Вопросы к экзамену

по курсу «Методы оптимизации в машинном обучении»,

осень 2017

1. Метод градиентного спуска. Стратегии выбора длины шага. Скорость сходимости метода для сильно выпуклых функций (с доказательством) ([1], разделы 2.2, 3.2, 3.3)
2. Метод Ньютона, его локальная скорость сходимости (с доказательством). Глобальная сходимость. Модификации метода Ньютона для невыпуклых задач оптимизации. ([1], разделы 3.3, 3.4)
3. Метод сопряжённых градиентов для решения системы линейных уравнений (с доказательством). ([1], раздел 5.1)
4. Неточный метод Ньютона (HFN), его скорость сходимости. Способы оценивания произведения гессиана на вектор ([1], раздел 7.1 + [3])
5. Самосогласованные функции: определение, основные примеры, правила комбинирования. Метод Ньютона для самосогласованных функций. Область квадратичной сходимости. Оценки скорости сходимости: глобальная + локальная фазы. (конспект + [13, раздел 4.1])
6. Квазиньютоновские методы оптимизации, методы SR1, BFGS, L-BFGS ([1], разделы 6.1, 6.2, 7.2)
7. Прямые методы оптимизации для выпуклых задач условной оптимизации с ограничениями вида равенств и неравенств (метод Ньютона и метод логарифмических барьеров). ([2], глава 10)
8. Прямо-двойственные методы внутренней точки для выпуклых задач условной оптимизации с ограничениями вида равенств и неравенств. ([2], глава 11)
9. Стандартные классы выпуклых задач: линейное программирование (LP), квадратичное программирование (QP), квадратичное программирование с квадратичными ограничениями (QCQP), коническое программирование второго порядка (SOCP), полуопределенное программирование (SDP). Примеры задач для каждого из классов. Доказательство вложенности классов друг в друга. Лемма о дополнении Шура. Применение метода барьеров для решения задач из стандартных классов.
10. Субградиентный метод, его скорость сходимости (с доказательством). Различные способы выбора длины шага. Субградиентный метод Поляка. ([4], [13, раздел 3.2.3], [16, раздел 5.3])
11. Техника сглаживания. Двойственность свойств сильной выпуклости и липшицевости градиента (с доказательством). Сглаживание с помощью сопряженной функции, оценки на качество аппроксимации. Обобщенное фенхелево представление. Оценка трудоемкости техники сглаживания, сравнение с субградиентным методом. [14]
12. Метод проекции градиента и проксимальный градиентный метод. Скорость сходимости (с доказательством). Подбор длины шага. Примеры применения метода. ([5], раздел 2.3)
13. Ускоренный проксимальный градиентный метод Нестерова, его скорость сходимости (с доказательством). Подбор длины шага. [6]
14. Стохастический субградиентный метод, его скорость сходимости (с доказательством). ([7] + [5, глава 13] + [17, раздел 14.1] + [15, раздел 2.4])
15. Методы стохастической оптимизации с линейной скоростью сходимости: SAG и SVRG [11, 12].
16. Метод SDCA, проксимальная версия метода. Примеры применения. [8]
17. Адаптивный выбор длины шага в стохастическом субградиентном методе (с доказательством). Метод AdaGrad (с доказательством). [15, раздел 3.3]

Теоретический минимум

Ниже перечислены вопросы, незнание ответа на которые во время экзамена автоматически влечёт неудовлетворительную итоговую оценку. Необходимо знать основные определения и формулировки утверждений/теорем; доказательства знать не требуется.

