

# Методы Оптимизации в Машинном Обучении

Факультет Компьютерных Наук  
ВШЭ

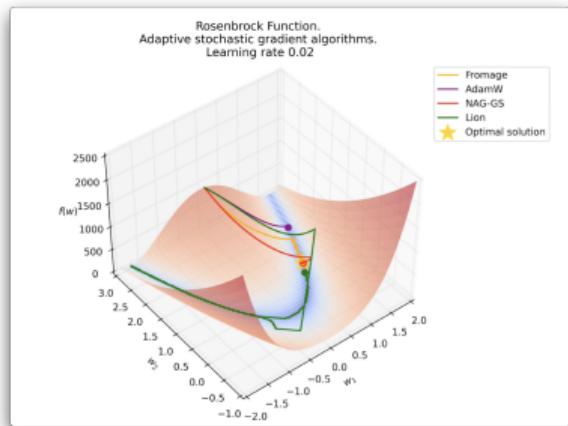


Описание курса

Даня Меркулов

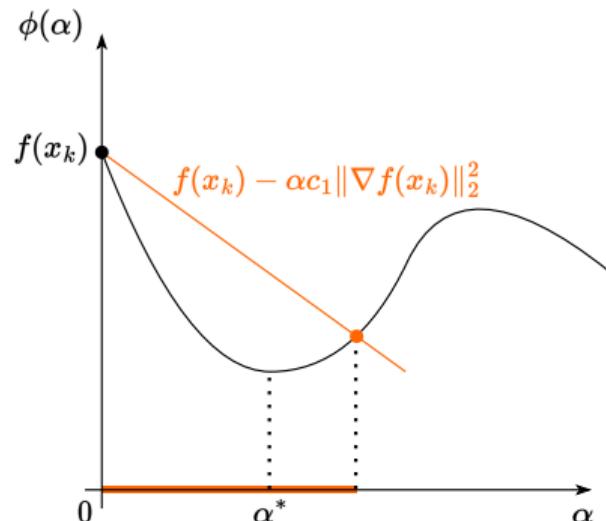
# Общая информация о курсе

- Лектор - Меркулов Даниил Максимович (Сколтех, МФТИ, ВШЭ)
- Курс для студентов 3 курса ПМИ ФКН ВШЭ. 1 лекция + 1 семинар в неделю.
- Старт учебной недели - 8 января. Длительность 20 недель.
- Курс охватывает темы выпуклой, невыпуклой, непрерывной оптимизации, особенно мотивированные задачами и приложениями в Машинном Обучении. Рассматриваются разные темы - от фундаментальных материалов до недавних исследований.



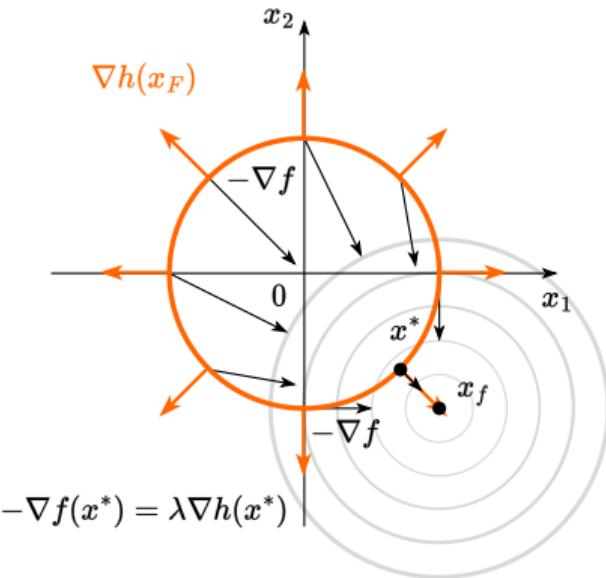
# I. База 1

1. Вспоминаем линейную алгебру. Некоторые матричные разложения. Скорость сходимости. Одномерная оптимизация. Неточная одномерная оптимизация.
2. Градиент. Гессиан. Матрично-векторное дифференцирование. Автоматическое дифференцирование. Вычислительный граф.
3. Выпуклость. Выпуклые множества. Выпуклые функции. Неравенство Йенсена. Сильно выпуклые функции. Условие Поляка - Лоясиевича.



