Методы выпуклой оптимизации высокого порядка

Даниил Селиханович

МФТИ
ИППИ им. Харкевича РАН

selihanovich.do@phystech.edu

10 декабря 2018 г.

План доклада

1. Методы высоких порядков выпуклой оптимизации
2. Применение для задачи классификации
3. Описание данных
4. Результаты
5. Выводы

Тензорный шаг Нестерова

В работе [1], вышедшей в марте 2018 года, рассматривается следующий итерационный процесс для решения задачи выпуклой оптимизации $\min\limits_{x \in \mathbb{E}} f(x)$:

$$x_{t+1} = T_{p,M}(x_t), t \geq 0,$$

где функция $f(x) \in \mathcal{F}_p$ - класс выпуклых и p-раз гладких функций, а $T_{p,M}(x_t) \in \arg\min_{y \in \mathbb{E}} \Omega_{x,p,M}(y)$,

$$\Omega_{x,p,M}(y) = f(x) + \sum_{i=1}^{p} \frac{1}{i!} D^i f(x)[y-x]^i + \frac{M}{(p-1)!} \frac{||y-x||^{p+1}}{p+1}$$

при $M \geq L_p$, L_p - константа Липшица для p-й производной функции f.

Случай задачи с целевой функцией logistic-loss

Рассматривается задача бинарной классификацией с логистической функцией потерь:

$$loss(w, y, X) = \sum_{i=1}^{m} \ln\Big(1 + \exp(-y_i \langle X_i, w \rangle)\Big) \to \min_{w \in \mathbb{R}^n},$$

где $y_i \in \{1, -1\}$ - классы объектов, m - число объектов, X - матрица объектов-признаков с константым признаком 1, n - размерность вектора в решающем правиле линейного классификатора:

$$y_{predict}(w, x) = sign\langle w, x \rangle.$$

Оценка констант Липшица для logistic-loss

Определение константы Липшица для p-й производной функции f:

$$||D^p f(x) - D^p f(y)|| \leq L_p ||x - y||,\ x, y \in \mathrm{dom}\, f, p \geq 1.$$

Для $p = 1 \rightarrow D^1 f(x) = \nabla f(x)$, для $p = 2 \rightarrow D^2 f(x) = \nabla^2 f(x)$. Можно показать, что в случае евклидовой нормы для такой функции выполнено:

$$L_1 \leq \frac{1}{4} \sum_{i=1}^{m} ||X_i||_2^2 = M_1;$$

$$L_2 \leq \frac{1}{6\sqrt{3}} \sum_{i=1}^{m} ||X_i||_2^3 \leq \frac{1}{10} \sum_{i=1}^{m} ||X_i||_2^3 = M_2.$$

Цель работы

Постановка задачи

- Применить тензорный метод Нестерова 2-го порядка в задаче бинарной классификации;
- Рассмотреть разные способы проведения тензорного шага;
- Сравнить результат работы с быстрым градиентным спуском;
- Сравнить все алгоритмы с реализацией логистической регрессии в sklearn.

Данные и их описание

Данные Covertype Data Set представляют собой задачу прогнозирования типа лесного покрова в штате Колорадо по картографическим признакам. Задача мультиклассовой классификации была предобработана в задачу бинарной.

- Всего 581012 объектов и 54 признака.
- Из 54 признаков 10 количественные и 44 бинарные.
- Количественные признаки нормированы на $[0, 1]$.
- В обучение вошли 20000 объектов, в тест - 30000.
- Распределение классов: 1-ый класс в train - 13978, в test - 7229, 2-й класс в train - 6022, в test = 22771.

График logistic-loss от числа итераций

Видно, что алгоритмы сопряжённых градиентов и LBFGS в тензорном шаге Нестерова дают очень похожий результат.