1. Общее определение производной функции (через линейное отображение). Основные свойства производной (правила суммы, произведения и композиции). Определения второй производной, градиента и гессиана.
2. Сублинейная, линейная, сверхлинейная и квадратичная скорости сходимости. Оценки на количество требуемых итераций для достижения заданной точности.
3. Понятия липшицевости градиента, выпуклой и сильно выпуклой функций. Соответствующие глобальные верхние и нижние оценки на функцию.
4. Выпуклые множества и функции. Операции, сохраняющие выпуклость. Простейшие свойства выпуклых функций (неравенство Йенсена, выпуклость надграфика, выпуклость множеств подуровней). Дифференциальные критерии выпуклости.
5. Определение самосогласованной функции. Основные примеры.
6. Субградиент и субдифференциал. Субдифференциальное исчисление (умножение на скаляр, композиция с аффинным преобразованием, теорема Моро--Рокафеллара, максимум конечного числа функций и теорема Данскина). ([3, разделы 3.1.5-3.1.6], [7, раздел 4], [16, раздел 9.1])
7. Сопряженная функция Фенхеля. Неравенство Фенхеля--Юнга. Теорема Фенхеля--Моро. Связь сопряженной функции с субдифференциалом. [4, раздел 7]
8. Сопряженная норма. Неравенство Гельдера. Основные примеры.
9. Необходимые и достаточные условия оптимальности в задачах безусловной гладкой и негладкой минимизации.
10. Условия Армихо и Вульфа для неточной одномерной оптимизации. Процедура бэктрекинга.
11. Схема методов градиентного спуска и Ньютона. Что такое неточный метод Ньютона?
12. Основные матричные разложения: спектральное разложение, сингулярное разложение, LU разложение, QR разложение, разложение Холецкого, LDL разложение.
13. Примеры быстрой и медленной работы методов градиентного спуска и Ньютона.
14. Схема метода сопряжённых градиентов.
15. Теорема Каруша--Куна--Таккера для общей задачи нелинейного программирования. [16, глава 9]
16. Двойственная задача и ее основные свойства. Двойственность Фенхеля. ([16, раздел 9.1.3], [4, раздел 6])
17. Проксимальное отображение. Связь проксимального отображения с евклидовой проекцией.
18. Общая схема квазиньютоновских методов. Отличия метода L-BFGS от BFGS.
19. Общая схема метода барьеров.
20. Общая схема прямо-двойственного метода внутренней точки.
21. Общая схема субградиентного метода.
22. Схема градиентного метода для композитной минимизации (проксимальный градиентный метод).
23. Схема стохастического субградиентного метода.

Литература:

1. J. Nocedal, S.J. Wright. Numerical Optimization. Springer, 2006.
2. S. Boyd, L. Vandenberghe. Convex Optimization, Cambridge University Press, 2004.
3. M. Schmidt. Limited-Memory Quasi-Newton and Hessian-Free Newton Methods for Non-Smooth Optimization // NIPS workshop on optimization for machine learning, 2010.
4. D. Bertsekas. Convex Analysis and Optimization, Athena Scientific, 2003.
5. Optimization for Machine Learning. Edited by Suvrit Sra, Sebastian Nowozin and Stephen J. Wright, MIT Press, 2011.
6. Yu. Nesterov. Gradient methods for minimizing composite objective function. CORE discussion paper, 2007.
7. M. Schmidt. Notes on Big-n Problems, 2012.
8. Shai Shalev-Shwartz, Tong Zhang. Accelerated Proximal Stochastic Dual Coordinate Ascent for Regularized Loss Minimization // ArXiv: 1309.2375, 2013.
11. R. Johnson, T. Zhang. Accelerating Stochastic Gradient Descent using Predictive Variance Reduction // NIPS, 2013.
12. M. Schmidt, N. Le Roux, F. Bach. Minimizing Finite Sums with the Stochastic Average Gradient // ArXiv: 1309.2388, 2013.
13. Ю. Е. Нестеров. Методы выпуклой оптимизации. МЦНМО, Москва, 2010.
14. Yu. Nesterov. Smooth minimization of non-smooth functions, Springer-Verlag 2004.
15. J. Duchi. Introductory Lectures on Stochastic Optimization, Park City Mathematics Institute, Graduate Summer School Lectures, July 2016.
16. Б. Поляк. Введение в оптимизацию, Наука, 1983.
17. A. Nemirovski. Efficient Methods in Convex Programming, 1994.