

**Санкт-Петербургский филиал федерального государственного
автономного образовательного учреждения высшего образования
«Национальный исследовательский университет
"Высшая школа экономики"»**

Факультет Санкт-Петербургская школа
физико-математических и компьютерных наук
Департамент информатики

Рабочая программа дисциплины
Дополнительные главы машинного обучения

Разработчик: Кураленок Игорь Евгеньевич, ikuralenok@hse.ru

Утверждена Академическим руководителем образовательной программы

«31» августа 2018 г.

А.В. Омельченко _____

Санкт-Петербург, 2018

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 определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Дополнительные главы машинного обучения», учебных ассистентов и студентов направления 01.03.02 «Прикладная математика и информатика» подготовки магистров, обучающихся по бакалаврской программе «Прикладная математика и информатика» и изучающих дисциплину «Дополнительные главы машинного обучения».

Рабочая программа дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ по направлению подготовки 01.03.02 «Прикладная математика и информатика» (уровень бакалавриата), утвержденным ученым советом Национального исследовательского университета «Высшая школа экономики», протокол от 03.03.2017 №02.
- Основной профессиональной образовательной программой «Прикладная математика и информатика» направления подготовки 01.03.02 «Прикладная математика и информатика»;

2. Цели освоения дисциплины

Целью освоения дисциплины «Дополнительные главы машинного обучения» является формирование у студентов теоретических знаний и практических навыков по основам работы с продвинутыми алгоритмами машинного обучения.

Дисциплина направлена на ознакомление студентов с теоретическими основами и основными принципами машинного обучения, овладение студентами инструментарием, моделями и методами машинного обучения, а также приобретение навыков исследователя данных (data scientist) и разработчика математических моделей, методов и алгоритмов анализа данных.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- Знать ключевые понятия, цели и задачи использования машинного обучения; методологические основы применения алгоритмов машинного обучения.
- Уметь визуализировать результаты работы алгоритмов машинного обучения, выбирать метод машинного обучения, соответствующий исследовательской задаче, интерпретировать полученные результаты.
- Иметь навыки (приобрести опыт) чтения и анализа академической литературы по применению методов машинного обучения, построения и оценки качества моделей.

В результате освоения дисциплины студент осваивает следующие компетенции:

Компетенция	Код по стандарту НИУ ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию	Форма контроля уровня сформированности компетенции
-------------	--------------------------	----------------------------------	---	---	--

				компетенции	
Способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1	РБ СД МЦ	Знает различные методы математической статистики, применяемые для выделения важных признаков и уменьшения размерности данных. Применяет данные алгоритмы для конкретных наборов данных. Оценивает значимость и корректность производимых изменений признаков.	Лекции, подготовка к практическим занятиям, работа на практических занятиях, самостоятельная работа	Проект, устный экзамен
Способен применять и модифицировать математические модели для решения задач в области профессиональной деятельности	ОПК-3	РБ СД МЦ	Знает различные алгоритмы и математические модели, применяемые в анализе распределения данных. Производит семплинг, используя различные методы машинного обучения. Выбирает целевые функции для интерпретации результатов машинного	Лекции, подготовка к практическим занятиям, работа на практических занятиях, самостоятельная работа	Проект, устный экзамен

исследованиям		МЦ	решения возникающих прикладных задач. Обрабатывает и интерпретирует результаты существующих программных решений.		
Способен разрабатывать программное и информационное обеспечение компьютерных систем, сервисов, вычислительных комплексов, баз данных	ПК-3	РБ СД МЦ	Знает основные библиотеки на языках Python и Java, используемые в решении задач из области машинного обучения. Реализует различные методы и алгоритмы машинного обучения на языках Python и Java. Участвует в разработке OpenSource библиотек для машинного обучения.	Лекции, подготовка к практическим занятиям, работа на практических занятиях, самостоятельная работа	Проект, устный экзамен
Способен грамотно и аргументировано публично представлять результаты своей научной и профессиональной деятельности, в т.ч. используя современные средства ИКТ	ПК-5	РБ СД МЦ	Знает основные требования к авторам при оформлении решения задач из соревнований по машинному обучению. Производит оценку результатов решения задачи методами машинного обучения. Владеет	Лекции, подготовка к практическим занятиям, работа на практических занятиях, самостоятельная работа	Проект, устный экзамен

			навыками публичной презентации результатов проекта.		
--	--	--	---	--	--

4. Место дисциплины в структуре образовательной программы

Для образовательной программы «Прикладная математика и информатика» направления подготовки 01.03.02 «Прикладная математика и информатика» настоящая дисциплина относится к дисциплинам по выбору блока дисциплин.

