

**Федеральное государственное автономное образовательное
учреждение высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет Компьютерных наук
Департамент больших данных и информационного поиска
Базовая кафедра Яндекс

**Рабочая программа дисциплины «Методы оптимизации в машинном
обучении»
для образовательной программы «Науки о данных»
направления подготовки 01.04.02 "Прикладная математика и информатика"
уровень магистра**

Разработчик(и) программы
Бабенко М.А., к.ф.-м.н. (maxim.babenko@gmail.com)
Одобрена на заседании базовой кафедры Яндекс
«_» «_» 2016 г.

Заведующий Кафедрой
М.А. Бабенко_

Утверждена Академическим советом образовательной программы
«_» «_» 2016 г., № протокола_

Академический руководитель образовательной программы
С.О. Кузнецов _

Москва, 2016

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения подразделения разработчика программы.



1. Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Методы оптимизации в машинном обучении», учебных ассистентов и студентов направления подготовки/специальности 01.04.02 «Прикладная математика и информатика», обучающихся по образовательной программе «Науки о данных».

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом федерального государственного автономного образовательного учреждения высшего профессионального образования «Национального исследовательского университета «Высшая школа экономики»;
- Образовательной программой подготовки магистра по направлению 01.04.02 «Прикладная математика и информатика», специализации «Анализ Интернет-данных».
- Объединенным учебным планом университета по образовательной программе «Науки о данных», утвержденным в 2016 г.

2. Цели освоения дисциплины

Дисциплина «Методы оптимизации в машинном обучении» предназначена для подготовки магистров 01.04.02 – Прикладная математика и информатика.

Настройка модели алгоритмов по данным – это задача оптимизации, от эффективности решения которой зависит практическая применимость метода машинного обучения. В эпоху больших данных многие классические алгоритмы оптимизации становятся неприменимы, т.к. здесь требуется решать задачи оптимизации функций за время меньшее, чем необходимо для вычисления значения функции в одной точке. Таким требованиям можно удовлетворить в случае грамотного комбинирования известных подходов в оптимизации с учётом конкретной специфики решаемой задачи. Курс посвящен изучению классических и современных методов решения задач непрерывной оптимизации (в том числе невыпуклой), а также особенностям применения этих методов в задачах оптимизации, возникающих в машинном обучении. Наличие у слушателей каких-либо предварительных знаний по оптимизации не предполагается, все необходимые понятия разбираются в ходе занятий. Основной акцент в изложении делается на практические аспекты реализации и использования методов. Целью курса является выработка у слушателей навыков по подбору подходящего метода для своей задачи, наиболее полно учитывающего её особенности.

Программа курса предусматривает лекции (26 часов) и практические занятия (38 часов).



3. Тематический план дисциплины

Целью данного курса является ознакомление слушателей с основными методами решения обратных задач интерпретации косвенного эксперимента. Приводятся примеры конкретных теоретических постановок задачи, а также разнообразные подходы к решению. Все темы курса снабжены как теоретическими заданиями, позволяющими глубже понять суть рассматриваемых понятий и методов, так и практическими заданиями, призванными дать возможность сопоставить теорию с практикой.

№	Название темы	Всего часов по дисциплине	Аудиторные часы		Самостоятельная работа
			Лекции	Сем. и практика	
1	Основные понятия и примеры задач.	15	2	3	10
2	Методы одномерной оптимизации	15	2	3	10
3	Методы многомерной оптимизации	22	3	4	15
4	Методы оптимизации с использованием глобальных верхних оценок, зависящих от параметра	17	3	4	10
5	Методы внутренней точки	17	3	4	10
6	Разреженные методы машинного обучения	17	3	4	10
7	Методы отсекающих плоскостей	17	3	4	10
8	Стохастическая оптимизация	22	3	4	15
	Итого	152	22	30	100



4. Компетенции обучающегося, формируемые в результате освоения дисциплины.

В результате освоения дисциплины студент осваивает следующие компетенции:

Компетенции	Код по ФГОС/НИУ	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции
Способен предлагать концепции, модели, изобретать и апробировать способы и инструменты профессиональной деятельности	СК-М2	Студент активно участвует в поиске и реализации поставленных задач	Практические занятия

5. Формы контроля и структура итоговой оценки

Текущий контроль - домашняя работа в первом модуле, контрольная работа в первом модуле.

Итоговый контроль – письменный экзамен (120 мин.)

Итоговая оценка вычисляется следующим образом:

$0,1 \cdot \text{оценка за домашнюю} + 0,2 \cdot \text{оценка за контрольную} + 0,7 \cdot \text{оценка за экзамен.}$

Таблица соответствия оценок по десятибалльной и системе зачет/незачет

Оценка по 10-балльной шкале	Оценка по 5-балльной шкале
1	Незачет
2	
3	
4	Зачет
5	
6	
7	
8	
9	
10	



Таблица соответствия оценок по десятибалльной и пятибалльной системе

По десятибалльной шкале	По пятибалльной системе
1 – неудовлетворительно 2 – очень плохо 3 – плохо	неудовлетворительно – 2
4 – удовлетворительно 5 – весьма удовлетворительно	удовлетворительно – 3
6 – хорошо 7 – очень хорошо	хорошо – 4
8 – почти отлично 9 – отлично 10 – блестяще	отлично – 5

6. Программа дисциплины «Методы оптимизации в машинном обучении»

Тема 1. Основные понятия и примеры задач.

- Градиент и гессиан функции многих переменных, их свойства, необходимые и достаточные условия безусловного экстремума;
- Матричные разложения, их использование для решения СЛАУ;
- Структура итерационного процесса в оптимизации, понятие оракула, критерии останова;
- Глобальная и локальная оптимизация, скорости сходимости итерационных процессов оптимизации;
- Примеры оракулов и задач машинного обучения со «сложной» оптимизацией.

