



Национальный исследовательский университет «Высшая школа экономики»
Программа дисциплины «Практические задачи машинного обучения»
для направления 01.03.02 «Прикладная математика и информатика» подготовки

**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
Практические задачи машинного обучения**

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

Авторы программы:

Симагин Д.А. (dasimagin@yandex-team.ru)

Одобрена на заседании департамента больших данных и информационного поиска

«_»_____2018 г.

Руководитель департамента

В.В.Подольский_____

Утверждена Академическим советом образовательной программы

«_»_____2018 г., № протокола_____

Академический руководитель образовательной программы

А.С.Конушин_____

Москва, 2018

*Настоящая программа не может быть использована другими подразделениями университета
и другими вузами без разрешения кафедры-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 данную дисциплину, учебных ассистентов и студентов направления подготовки 01.03.02 «Прикладная математика и информатика» образовательной программы «Прикладная математика и информатика», изучающих дисциплину «Практические задачи машинного обучения».

Программа разработана в соответствии с:

- образовательным стандартом федерального государственного автономного образовательного учреждения высшего образования НИУ ВШЭ по направлению подготовки 01.03.02 «Прикладная математика и информатика» (квалификация – бакалавр);
- образовательной программой «Прикладная математика и информатика» по направлению подготовки 01.03.02 «Прикладная математика и информатика» (уровень подготовки: Бакалавр);
- рабочим учебным планом университета по образовательной программе «Прикладная математика и информатика», утвержденным в 2018 г.

2. Цели освоения дисциплины

Основная цель освоения дисциплины «Прикладное машинное обучение» – дать студентам базовые знания в различных разделах машинного обучения. Предложенный материал позволит более успешно пройти соответствующие профильные курсы. К изучению будут предложены материалы, основанные на наиболее свежих научных исследованиях, что поможет студенту в выборе направления для дальнейшей исследовательской деятельности. Практической уклон зайти позволит получить навыки в решении реальных практических задач.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- Знать понятия и методы машинного обучения, которые могут быть полезны для дальнейшего изучения соответствующих дисциплин, а также для применения в профессиональной деятельности;
- Уметь пользоваться методами машинного обучения и популярными программными пакетам для решения практических задач машинного обучения;
- Ориентироваться в наиболее популярных направлениях исследований в машинном обучении.

В результате освоения дисциплины студент осваивает следующие 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
Способен учиться, приобретать новые знания, умения, в том числе в области, отличной от профессиональной	УК-1	МЦ, СД	Умение учиться и приобретать новые знания и умения в области машинного обучения	Лекции, семинары и участие в соревнованиях по машинному обучению
Способен вести исследовательскую деятельность, включая анализ проблем, постановку целей и задач, выделение объекта и предмета исследования, выбор способа и методов исследования, а также оценку его качества	УК-6	РБ	Умение формализовать проблемы и изучение методов исследований машинного обучения	Лекции, семинары и участие в соревнованиях по машинному обучению



Способен описывать проблемы и ситуации профессиональной деятельности, используя язык и аппарат математики	ПК-1	РБ, СД	Умение описывать неформальные условия задач в формальных математических терминах	Практические занятия, домашние задания
Способен понимать, совершенствовать и применять современный математический аппарат	ПК-3	РБ	Умение использовать аппарат машинного обучения при решении задач и понимать теоретический материал	Лекции, семинары и участие в соревнованиях по машинному обучению
Способен формализовать и алгоритмизировать поставленную задачу	ПК-4	РБ, СД	Умение формализовывать математическую задачу и строить алгоритм для её решения	Лекции, семинары и участие в соревнованиях по машинному обучению
Способен вести письменную и устную коммуникацию на английском языке в рамках профессионального и научного общения, как межличностного, так и группового	ПК-10	СД	Умение излагать свои решения в письменном виде и совершенствование коммуникации на английском языке при обсуждении и решении задач	Лекции и семинары
Способен грамотно и аргументировано публично представлять результаты своей научной и профессиональной деятельности, в т.ч. используя современные средства ИКТ.	ПК-15		Умение излагать идеи, формулировки и доказательства теорем, решения задач по машинному обучению.	Лекции, семинары и участие в соревнованиях по машинному обучению

4. Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к базовой части профессионального цикла, является самостоятельной учебной дисциплиной, относится к математическому и естественнонаучному циклу дисциплин. Для направления 01.03.02 «Прикладная математика и информатика» дисциплина является дополнительной.

