**Санкт-Петербургский государственный университет**

**Р А Б О Ч А Я П Р О Г Р А М М А**

**УЧЕБНОЙ ДИСЦИПЛИНЫ**

Машинное обучение

Machine Learning

**Язык(и) обучения**

русский

Трудоемкость в зачетных единицах: 3

Регистрационный номер рабочей программы: 027269

Санкт-Петербург,

2020

**Раздел 1. Характеристики учебных занятий**

**1.1. Цели и задачи учебных занятий**

Целью курса «Машинное обучение» является введение в обширную область прикладной математики, занимающейся моделированием и оптимизацией в условиях неполных и неточных данных. Одними из наиболее важных аспектов предлагаемого курса и его отличительной особенностью является широта охвата и практическая направленность. Большой охват достигается за счет меньшей подробности изложения: многие утверждения приводится без строгих доказательств. Практическая направленность выражается как в том, что рассматриваются работающие методы, доказавшие свою состоятельность в реальных условиях, так и в организации практических занятий вокруг экспериментов с реальными данными.

Еще одной отличительной особенностью и целью курса является составление у слушателя целостной картины актуальных исследований в области машинного обучения. Поэтому в курсе рассматриваются не только ставшие классическими, но и перспективные направления развития машинного обучения. Материалы для этой части курса взяты и обновляются по результатам ведущих мировых конференций (ICML, CIKM, NIPS, и т.д.).

**1.2. Требования подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)**

Программа курса предназначена обучающимся 1 курса магистратуры и рассчитана на обучающихся, изучавших:

• общий курс математического анализа;

• линейную алгебру;

• методы вычислений;

• углубленное программирование;

• теорию вероятностей;

• математическую статистику;

• введение в машинное обучение.

Обязательным условием к слушателям курса является способность к самостоятельному программированию и доступ к соответствующей технике. Для чтения рекомендованной литературы нужен английский язык.

**1.3. Перечень результатов обучения (learning outcomes)**

После прослушивания курса, предполагается, что обучающийся получит следующие знания и умения:

* понимание задач машинного обучения, отличия предмета от смежных областей;
* навыки по моделированию практических задач и перевод их на язык оптимизации;
* способность оценить потенциал использования того или иного подхода в условиях поставленной задачи и предоставленных данных;
* навыки практической реализации решений на одном из языков программирования (Octave/Python/Java);
* навыки анализа полученных результатов и возможных путей их улучшения;
* ориентирование в актуальных проблемах машинного достаточное для формулирования собственных исследовательских задач.

Курс формирует следующие компетенции:

* ПКА-2 – способен учитывать знания проблем и тенденций развития рынка ПО в профессиональной деятельности;
* ПКП-1 – способность проводить научные исследования на основе существующих методов в конкретной области профессиональной деятельности;
* ПКП-7 – способен использовать основные концептуальные положения функционального, логического, объектно-ориентированного и визуального направлений программирования, методы, способы и средства разработки программ в рамках этих направлений;
* ПКП-8 – способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования

**1.4. Перечень и объём активных и интерактивных форм учебных занятий**

Занятия в активных и интерактивных формах (общим объёмом 25 ак. ч):

* семинары (15 ак. ч.);
* лекции, предполагающие активную дискуссию с преподавателем (10 ак. ч.).

**Раздел 2. Организация, структура и содержание учебных занятий**

**2.1. Организация учебных занятий**

**2.1.1 Основной курс**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся | | | | | | | | | | | | | | | | | | |
| Код модуля в составе дисциплины,  практики и т.п. | Контактная работа обучающихся с преподавателем | | | | | | | | | | | | Самостоятельная работа | | | | Объём активных и интерактивных  форм учебных занятий | Трудоёмкость |
| лекции | семинары | консультации | практические  занятия | лабораторные работы | контрольные работы | коллоквиумы | текущий контроль | промежуточная  аттестация | итоговая аттестация | под руководством преподавателя | в присутствии  преподавателя | сам. раб. с использованием  методических материалов | текущий контроль (сам.раб.) | промежуточная аттестация (сам.раб.) | итоговая аттестация  (сам.раб.) |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | | | | | | | | | | | | | |
| Форма обучения: очная | | | | | | | | | | | | | | | | | | |
| Семестр 2 | 15 | 15 | 2 |  |  |  |  |  | 2 |  |  |  | 44 |  | 30 |  | 25 | 3 |
|  | 2-100 | 2-30 | 1-100 |  |  |  |  |  | 1-100 |  |  |  | 1-1 |  | 1-1 |  |  |  |
| ИТОГО | 15 | 15 | 2 |  |  |  |  |  | 2 |  |  |  | 44 |  | 30 |  |  | 3 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Виды, формы и сроки текущего контроля успеваемости и промежуточной аттестации | | | | | | |
| Код модуля в составе дисциплины, практики и т.п. | Формы текущего контроля успеваемости | | Виды промежуточной аттестации | | Виды итоговой аттестации  (только для программ итоговой аттестации и дополнительных образовательных программ) | |
| Формы | Сроки | Виды | Сроки | Виды | Сроки |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | |
| Форма обучения: очная | | | | | | |
| Семестр 2 |  |  | экзамен, устно, традиционная форма | по графику промежуточной аттестации |  |  |