Тема 2. Методы одномерной оптимизации

- Минимизация функции без производной: метод золотого сечения, метод парабол;
- Гибридный метод минимизации Брента;
- Методы решения уравнения : метод деления отрезка пополам, метод секущей;
- Минимизация функции с известной производной: кубическая аппроксимация и модифицированный метод Брента;
- Поиск ограничивающего сегмента;
- Условия Армихо-Голдштайна-Вольфа для неточного решения задачи одномерной оптимизации;
- Неточные методы одномерной оптимизации, backtracking.



Тема 3. Методы многомерной оптимизации.

- Методы линейного поиска и доверительной области;
- Метод градиентного спуска: наискорейший спуск, спуск с неточной одномерной оптимизацией, скорость сходимости метода для сильно-выпуклых функций с липшицевым градиентом, зависимость от шкалы измерений признаков;
- Сходимость общего метода линейного поиска с неточной одномерной минимизацией;
- Метод Ньютона: схема метода, скорость сходимости для выпуклых функций с липшицевым гессианом, подбор длины шага, способы коррекции гессиана до положительно-определённой матрицы;
- Метод сопряженных градиентов для решения систем линейных уравнений, скорость сходимости метода, предобуславливание;
- Метод сопряженных градиентов для оптимизации неквадратичных функций, стратегии рестарта, зависимость от точной одномерной оптимизации;
- Неточный (безгессианный) метод Ньютона: схема метода, способы оценки произведения гессиана на вектор через вычисление градиента;
- Применение неточного метода Ньютона для обучения линейного классификатора и нелинейной регрессии, аппроксимация Гаусса-Ньютона и адаптивная стратегия Levenberg-Marquardt;
- Квазиньютоновские методы оптимизации: DFP, BFGS и L-BFGS;

Тема 4. Методы оптимизации с использованием глобальных верхних оценок, зависящих от параметра

- Вероятностная модель линейной регрессии с различными регуляризациями: квадратичной, L1, Стьюдента;
- Идея метода оптимизации, основанного на использовании глобальных оценок, сходимость;
- Пример применения метода для обучения LASSO;
- Построение глобальных оценок с помощью неравенства Йенсена, EM-алгоритм, его применение для вероятностных моделей линейной регрессии;
- Построение оценок с помощью касательных и замены переменной;
- Оценка Jaakkola-Jordan для логистической функции, оценки для распределений Лапласа и Стьюдента;
- Применение оценок для обучения вероятностных моделей линейной регрессии;
- Выпукло-вогнутая процедура, примеры использования.

Тема 5. Методы внутренней точки.

- Необходимые и достаточные условия оптимальности в задачах условной оптимизации, условия Куна-Таккера и условия Джона, соотношение между ними;
- Выпуклые задачи условной оптимизации, двойственная функция Лагранжа, двойственная задача оптимизации;
- Решение задач условной оптимизации с линейными ограничениями вида равенство, метод Ньютона;
- Прямо-двойственный метод Ньютона, неточный вариант метода;



- Метод логарифмических барьерных функций;
- Методы первой фазы;
- Прямо-двойственный метод внутренней точки;
- Использование методов внутренней точки для обучения SVM.

Тема 6. Разреженные методы машинного обучения

- Модели линейной/логистической регрессии с регуляризациями L1 и L1/L2;
- Понятие субградиента выпуклой функции, его связь с производной по направлению, необходимое и достаточное условие экстремума для выпуклых негладких задач безусловной оптимизации;
- Метод наискорейшего субградиентного спуска;
- Проксимальный метод;
- Метод покоординатного спуска и блочной покоординатной оптимизации.

Тема 7. Методы отсекающих плоскостей.

- Понятие отделяющего оракула, базовый метод отсекающих плоскостей (cutting plane);
- Надграфная форма метода отсекающих плоскостей;
- Bundle-версия метода отсекающих плоскостей, зависимость от настраиваемых параметров;
- Применение bundle-метода для задачи обучения SVM;
- Добавление эффективной процедуры одномерного поиска;
- Реализация метода с использованием параллельных вычислений и в условиях ограничений по памяти.

Тема 8. Стохастическая оптимизация

- Общая постановка задачи стохастической оптимизации, пример использования;
- Задачи минимизации среднего и эмпирического риска;
- Метод стохастического градиентного спуска, две фазы итерационного процесса, использование усреднения и инерции;
- Стохастический градиентный спуск как метод оптимизации и как метод обучения;
- Метод SAG;
- Применение стохастического градиентного спуска для SVM (алгоритм PEGASOS).



7. Список литературы

Основная литература

1. Optimization for Machine Learning. Edited by Suvrit Sra, Sebastian Nowozin and Stephen J. Wright, MIT Press, 2012.
2. J. Nocedal, S.J. Wright. Numerical Optimization, Springer, 2006.
3. S. Boyd, L. Vandenberghe. Convex Optimization, Cambridge University Press, 2004.
4. Yu. Nesterov. Introductory Lectures on Convex Programming, 1998.
A. Antoniou, W.-S. Lu. Practical Optimization: Algorithms and Engineering Applications, Springer, 2007.
5. Numerical Recipes. The Art of Scientific Computing, Cambridge University Press, 2007.
6. M. Schmidt. Limited-Memory Quasi-Newton and Hessian-Free Newton Methods for Non-Smooth Optimization // NIPS workshop on optimization for machine learning, 2010.
7. D. Bertsekas. Convex Analysis and Optimization, Athena Scientific, 2003.
8. M. Schmidt. Notes on Big-n Problems, 2012.