Для освоения дисциплины необходимы компетенции, полученные в ходе изучения дисциплин:

- Машинное обучение;
- Методы оптимизации.

5. Тематический план учебной дисциплины

Курс рассчитан на 44 часов аудиторной нагрузки, из них 30 часов лекций и 14 часов практических занятий, общим объемом 3 зачетных единицы (114 часа).

№	Название раздела	Всего часов	Аудиторные часы			Самостоятельная работа
			Лекции	Семинары	Практические занятия	
1	Анализ размерности данных и работа с признаками	36	10	0	4	22
2	Анализ распределения данных	39	10	0	5	24
3	Интерпретация результатов	39	10	0	5	24
ИТОГО		114	30	0	14	70

6. Содержание дисциплины

<u>Раздел 1</u> Анализ размерности данных и работа с признаками	
Тема 1	Анализ размерности данных. Понятие эффективной размерности. Уменьшение размерности данных. Выделение важных признаков.
Тема 2	Работа с признаками. Работа с последовательностями (текст, звук, сигнал), изображениями. Выделение признаков. Коллаборативная фильтрация.

<u>Раздел 2</u> Анализ распределения данных	
Тема 1	Анализ распределения признаков в данных и сэмплинг. Максимизация ожидания (EM), гауссовский сэмплинг (GS), методы Монте Карло с марковскими цепями (MCMC).
Тема 2	Сложность моделей машинного обучения. Функция роста. Теорема Вапника-Червоненкиса. Размерность Вапника-Червоненкиса. PAC-Bayes анализ.
<u>Раздел 3</u> Интерперетация результатов	
Тема 1	Интерпретация результатов машинного обучения. Целевые функции. Практическая оценка качества обучения. Теоретическая оценка качества обучения.
Тема 2	Работа с экспертными оценками больших данных. Фильтрация экспертных оценок больших данных. Согласие экспертов. Репутация экспертов.

7. Оценочные средства

7.1. Формы контроля знаний студентов

Тип контроля	Форма контроля	4 год		Параметры
		1 модуль	2 модуль	
Текущий	Проект		*	Проектное задание
Итоговый	Устный экзамен		*	Экзамен в устной форме

7.2. Критерии и шкалы оценки, примеры заданий

7.2.1. Текущий контроль

Текущий контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств.

ПРОЕКТ

В процессе изучения дисциплины обучающиеся должны применить пройденные методы для разбора одной из статей по машинному обучению. В качестве статей выбираются любые публикации конференций по машинному обучению.

Примерный список приемлемых конференций:

1. NIPS (Neural Information Processing Systems)
2. ICML (International Conference on Machine Learning)
3. CIKM (Conference on Information and Knowledge Management)
4. KDD (Knowledge discovery in databases)
5. WSDM (Web Search and Data Mining)

Критерии оценивания и шкала оценки проекта

Оценка	Критерии выставления оценки
«Отлично» (8-10)	Проведен полный разбор статьи, все математические выкладки проведены самостоятельно; проведен анализ границ применимости метода и научности статьи; студент ответил на все вопросы по статье
«Хорошо» (6-7)	Проведен полный разбор статьи, все математические выкладки проведены самостоятельно; проведен анализ границ применимости метода; студент не ответил на один из вопросов по статье
«Удовлетворительно» (4-5)	Проведен полный разбор статьи, однако не все выкладки проделаны самостоятельно; студент ответил не на все вопросы
«Неудовлетворительно» (0-3)	Проведен неполный разбор статьи; студент не может ответить на вопросы; изложение лишено логики

7.2.2. Итоговый контроль по дисциплине

Проверка качества освоения дисциплины производится в форме устного экзамена.

УСТНЫЙ ЭКЗАМЕН

Экзамен проводится в формате разбора научной статьи по анализу данных и машинному обучению. Экзаменуемый должен продемонстрировать знание предмета на достаточном уровне для интерпретации современной научной литературы:

Примеры статей:

1. Wilson, Ashia C., et al. "The marginal value of adaptive gradient methods in machine learning."
2. Jain, Prateek, and Purushottam Kar. "Non-convex optimization for machine learning."
3. Amit, Ron, and Ron Meir. "Meta-Learning by Adjusting Priors Based on Extended PAC-Bayes Theory."

