



Национальный исследовательский университет «Высшая школа экономики»
Программа дисциплины «Машинное обучение 2» для направления 01.03.02 «Прикладная
математика и информатика» подготовки бакалавра

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

Факультет Компьютерных наук
Департамент больших данных и информационного поиска

**Рабочая программа дисциплины
Машинное обучение 2**

для образовательной программы «Прикладная математика и информатика»
направления подготовки 01.03.02 «Прикладная математика и информатика»
уровень - бакалавр

Разработчик программы
Соколов Е.А., старший преподаватель, esokolov@hse.ru

Одобрена на заседании департамента больших данных и информационного поиска
«__»_____2017 г.

Руководитель департамента
В.В.Подольский _____

Утверждена Академическим советом образовательной программы
«__»_____2017 г., № протокола _____

Академический руководитель образовательной программы
А.С.Конущин _____

Москва, 2017

*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 минимальные требования к знаниям и умениям студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Машинное обучение 2», учебных ассистентов и студентов направления подготовки 01.03.02 «Прикладная математика и информатика», обучающихся по образовательной программе «Прикладная математика и информатика»

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом ФГАОУ ВО НИУ ВШЭ по направлению подготовки 01.03.02 Прикладная математика и информатика <https://www.hse.ru/data/2017/09/04/1321436546/2.03.02%20%D0%9F%D1%80%D0%B8%D0%BA%D0%BB%D0%B0%D0%B4%D0%BD%D0%B0%D1%8F%20%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%20%D0%B8%20%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0.pdf> ;
- Образовательной программой «Прикладная математика и информатика», направление подготовки 01.03.02 «Прикладная математика и информатика»
- Объединенным учебным планом университета по образовательной программе «Прикладная математика и информатика», утвержденным в 2017 г.

2 Цели освоения дисциплины

Целями освоения дисциплины «Машинное обучение 2» являются:

- Ознакомление студентов с современными методами машинного обучения — в частности, с методами обучения без учителя, генеративными моделями, активным обучением и т.д.
- Формирование у студентов практических навыков работы с данными и решения прикладных задач анализа данных.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент осваивает компетенции:

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен выявлять научную сущность проблем в профессиональной области.	УК-2	РБ	Умение формулировать профессиональные задачи на языке машинного обучения	Лекции, семинары, практические задания	Практические задания, контрольная работа, экзамен
Способен решать проблемы в профессиональной деятельности на основе анализа и синтеза	УК-3	РБ	Умение решать задачи машинного обучения	Лекции, семинары, практические задания	Практические задания, контрольная работа, экзамен
Способен описывать проблемы и ситуации профессиональной	ПК- 1	РБ, СД	Умение формулировать профессиональные	Лекции, семинары, практические задания,	Теоретические задания, контрольная



деятельности, используя язык и аппарат математики			задачи на языке машинного обучения	теоретические задания	работа, экзамен
Способен понимать, совершенствовать и применять современный математический аппарат	ПК-3	СД	Умение использовать математические методы для построения и обучения моделей	Лекции, семинары, практические задания, теоретические задания	Практические задания, теоретические задания, контрольная работа, экзамен
Способен провести сбор, обработку и анализ данных с использованием существующих методов машинного обучения	ПК-7	СД	Умение проводить сбор, обработку и анализ данных с использованием существующих методов машинного обучения	Семинары, практические задания	Практические задания
Способен разработать математическую модель и провести её анализ для поставленной теоретической или прикладной задачи	ПК-8	РБ, СД	Умение разработать математическую модель для задачи машинного обучения	Семинары, практические задания	Практические задания
Способен разработать и реализовать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи на основе математической модели	ПК-9	СД	Умение реализовывать методы машинного обучения	Семинары, практические задания	Практические задания

Виды и задачи профессиональной деятельности

Исследование и разработка математических моделей и методов, алгоритмов и программного обеспечения по тематике проводимых научно-исследовательских проектов;	НИД-3
Применение математических методов и наукоемких технологий для изучения и моделирования сложных систем, в частности, в области обработки и анализа данных, экономики, социологии, физики, наук о жизни и др.;	НИД-4
Разработка математических методов для анализа и построения моделей по тематике выполняемых научно-исследовательских прикладных задач или опытно-конструкторских работ	ПД-1



4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к дисциплинам профессионального цикла, вариативной части профиля.

Для специализации «Машинное обучение и приложения» настоящая дисциплина является обязательной.

Для освоения учебной дисциплины студенты должны владеть знаниями и компетенциями следующих дисциплин:

- Математический анализ
- Линейная алгебра и геометрия
- Теория вероятностей и математическая статистика
- Алгоритмы и структуры данных
- Машинное обучение 1

Основные положения дисциплины должны быть использованы в дальнейшем при изучении дисциплин:

- Машинное обучение на больших данных
- Байесовские методы машинного обучения
- Глубинное обучение
- Компьютерное зрение

5 Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы		Самостоятельная работа
			Лекции	Семина-ры	
1	Обучение с учителем	46	8	8	30
2	Обучение без учителя	62	16	16	30
3	Метрические методы	18	4	4	10
4	Ранжирование	18	4	4	10
5	Адаптация признакового пространства	14	2	2	10
6	Оптимизация гиперпараметров	14	2	2	10
7	Теория статистического обучения	18	4	4	10
	ИТОГО	190	40	40	110



6 Формы контроля знаний студентов

Тип контроля	Форма контроля	3 год				
		1	2	3	4	
Текущий	Домашнее задание			*		
	Домашнее задание			*		
	Домашнее задание				*	
	Домашнее задание				*	
	Контрольная работа			*		Письменный
Итоговый	Экзамен				*	Письменный

7 Критерии оценки знаний, навыков

В курсе предусмотрено несколько форм контроля знания:

- Самостоятельные работы на семинарах, проверяющие знание основных фактов с лекций и выполнение теоретических домашних заданий
- Теоретические домашние задания
- Практические домашние работы на Python, формирующие навыки работы с основными инструментами анализа данных, а также помогающие освоить основные концепции машинного обучения
- Соревнования по анализу данных, направленные на приобретения опыта решения практических задач анализа данных
- Контрольная работа
- Письменный экзамен

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

8 Содержание дисциплины

1. Обучение с учителем (4 лекции, 4 семинара)

Продвинутые методы обучения линейных моделей. Хэширование признаков и случайные проекции. Методы отбора признаков. Обобщённые линейные модели. Современные имплементации градиентного бустинга. Ядровые методы.