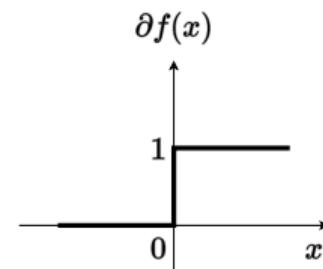
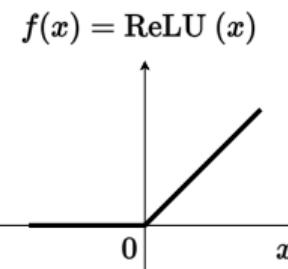
## II. База 2

4. Условия оптимальности. Функция Лагранжа. Множители Лагранжа. Теорема Каруша - Куна - Таккера.
5. Двойственность. Введение в двойственность. Двойственная задача. Анализ чувствительности.
6. Задача линейного программирования. Симплекс метод.



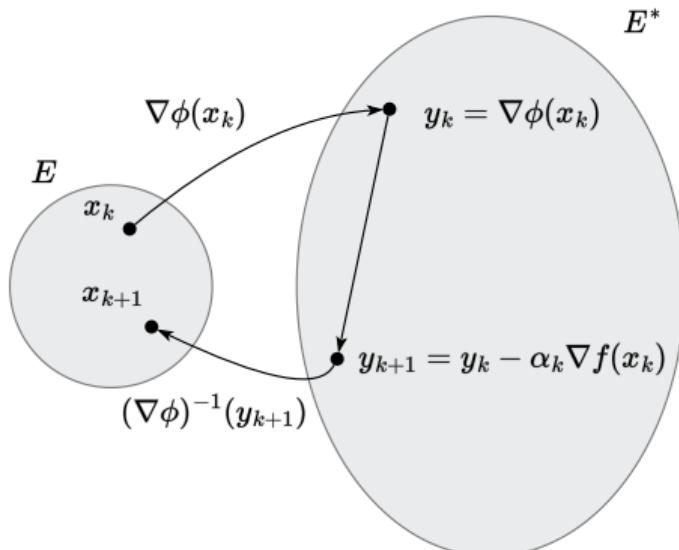
### III. Методы первого порядка

7. Градиентный спуск. Теоремы сходимости в гладком случае (выпуклые, сильно выпуклые, PL). Верхние и нижние оценки сходимости.
8. Ускоренные градиентные методы. Метод Поляка, Нестерова.
9. Субградиент. Субдифференциал. Субградиентный спуск. Теоремы сходимости в негладком случае. Особенности работы градиентного метода в практических негладких задачах.



## IV. Ускорения, ограничения, более высокие порядки

10. Проксимальный градиентный метод. Метод Франк-Вульфа.  
Метод проекции градиента.
11. Метод сопряженных градиентов.
12. Метод натурального градиента. K-FAC. Идея метода зеркального спуска.
13. Метод Ньютона. Квазиньютоновские методы.
14. Введение в методы внутренней точки.



# V. Оптимизация в нейронных сетях и недавние исследования

15. Стохастический градиентный спуск. Адаптивные стохастические градиентные алгоритмы: AdaDelta, RMSProp, Adam, Nadam.
16. Методы редукции дисперсии: SAG, SVRG, SAGA.
17. Градиентный поток и диффузия. Методы оптимизации в непрерывном времени.
18. Седловые задачи. Минимакс. Обучение GAN.
19. Обобщающая способность моделей машинного обучения. Neural Tangent Kernel. Mode connectivity.
20. Инициализация. Регуляризация. Residual connections. Вопросы обучения больших моделей. Lars, Lamb. Расписания learning rate. Warm-up. Клиппинг.

