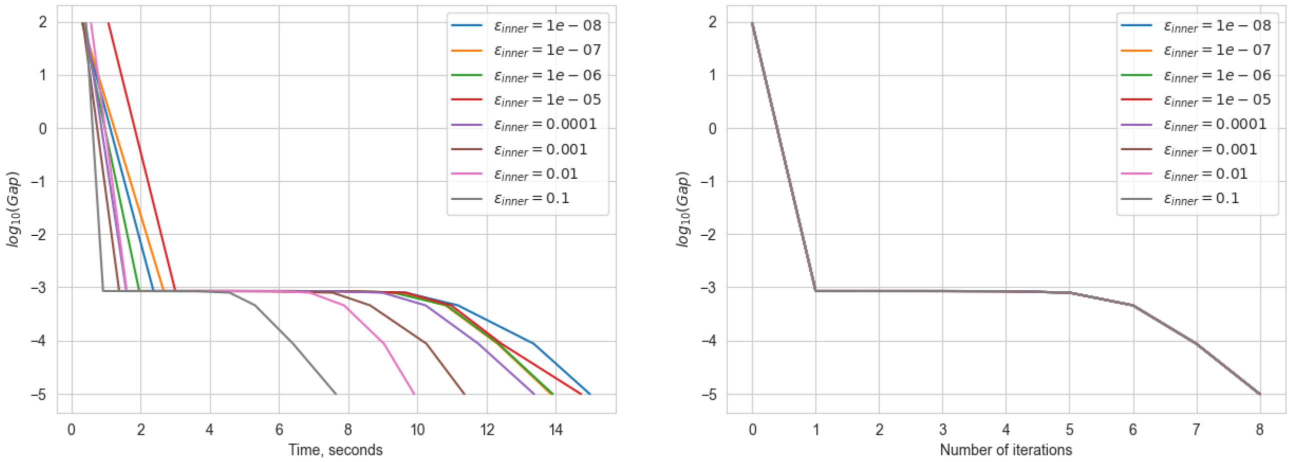

Рис. 2: Зависимость гарантированной точности по зазору от реального времени выполнения и количества итераций

Здесь наблюдается интересный факт, который говорит нам о том, что внутренняя точность никак не влияет на количество внешних итераций на тестовых данных. А вот на время выполнения влияет довольно сильно, при низкой точности оно в разы ниже, но метод до сих пор сходится!

2.3 Чувствительность метода к размерности пространства

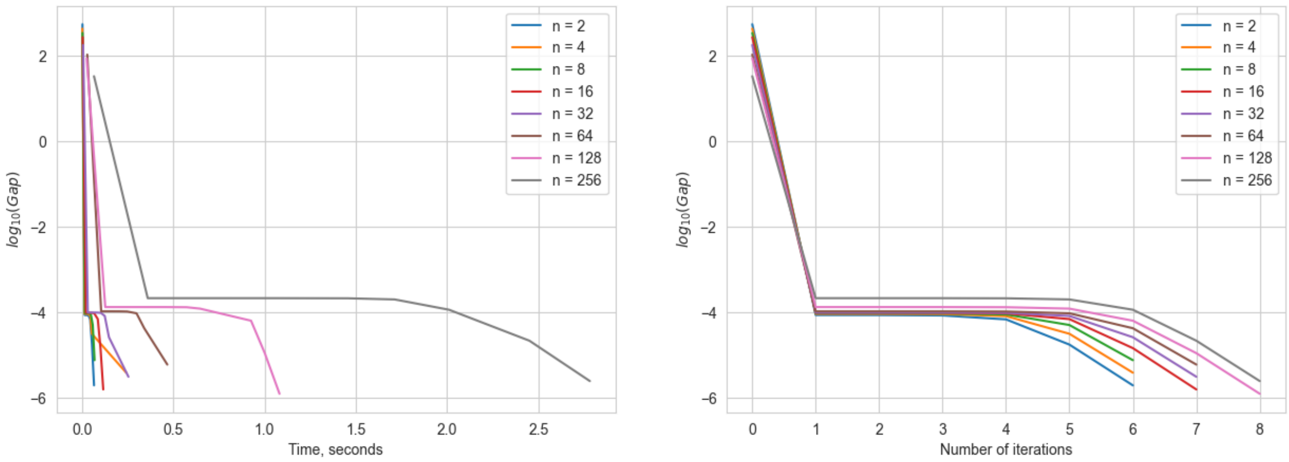

Рис. 3: Зависимость гарантированной точности по зазору от реального времени выполнения и количества итераций

Видно, что методу требуется намного больше времени и большее число итераций при увеличении размерности пространства. Делаем вывод, что цена итерации сильно возрастает при увеличении размерности векторов в датасете.