Для освоения учебной дисциплины, студенты должны владеть знаниями и навыками в объеме программы средней школы по математике и освоить учебные курсы:

Дискретная математика
Математический анализ 1
Теория вероятностей

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

Машинное обучение 1;
Непрерывная оптимизация;
Машинное обучение 2;
Машинное обучение на больших данных;
Основные методы анализа данных;
Машинное обучение и майнинг данных;
Введение в теорию статистического обучения;
Байесовские методы машинного обучения;



5. Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы			Самостоятельная работа
			Лекции и	Семинары	Практические занятия	
	Введение в машинное обучение	10	2	2		8
1	Линейные модели	30	6	6		18
2	Байесовский классификатор и оценка распределения	20	4	4		12
3	Решающие деревья и композиции	24	4	4		16
4	Нейронные сети. Сверточные сети и их применение для обработки изображений.	30	6	6		18
6						
7	Обучение представлений слов	18	2	2		14
8	Обработка текстов. Обучение представлений слов. Рекуррентные сети. Задача перевода.	22	4	4		14
9	Обучение с подкрепление	26	4	4		18
10	Обзор современных направлений исследований	8	4	4		0
	Итого:	190	36	36		118

6. Формы контроля знаний студентов

Участие в соревнованиях по машинному обучению. Уровень знаний студента определяется на основе качества его решений. В конце курса проводится итоговый устный экзамен.

7. Критерии оценки знаний, навыков

Для прохождения контроля студент должен продемонстрировать понимание основных определений, знание методов машинного обучения и умение применять полученные знания для решения практических задач. Оценки за соревнования по машинному обучению и устный экзамен выставляются по 10-ти балльной системе.

За участие в соревновании допускается оценка 0 в тех случаях, когда студент не предложил решения или его качество решения низко (пороговое значение определяется в зависимости от задачи). Оценка 0 также ставится в случаях доказанного списывания.

Пересдача и комиссия курса проходит в форме решения одного из проведенных ранее соревнований в рамках курса (по выбору студента). Для получения оценки «зачет» и выше решение должно иметь качество на тестовой выборке не ниже оговоренного заранее порога.

Оценка 0 за экзамен допускается в случае неявки студента или доказанного списывания.



8. Содержание дисциплины

1. Введение в машинное обучение, основная терминология, варианты постановки задач.
2. Задача регрессии и классификации на примере метода ближайших соседей.
3. Линейные модели классификации и регрессии.
4. Метод опорных векторов для бинарной классификации, замена скалярного произведения ядерной функцией.
5. Байесовский классификатор и оценка плотности распределения. Смеси распределений, EM-алгоритм.
6. Решающие деревья. Композиция деревьев, бэггинг и бустинг.
7. Метрики качества классификации.
8. Задача ранжирования. Применение решающих деревьев.
9. Матричные разложения. Рекомендательные системы, коллаборативная фильтрация.
10. Введение в нейронные сети. Алгоритм обратного распространения ошибки. Универсальная теорема аппроксимации.
11. Сверточные сети в задачах классификации, детекции и сегментации изображений. Популярные архитектуры.
12. Обучение представлений для слов (модель word2vec).
13. Рекуррентные сети и их применимость для задач обработки текстов. Задача перевода, метрики качества.
14. Введение в обучение с подкреплением. Задача многорукого бандита. Марковский процесс принятия решения.
15. Алгоритм SARSA и Q-обучения.
16. Глубинное обучение с подкреплением (DQN).
17. Обзор современных результатов.

9. Порядок формирования оценок по дисциплине

Оценка состоит из двух частей: накопленная и экзамен. В ходе курса студенты примут участие в серии соревнований, основанных на реальных данных. Цель конкурса – предложить модель, которая набирает наибольшее количество очков относительно предложенной метрики. Оценка за каждое соревнование определяется итоговым качеством на закрытой тестовой выборке. Накопленная оценка является средневзвешенной по результатам всех конкурсов. Экзамен проводится в устной форме.

$$O_{\text{итог}} = 0.7 * O_{\text{накопленная}} + 0.3 O_{\text{экзамен}}$$

В вычислениях текущие оценки и промежуточные величины не округляются. Результат вычисляется точно и округляется только в момент выставления итоговой оценки. Итоговая оценка округляется вверх.

Перевод в 5-балльную шкалу осуществляется по правилу:

Оценка по 10-балльной шкале	Оценка по 5-балльной шкале
1, 2, 3	Неудовлетворительно
4, 5	Удовлетворительно
6, 7	Хорошо
8, 9, 10	Отлично