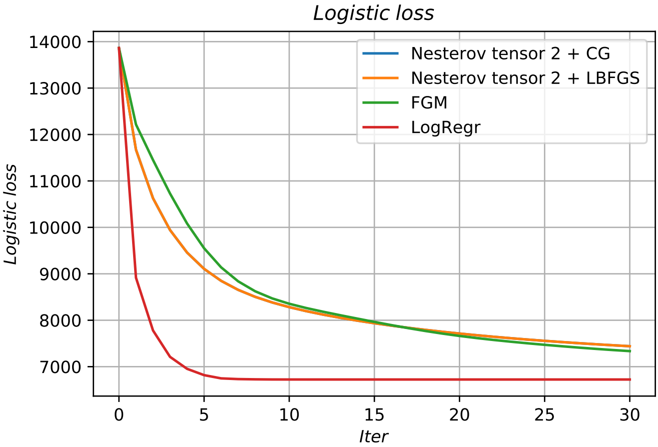
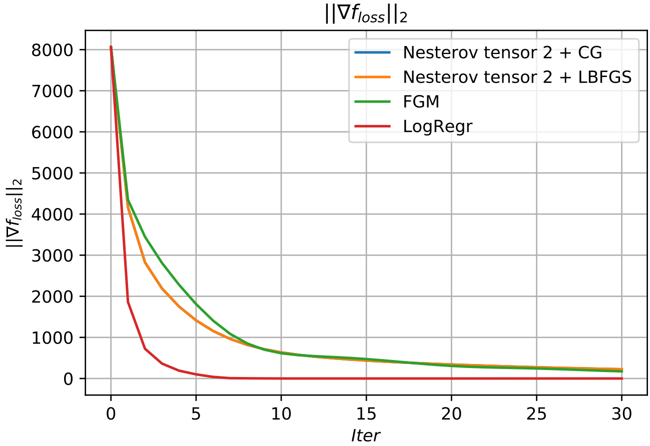


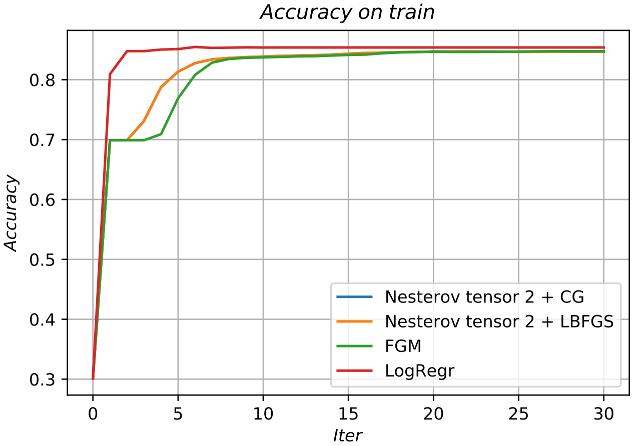
График нормы градиента от числа итераций

Методы высоких порядков работают не намного лучше быстрого градиентного спуска.



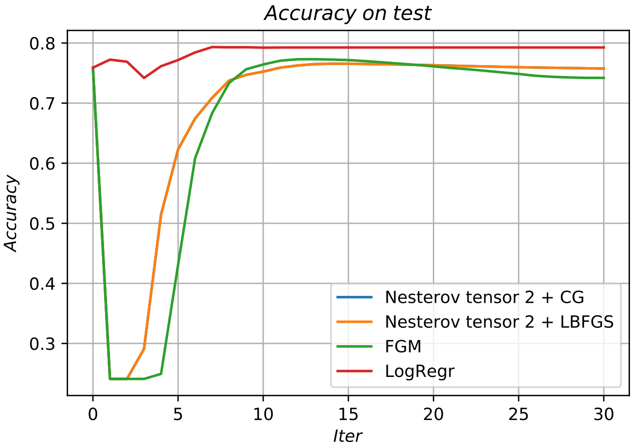
Точность на train

Через большое число итераций все методы дают практически одинаковую точность на train.



Точность на test

FGM и метод 2-го порядка работают хуже на тестовых данных, чем реализованный в библиотеке LIBLINEAR.



Выводы

- Методы высоких порядков могут использоваться в задачах ML.
- С повышением порядка метода увеличиваются вычислительные расходы по времени и памяти.
- Константы Липшица стоит подбирать адаптивно.
- Можно выбирать разные алгоритмы для выполнения тензорного шага.

Литература

- *Nesterov Y.* Implementable tensor methods in unconstrained convex optimization. – Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), 2018. – №. 2018005.
- *Родоманов А. О., Кропотов Д. А., Ветров Д. П.* Анализ быстрого градиентного метода Нестерова для задач машинного обучения с L_1-регуляризацией.
- Современные численные методы оптимизации. Метод универсального градиентного спуска: учебное пособие / А. В. Гасников. – М. : МФТИ, 2018. – 181с. – Изд. 2-е, доп.
- *Гасников А. В. и др.* Обоснование гипотезы об оптимальных оценках скорости сходимости численных методов выпуклой оптимизации высоких порядков.

Благодарность

Выражаю благодарность Е. А. Воронцовой и А. В. Гасникову за руководство проектом и обсуждения результатов!

Спасибо за внимание! Вопросы?