2. Обучение без учителя (8 лекций, 8 семинаров)

Ядра, их теоретические основы и примеры применения для сложных объектов. Аппроксимация признаков в спрямляющем пространстве. EM-алгоритм. Тематическое моделирование, LDA. Генеративные модели. Продвинутые методы кластеризации. Лапласианы графов в машинном обучении. Матричные разложения и факторизационные машины; их применения в машинном обучении. Методы обнаружения аномалий, одноклассовые методы.

3. Метрические методы (2 лекции, 2 семинара)

Метод k ближайших соседей в классификации и регрессии. Метрики на сложных объектах. Метрики на текстах на основе представлений слов. Методы поиска ближайших соседей. Locality-sensitive hashing. Обучение метрик, learning to hash.

4. Ранжирование (2 лекции, 2 семинара)

Обучение ранжированию. Метрики качества ранжирования. Поточечный, попарный и списочный подходы. Краудсорсинг при сборе разметки, оптимальная агрегация меток.



5. Задача адаптации признакового пространства (domain adaptation), transfer learning (1 лекция, 1 семинар)
6. Оптимизация гиперпараметров, гауссовские процессы в оптимизации (1 лекция, 1 семинар)
7. Теория статистического обучения (2 лекции, 2 семинара)

9 Оценочные средства для текущего контроля и аттестации студента

Примеры вопросов к экзамену и контрольной работе

1. Теорема о представлении (Representer theorem) с доказательством.
2. Аппроксимация спрямляющего пространства: метод Нистрома и метод случайных признаков Фурье.
3. Скрытые переменные, полное и неполное правдоподобие на примере смеси распределений. EM-алгоритм (описание шагов).
4. Дивергенция Кульбака-Лейблера, её неотрицательность. Вывод E- и M-шагов через разложение логарифма неполного правдоподобия на нижнюю оценку и KL-дивергенцию.
5. Ядровой метод главных компонент.
6. Двойственная задача SVM (с выводом). Какие объекты называются периферийными, опорными граничными, опорными нарушителями?
7. Лапласиан графа, его свойства. Связь кратности нулевого собственного значения с числом компонент связности (с доказательством). Алгоритм спектральной кластеризации.
8. Методы частичного обучения: self-training, моделирование смесью гауссиан, трансдуктивный SVM, графовый подход на основе лапласианов.
9. Несбалансированная классификация: undersampling, oversampling, SMOTE.
10. Одноклассовая классификация: непараметрические методы (восстановление плотности), одноклассовый SVM.
11. User-based и item-based рекомендательные системы.
12. Модели со скрытыми переменными (latent factor model, LFM) для построения рекомендаций. Обучение LFM: стохастический градиентный спуск, ALS, HALS.
13. Неявная информация в рекомендательных системах, implicit ALS.
14. Факторизационные машины. FFM. Способы обучения моделей.
15. Метрики качества рекомендаций.
16. Оценки максимального правдоподобия на параметры многомерного нормального распределения.
17. All-subsequences-kernel, алгоритм его вычисления за $O(|s||t|)$.
18. Вероятностная постановка задачи классификации, функционал среднего риска. Оптимальное байесовское решающее правило, наивный байесовский классификатор.

Примеры домашних заданий и задач к контрольной работе и экзамену

См. репозиторий курса: <https://github.com/esokolov/ml-course-hse/tree/master/2016-spring>

10 Порядок формирования оценок по дисциплине

Результирующая оценка по дисциплине рассчитывается по формуле

$$O_{\text{итог}} = 0.7 O_{\text{накопл}} + 0.3 O_{\text{экз}}$$

Накопленная, экзаменационная и итоговая оценки округляются арифметически.

Накопленная оценка рассчитывается по формуле

$$O_{\text{накопл}} = 0.2 O_{\text{самост}} + 0.4 O_{\text{практдз}} + 0.2 O_{\text{теордз}} + 0.2 O_{\text{контрольная}}$$

Оценка за домашние задания рассчитывается как среднее значение оценок за все выданные



домашние задания. Оценка за самостоятельную работу рассчитывается как среднее значение оценок за все проверочные работы, проведённые на семинарских занятиях. В конце семестра разрешается переписать все самостоятельные работы, пропущенные по уважительной причине.

Студенту, получившему отличную накопленную оценку, данная оценка может быть выставлена в качестве итоговой при условии, что им набрано определённое количество дополнительных баллов, установленное преподавателем.

11 Учебно-методическое и информационное обеспечение дисциплины

11.1 Базовые учебники

1. Hastie T., Tibshirani R, Friedman J. [The Elements of Statistical Learning \(2nd edition\)](#). Springer, 2009.
2. Bishop C. M. [Pattern Recognition and Machine Learning](#). Springer, 2006.

11.2 Дополнительная литература

1. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of Machine Learning. MIT Press, 2012.
2. Murphy K. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.
3. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014.