**2.2. Структура и содержание учебных занятий**

Модуль 1 (Ближайшие соседи)

1. Instance based learning (IBL). Проклятие размерности.

2. Практические аспекты реализации IBL.

Обучение метрикам (по Brian Kulis).

Модуль 2 (Уменьшение размерности)

1. Обзор методов уменьшения размерности. Johnson-Lindenstrauss Lemma.

2. Выбор свойств (feature selection).

3. Выделение новых свойств: PCA, ICA.

4. Кластерный анализ. Самоорганизующиеся сети Кохоннена.

Модуль 3 (Нейронные сети)

1. Персептронные сети. Обратное распространение ошибки.

2. Сети Хопфилда, Bolzman machine. Обзор глубокого обучения (Deep Learning).

Модуль 4 (Деревья решений и их ансамбли)

1. Деревья решений: ID3/C4.5. CART, забывчивые деревья.

2. Bagging. Случайные леса. Boosting на примере AdaBoost.

3. Gradient Boosting. MatrixNet.

**Раздел 3. Обеспечение учебных занятий**

**3.1. Методическое обеспечение**

**3.1.1 Методические указания по освоению дисциплины**

Методические материалы включают в себя следующие типы материалов — литература, Интернет-ресурсы, учебные пособия, с опорой на которые проводится аудиторная работа.

**3.1.2 Методическое обеспечение самостоятельной работы**

Самостоятельная работа обучающегося, как вид деятельности, стимулирующий активность, самостоятельность, познавательный интерес с целью поиска необходимой информации, приобретения знаний, использования этих знаний для решения учебных, научных и профессиональных задач, представляет собой важную составляющую учебного процесса, которой отводится не менее половины учебного времени при очной форме обучения. Время, отводимое на самостоятельную работу, должно использоваться обучающимися для наиболее полного освоения учебной дисциплины. Следовательно, организация эффективной внеаудиторной самостоятельной работы в процессе обучения требует создания условий, призванных обеспечить рациональное и планомерное управление учебной деятельностью, протекающей в отсутствие преподавателя.

К числу методических пособий относятся:

- задания для самостоятельного выполнения;

- литература по теме курса;

- сайт поддержки курса в сети интернет, на котором обучающиеся публикуют свои самостоятельные работы и имеют возможность задавать вопросы преподавателю и друг другу.

Роль преподавателя в организации самостоятельной работы состоит в координации действий обучающихся в освоении дисциплины, в методическом и организационном обеспечении учебного процесса. Взаимодействие между преподавателем и обучающимся осуществляется в форме консультаций, а также через сайт поддержки курса. Преподаватели также оказывают помощь обучающимся по планированию и организации самостоятельной работы.

**3.1.3 Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания**

Экзамен проводится в устной форме. Билет состоит из двух вопросов, на подготовку ответа на которые даётся не менее одного академического часа (при подготовке можно пользоваться литературой). После ответа на вопросы билета преподаватель вправе задать дополнительные вопросы по любой теме из списка вопросов, вынесенных на экзамен. Количество и содержание дополнительных вопросов – на усмотрение преподавателя, принимающего экзамен. Результирующая оценка определяется по следующим критериям:

