

## Вопросы к экзамену

### по курсу «Методы оптимизации в машинном обучении», осень 2016

1. Методы одномерной оптимизации без производной: метод золотого сечения, парабол и Брента. ([6], разделы 10.1, 10.2 + [7], раздел 5)
2. Методы одномерной оптимизации с производной: метод деления отрезка пополам, секущей и Брента ([6], разделы 9.1, 9.2 + [7], раздел 4)
3. Метод градиентного спуска. Различные стратегии выбора длины шага. Скорость сходимости метода для сильно-выпуклых функций ([1], разделы 2.2, 3.2, 3.3)
4. Метод Ньютона, его скорость сходимости. Модификации метода Ньютона для невыпуклых задач оптимизации. ([1], разделы 3.3, 3.4)
5. Метод сопряженных градиентов для решения СЛАУ, предобуславливание ([1], раздел 5.1)
6. Метод сопряжённых градиентов для произвольной функции. Схемы Флетчера-Ривса, Полака-Рибье, стратегия рестарта ([1], раздел 5.2)
7. Неточный метод Ньютона (HFN), его скорость сходимости. Способы оценивания произведения гессиана на вектор ([1], раздел 7.1 + [3])
8. Квази-ньютоновские методы оптимизации, методы SR1 и BFGS, метод L-BFGS ([1], разделы 6.1, 6.2, 7.2)
9. Прямые методы оптимизации для выпуклых задач условной оптимизации с ограничениями вида равенств и неравенств (метод Ньютона и метод логарифмических барьеров). ([2], глава 10)
10. Прямо-двойственные методы для выпуклых задач условной оптимизации с ограничениями вида равенств и неравенств. Методы первой фазы ([2], глава 11)
11. Негладкие выпуклые задачи оптимизации. Субградиентный спуск, его скорость сходимости.
12. Разреженные линейные модели для задач регрессии/классификации. Проксимальный градиентный метод. Выбор длины шага. ([5], раздел 2.3)
13. Стохастический градиентный спуск, его скорость сходимости. ([8] + [5], глава 13 + [4], раздел 8.2)
14. Методы стохастической оптимизации с линейной скоростью сходимости: SAG и SVRG [11, 12].
15. Суррогатная оптимизация (оптимизация с использованием глобальных верхних оценок). Способы получения оценок. Примеры оценок. Пример применения метода для задачи LASSO ([10])

### 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. Numerical Recipes. The Art of Scientific Computing, Cambridge University Press, 2007.
7. R. Brent. Algorithms for minimization without derivatives. Prentice-Hall, 1973.
8. M. Schmidt. Notes on Big-n Problems, 2012.
9. J. Martens. Deep Learning via Hessian-free Optimization // ICML, 2010.
10. J. Mairal. Optimization with First-Order Surrogate Functions // Arxiv, 1305.3120, 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.

### Теоретический минимум

Ниже перечислены вопросы, незнание ответа на которые во время экзамена автоматически влечёт неудовлетворительную итоговую оценку.

1. Определение градиента, субградиента и гессиана функции многих переменных.
2. Необходимые и достаточные условия оптимальности в задачах безусловной и условной оптимизации.
3. Определение функции с липшицевым градиентом, выпуклой и сильно выпуклой функции.
4. Условия Армихо и Вольфа для неточной одномерной оптимизации. Процедура backtracking.
5. Схема методов градиентного спуска и Ньютона. Что такое неточный метод Ньютона?
6. Примеры задач быстрой и медленной работы методов градиентного спуска и Ньютона.
7. Схема метода сопряжённых градиентов. Что такое предобуславливание?
8. Двойственная задача условной оптимизации, её свойства.
9. Модели разреженной линейной/логистической регрессии, метода опорных векторов.
10. Определение сублинейной, линейной и квадратичной скорости сходимости, оценки на количество требуемых итераций.
11. Проксимальный оператор, сопряженные функции.
12. Общая схема BFGS и L-BFGS.
13. Общая схема метода логарифмических барьеров.
14. Общая схема прямо-двойственного метода внутренней точки.