2.4 Чувствительность метода к размеру выборки

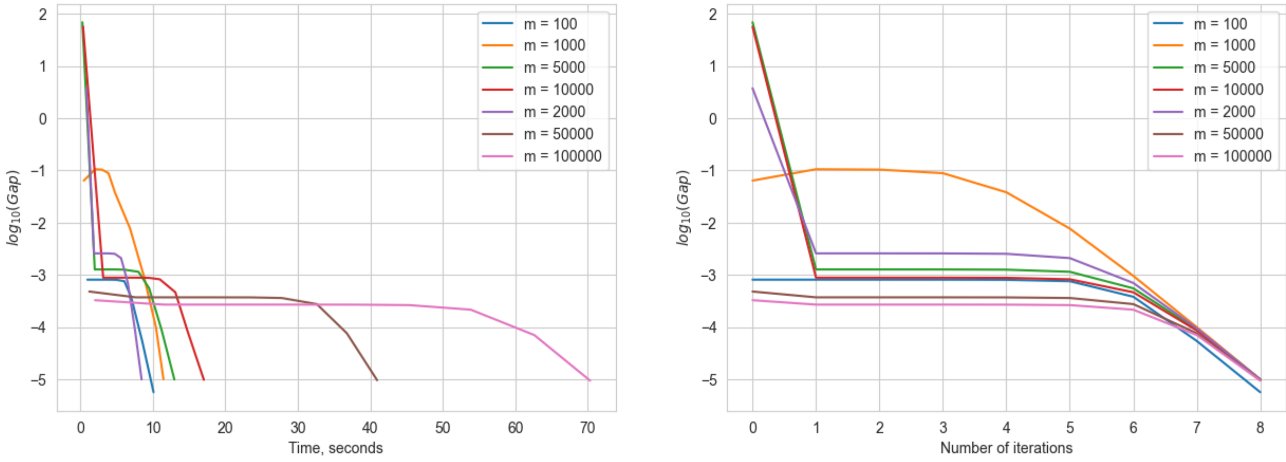

Рис. 4: Зависимость гарантированной точности по зазору от реального времени выполнения и количества итераций

Если посмотреть на второй график можно сделать вывод, что при увеличении размера выборки начальная гарантированная точность по зазору становится меньше, но уменьшается она намного медленнее. Но при любом размере выборки методу требуется одинаковое количество итераций. Если взглянуть на левый график опять же видно, что цена итерации сильно возросла - время выполнения возрастает при увеличении датасета!

2.5 Чувствительность метода к параметру регуляризации λ

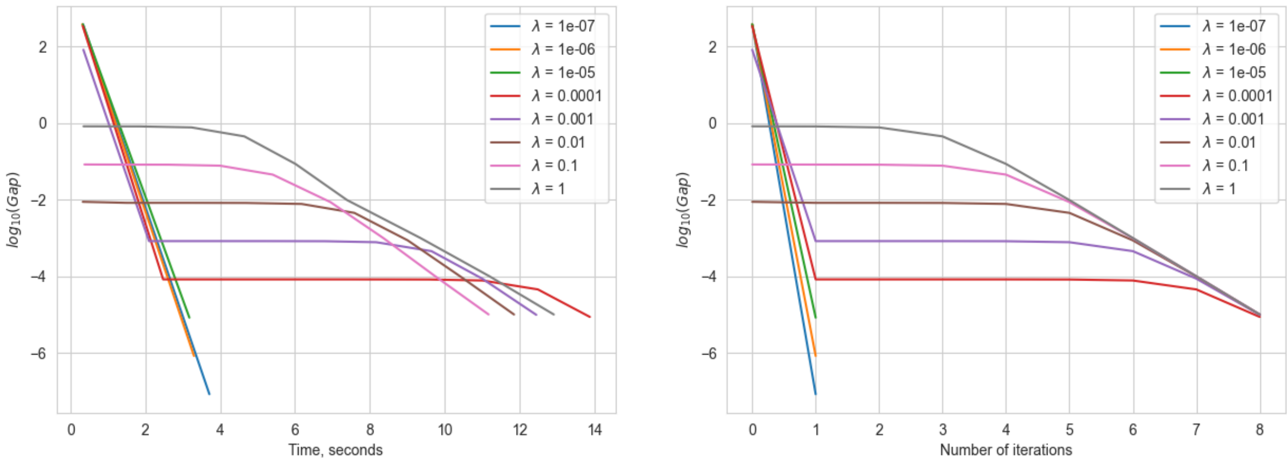

Рис. 5: Зависимость гарантированной точности по зазору от реального времени выполнения и количества итераций

Вспомним, что при увеличении параметра регуляризации начинает больше влиять сумма весов, а следовательно и задача становится более сложной именно по условиям. Если взять маленькую регуляризацию порядка $1e-7$, то эта задача будет подобна задаче обычной линейной регрессии, которая решается аналитически в один шаг взятия обратной матрицы. Это и видно на правом графике - при маленьких параметрах регуляризации метод решает задачу за один шаг внешней итерации по t! Если взять чуть большие параметры регуляризации, ситуация меняется. Методу требуется намного больше итераций и времени для достижения желаемой точности. Видно, что при увеличении параметра регуляризации методу требуется большее реальное время для сходимости, но то же самое количество итераций.

3 Выводы

Был создан оракул LASSO и написан метод логарифмических барьеров для задачи условной оптимизации, который оптимизирует функцию потерь оракула LASSO. Были проведены тесты на синтетических данных на чувствительность метода к различным гипер-параметрам и параметрам датасета.

Интерпретируя полученные результаты, для практики можно сделать вывод о том, что оракул LASSO не подойдет для слишком больших датасетов с большими размерностями пространства из-за своей вычислительной сложности, в таком случае стоит отдать предпочтение L2 регуляризации весов регрессии.

Также полезным замечанием будет то, что метод не сильно страдает от снижения точности на внутренних итерациях, а также довольно робастный к большому шагу на параметр t.