|  |  |  |
| --- | --- | --- |
| Дан полный ответ на вопросы по билету, приведены доказательства необходимых утверждений, обнаружено всестороннее, глубокое и систематическое знание учебного материала. Обучающийся усвоил взаимосвязь основных понятий курса, проявил творческие способности в понимании и использовании учебного материала. | A | отлично |
| Допущены неточности в ответе, непринципиальные ошибки, исправленные самостоятельно после наводящих вопросов, доказательства утверждений приведены схематично. Обучающийся способен к самостоятельному пополнению и обновлению знаний по курсу в ходе дальнейшей учебы и профессиональной деятельности. | B | хорошо |
| Допущены неточности в ответе, доказательства утверждений приведены схематично. Обучающийся способен к самостоятельному пополнению и обновлению знаний по курсу в ходе дальнейшей учебы и профессиональной деятельности. | C |
| Допущены существенные ошибки, но обучающийся обладает необходимыми знаниями для их устранения под руководством преподавателя. Знание учебного материала в объеме, необходимом для дальнейшей учебы и профессиональной деятельности (основные определения, формулы, формулировки утверждений). Обучающийся знаком с основной литературой, рекомендованной программой. | D | удовлетворительно |
| Допущены грубые ошибки. Знание учебного материала в объеме, необходимом для дальнейшей учебы и профессиональной деятельности (основные определения, формулы, формулировки утверждений). | E |
| В остальных случаях | F | неудовлетворительно |

**3.1.4 Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)**

*Примерный список вопросов к экзамену*:

1. Instance based learning (IBL). Проклятие размерности.
2. Практические аспекты реализации IBL.
3. Обучение метрикам (по Brian Kulis).
4. Обзор методов уменьшения размерности. Johnson-Lindenstrauss Lemma.
5. Выбор свойств (feature selection).
6. Выделение новых свойств: PCA, ICA.
7. Кластерный анализ. Самоорганизующиеся сети Кохоннена.
8. Персептронные сети. Обратное распространение ошибки.
9. Сети Хопфилда, Bolzman machine. Обзор глубокого обучения (Deep Learning).
10. Деревья решений: ID3/C4.5. CART, забывчивые деревья.
11. Bagging. Случайные леса. Boosting на примере AdaBoost.
12. Gradient Boosting. MatrixNet.

**3.1.5 Методические материалы для оценки обучающимися содержания и качества учебного процесса**

Оценка обучающимися содержания и качества учебного процесса по дисциплине осуществляется в установленном в СПбГУ порядке.

**3.2. Кадровое обеспечение**

**3.2.1 Образование и (или) квалификация штатных преподавателей и иных лиц, допущенных к проведению учебных занятий**

К преподаванию дисциплины могут быть допущены преподаватели, имеющие диплом о высшем образовании по соответствующему направлению.

**3.2.2 Обеспечение учебно-вспомогательным и (или) иным персоналом**

Специальных требований нет.

**3.3. Материально-техническое обеспечение**

**3.3.1 Характеристики аудиторий (помещений, мест) для проведения занятий**

В аудиториях, где проводятся занятия, необходимо наличие досок и средств письма на них.

**3.3.2 Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования**

Аудитории для проведения занятий должны быть оснащены проекционной техникой и компьютером.

**3.3.3 Характеристики специализированного оборудования**

Специальных требований нет

**3.3.4 Характеристики специализированного программного обеспечения**

Специальных требований нет

**3.3.5 Перечень и объёмы требуемых расходных материалов**

Специальных требований нет

**3.4. Информационное обеспечение**

**3.4.1 Список обязательной литературы**

Не требуется

**3.4.2 Список дополнительной литературы**

Не требуется

**3.4.3 Перечень иных информационных источников**

1. T. Hastie, R. Tibshirani, J. Friedman “The Elements of Statistical Learning”. Springer. 2009.
2. Нестеров Ю.Е. «Методы выпуклой оптимизации», М.: Издательство МЦНМО, 2010.
3. R. Michalski, J. Carbonell, and T. Mitchell (eds.), “Machine Learning: An Artificial Intelligence” Approach, Tioga Press, 1983.

**Раздел 4. Разработчики программы**

Михайлова Елена Георгиевна, к.ф.-м.н., доцент кафедры информационно-аналитических систем СПбГУ, e.mikhaylova@spbu.ru. +7 (921) 9979791

По материалам и при поддержке Кураленка Игоря Евгеньевича, к.ф.-м.н., руководителя отдела мобильного поиска ООО «Яндекс», [solar@yandex-team.ru](mailto:solar@yandex-team.ru), +7 (812) 633-3600, доб. 8523.