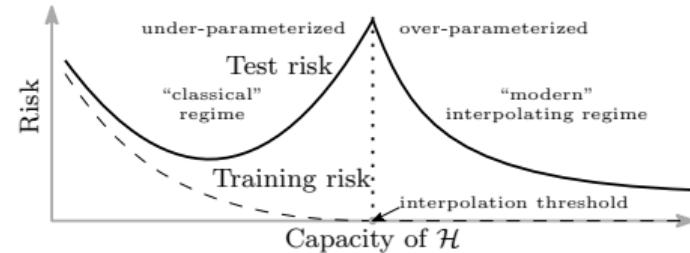
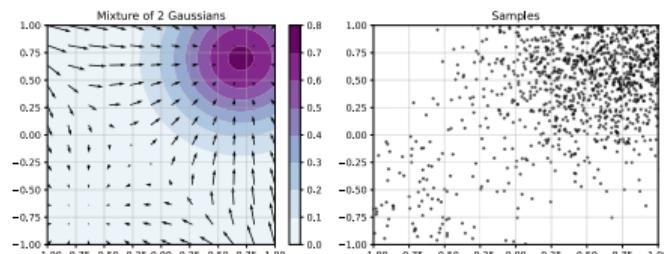


Рис.: Source: [1]



## Оценивание

Оценка за курс вычисляется по следующей формуле:

$$\text{grade} = \text{round} \left( \min \left( 10, \begin{bmatrix} 0.15 \\ 0.35 \\ 0.25 \\ 0.25 \end{bmatrix}^T \begin{bmatrix} \text{Test} \\ \text{HW} \\ \text{Colloquium} \\ \text{Exam} \end{bmatrix} + 0.5 \min \left( 1, \frac{\text{Commits}}{10} \right) + 0.5 \cdot \text{Seminar\_Attendance} \right) \right)$$

- Test, HW, Colloquium, Exam - оценки за соответствующие активности от 0 до 10.
- Тесты проводятся (по возможности) на каждой лекции по материалам предыдущей лекции. За пропущенный по неуважительной причине тест ставится 0.
- Оценки за тесты и коллоквиум являются блокирующими (если набрать меньше 3.5 из 10, то курс не сдан).
- Домашние работы выдаются по темам каждого занятия в конце лекции.
- Commits - количество принятых коммитов в главный репозиторий с учебными материалами.
- Seminar\\_Attendance = 1, если пропущены менее 25% семинаров не по уважительной причине. Иначе Seminar\\_Attendance = 0.

## Коллоквиум

- Коллоквиум пройдёт в конце учебного года (точная дата будет объявлена позднее) и будет включать в себя только материалы по темам прошедших лекций.
- Оценка за коллоквиум складывается из 4 частей
  - Вопросы по формулировкам - 2 балла
  - Теорема с доказательством - 3 балла
  - Решение задачи - 3 балла
  - Дополнительный вопрос - 2 балла
- Сначала выдаются 4 случайных определения/формулировки из списка. На подготовку дается 10 минут. При правильном ответе хотя бы на 3 из 4 определений/формулировок коллоквиум продолжается дальше, и вы получаете  $x - 2$  баллов, где  $x$  – число верно отвеченных вопросов. В противном случае за коллоквиум выставляется 0 баллов.
- При успешной сдаче определений вам выдается билет, содержащий теоретический вопрос на доказательство, а также задачу. На подготовку к ответу дается 40 минут. Теоретический вопрос на доказательства будет по теоремам из списка. Для подготовки к задачам советуем повторить домашние задания, а также задачи с семинаров. В процессе беседы по предыдущим пунктам экзаменатор может задавать уточняющие вопросы.
- После ответа на предыдущие этапы экзаменатор задает дополнительный вопрос, например, задачу или вопрос, связанный с теорией. Ответ на дополнительный вопрос оценивается в 2 балла.

# Экзамен

- Письменный экзамен проводится в конце учебного года и содержит в себе практические и теоретические вопросы по всем темам, которые обсуждали на лекциях и семинарах.
- На экзамене нельзя пользоваться справочными материалами. При себе необходимо иметь только ручку для написания экзамена.

# Материалы

Книги:

- Boyd S. P., Vandenberghe L. Convex optimization. – Cambridge university press, 2004.
- Nocedal J., Wright S. J. (ed.). Numerical optimization. – New York, NY : Springer New York, 1999.
- Nesterov Y. et al. Lectures on convex optimization. – Berlin : Springer, 2018. – Т. 137. – С. 576.
- Жадан В. Г. Методы оптимизации. Части 1, 2, 3 //М.: МФТИ. – 2014.

Ссылки

[1] [Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal.](#)

Reconciling modern machine-learning practice and the classical bias–variance trade-off.

*Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019.