Критерии оценивания и шкала оценки устного экзамена

Оценка	Критерии выставления оценки
«Отлично» (8-10)	Проведен полный разбор статьи, все математические выкладки проведены самостоятельно; проведен анализ границ применимости метода и научности статьи; студент ответил на все вопросы по статье
«Хорошо» (6-7)	Проведен полный разбор статьи, все математические выкладки проведены самостоятельно; проведен анализ границ применимости метода; студент не ответил на один из вопросов по статье
«Удовлетворительно» (4-5)	Проведен полный разбор статьи, однако не все выкладки проделаны самостоятельно; студент ответил не на все вопросы
«Неудовлетворительно» (0-3)	Проведен неполный разбор статьи; студент не может ответить на вопросы; изложение лишено логики

7.3 Порядок формирования оценок по дисциплине

Результирующая оценка за дисциплину рассчитывается следующим образом:

$$O_{\text{Результирующая}} = 0,5 O_{\text{проект}} + 0,5 O_{\text{экзамен}}$$

Действует следующий способ округления оценки: при значениях от 0,1 до 0,4 оценка

округляется в меньшую сторону, от 0,5 до 0,9 – в большую.

На экзамене студенту не предоставляется возможность получить дополнительный балл для компенсации оценки за текущий контроль.

8. Образовательные технологии

Основными образовательными технологиями являются: интерактивные лекции, работа в группах на практических занятиях.

9. Учебно-методическое и информационное обеспечение дисциплины

9.1 Основная литература

1. Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning Series) / Murphy, Kevin P. MIT Press. 2014
2. Encyclopedia of Machine Learning and Data Mining / Sammut. Springer. 2016

9.2 Дополнительная литература

1. Machine Learning and Knowledge Discovery in Databases / Peter A. Flach; Tijl Bie; Nello Cristianini. Springer Berlin Heidelberg. 2012
2. Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer series in statistics) / Hastie, Trevor; Tibshirani, Robert; Friedman, J. H.; Penn. Springer. 2009
3. Building Machine Learning Systems with Python: Get More From Your Data Through Creating Practical Machine Learning Systems with Python / Luis Pedro Coelho; Willi Richert. Packt Publishing. 2015

10. Рекомендации для самостоятельной работы студентов

Самостоятельная работа может рассматриваться как организационная форма обучения – система педагогических условий, обеспечивающих управление учебной деятельностью по освоению знаний и умений в области учебной деятельности без посторонней помощи. Студенту нужно четко понимать, что самостоятельная работа – не просто обязательное, а необходимое условие для получения знаний по дисциплине и развитию компетенций, необходимых в будущей профессиональной деятельности.

Самостоятельная работа проводится с целью:

- систематизации и закрепления полученных на лекциях теоретических знаний;
- углубления и расширения теоретических знаний;
- формирования умений использовать нормативную, правовую, справочную документацию и специальную литературу;
- развития познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;
- формирования самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- формирования практических (общеучебных и профессиональных) умений и навыков;
- развития исследовательских умений;
- получения навыков эффективной самостоятельной профессиональной (практической и научно-теоретической) деятельности.

В учебном процессе выделяют два вида самостоятельной работы:

- аудиторная;
- внеаудиторная.

Аудиторная самостоятельная работа по дисциплине выполняется на учебных занятиях под непосредственным руководством преподавателя и по его заданию.

Внеаудиторная самостоятельная работа - планируемая учебная работа студентов, выполняемая во внеаудиторное время по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Самостоятельная работа, не предусмотренная программой учебной дисциплины, раскрывающей и конкретизирующей ее содержание, осуществляется студентом инициативно, с целью реализации собственных учебных и научных интересов.

Для более эффективного выполнения самостоятельной работы по дисциплине преподаватель рекомендует источники для работы, характеризует наиболее рациональную методику самостоятельной работы, демонстрирует ранее выполненные студентами работы и т. п.

Виды заданий для внеаудиторной самостоятельной работы, их содержание и характер могут иметь вариативный и дифференцированный характер, учитывать индивидуальные особенности студента.

Самостоятельная работа может осуществляться индивидуально или группами студентов online и на занятиях в зависимости от цели, объема, конкретной тематики самостоятельной работы, уровня сложности.

Контроль результатов внеаудиторной самостоятельной работы осуществляется в пределах времени, отведенного на обязательные учебные занятия по дисциплине на практических занятиях.

11. Материально-техническое обеспечение дисциплины и информационные технологии, используемые при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения информационных справочных систем (при необходимости).

Для проведения всех занятий используется проектор и компьютер для проекции слайдов. Для самостоятельной работы необходимо следующее ПО: Oracle Java, Python.

12. Особенности организации обучения для лиц с ограниченными возможностями здоровья

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) могут предлагаться следующих варианты восприятия учебной информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

1) для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации.

2) для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации.

3) для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.