**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: 9

Регистрационный номер рабочей программы: 057548

Санкт-Петербург

2020

**Раздел 1. Характеристики учебных занятий**

**1.1. Цели и задачи учебных занятий**

Сообщение сведений об основных принципах теории машинного обучения. Усвоение основных идей и понятий машинного обучения.

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

Курсы "Математическая статистика" и "Теоретическая информатика".

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

Обучающийся должен овладеть теоретическим материалом в объеме, предусмотренном программой, уметь применять полученные знания при решении теоретических и прикладных задач. Дисциплина участвует в формировании компетенций обучающихся по образовательной программе, установленных учебным планом для данной дисциплины.

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

Лекции 96 часов, практические занятия 94 часа, промежуточная аттестация (1 зачёт и 2 экзаменa) 6 часов.

**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 6 | 48 |  | 2 | 48 |  |  |  |  | 2 |  |  |  | 52 |  | 28 |  | 52 | 5 |
|  | 2-100 |  | 2-100 | 10-25 |  |  |  |  | 2-100 |  |  |  | 1-1 |  | 1-1 |  |  |  |
| Семестр 7 | 48 |  | 2 | 46 |  |  |  |  | 4 |  |  |  | 18 |  | 26 |  | 52 | 4 |
|  | 2-100 |  | 2-100 | 10-25 |  |  |  |  | 2-100 |  |  |  | 1-1 |  | 1-1 |  |  |  |
| ИТОГО | 96 |  | 4 | 94 |  |  |  |  | 6 |  |  |  | 70 |  | 54 |  |  | 9 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 6 |  |  | экзамен, устно, традиционная форма | по графику промежуточной аттестации |  |  |
| Семестр 7 |  |  | зачёт, устно, традиционная форма, экзамен, устно, традиционная форма | по графику промежуточной аттестации, по графику промежуточной аттестации |  |  |

**2.2. Структура и содержание учебных занятий**

Период обучения (модуль): **Семестр 6**

|  |  |  |  |
| --- | --- | --- | --- |
| № п/п | Наименование темы (раздела, части) | Вид учебных занятий | Количество часов |
| 1 | **Вероятностные методы в искусственном интеллекте** | Лекции | 12 |
| практические занятия | 12 |
| в присутствии преподавателя |  |
| по методическим материалам | 13 |
| 2 | Байесовский вывод | Лекции | 24 |
| практические занятия | 24 |
| в присутствии преподавателя |  |
| по методическим материалам | 26 |
| 3 | Скрытые марковские модели | Лекции | 12 |
| практические занятия | 12 |
| в присутствии преподавателя |  |
| по методическим материалам | 13 |
| 4 | Экзамен | промежуточная аттестация (с.р.) | 22 |
| промежуточная аттестация (ауд.) | 2 |

Раздел 1: Вероятностные методы в искусственном интеллекте.

1. Введение. История искусственного интеллекта. Вспоминаем теорию вероятностей. Теорема Байеса и машинное обучение. Что умеет делать машинное обучение.

2. Правило Лапласа. Априорные распределения. Сопряжённые априорные распределения.

Раздел 2: Байесовский вывод.

1. Наименьшие квадраты и ближайшие соседи. Линейная регрессия. Логистическая регрессия.

2. Статистическая теория принятия решений. Разложение bias-variance-noise. Оверфиттинг. Регуляризация: гребневая регрессия. Линейная регрессия по-байесовски.

3. Линейная регрессия: разные формы регуляризаторов. Лассо-регрессия. Эквивалентные ядра. Проклятие размерности.

4. Задачи классификации. Линейный дискриминант Фишера. Наивный байесовский классификатор: мультиномиальный и многомерный.

5. Логистическая регрессия: как обучать. Мультиклассовая логистическая регрессия. Аппроксимация по Лапласу. Пробит. Логистическая регрессия по-байесовски.

6. Метод опорных векторов (SVM). Трюк с ядрами.

7. Варианты SVM. SVM по-байесовски: relevance vector machines.

8. Кластеризация: иерархическая, методами теории графов. Алгоритм EM для кластеризации.

Раздел 3: Скрытые марковские модели.

9. Скрытые марковские модели.

10. Комбинация моделей: усреднение, бутстрап, бэггинг. Бустинг: AdaBoost.

11. Обучение ранжированию: постановка задачи, RankBoost. LambdaRank.

12. Рекомендательные системы: метод ближайших соседей, сингулярное разложение матриц.

Период обучения (модуль): **Семестр 7**

|  |  |  |  |
| --- | --- | --- | --- |
| № п/п | Наименование темы (раздела, части) | Вид учебных занятий | Количество часов |
| 1 | Графические модели | Лекции | 14 |
| практические занятия | 14 |
| в присутствии преподавателя |  |
| по методическим материалам | 16 |
| 2 | Нейронные сети | Лекции | 34 |
| практические занятия | 32 |
| в присутствии преподавателя |  |
| по методическим материалам | 36 |
| 3 | Зачёт | промежуточная аттестация (с.р.) | 30 |
| промежуточная аттестация (ауд.) | 2 |
| 4 | Экзамен | промежуточная аттестация (с.р.) | 30 |
| промежуточная аттестация (ауд.) | 2 |

Раздел 1: Графические модели.

1. Графические модели: определения, обозначения, примеры.

2. Маргинализация в общем виде, вывод на графе без циклов.

3. Вывод на графе с циклами: вариационные приближения. Алгоритм EM в общем виде.

4. Сэмплирование как метод приближённого вычисления. Методы сэмплирования.

5. Тематическое моделирование и модель LDA.

6. Вывод в моделях со сложными факторами: Expectation Propagation. Байесовские рейтинг-системы.

Раздел 2: Нейронные сети.

1. Нейронные сети: перцептрон. Виды функций активации. Обучение одного перцептрона. История развития нейронных сетей.

2. Как обучать сети I: градиентный спуск. Обратное распространение градиента на графе вычислений.

3. Как обучать сети II: как сделать градиентный спуск быстрее и лучше. Метод моментов, методы второго порядка и другие трюки.

4. Регуляризация в нейронных сетях. Дропаут и его мотивация. Другие методы.

5. Рекуррентные сети: базовые архитектуры, LSTM, GRU.

6. Свёрточные сети: архитектуры, как обучать, для чего они нужны.

7. Глубокие сети для обработки текстов I: распределённые представления слов.

8. Глубокие сети для обработки текстов II: рекурсивные нейронные сети, сети со стеком, сети с памятью.

9. Обучение с подкреплением: основы, определения, классические алгоритмы.

10. Как работает AlphaGo: обучение с подкреплением на глубоких сетях. DQN.

11. Соединяем байесовский вывод и глубокие сети: нейробайесовские методы.

**Раздел 3. Обеспечение учебных занятий**

**3.1. Методическое обеспечение**

**3.1.1 Методические указания по освоению дисциплины**

Посещение лекций и практических занятий

**3.1.2 Методическое обеспечение самостоятельной работы**

Основная литература

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

**Методика проведения зачёта**

Зачет проводится в устной форме. Для получения зачета необходимо решить 60% задач, предлагаемых в течение семестра. В случае, если к моменту проведения зачета студент решил меньшее количество задач, на зачете ему предлагаются задачи, аналогичные по тематике и сложности. Задачи даются в форме домашних заданий с устной сдачей («листочки»), письменных домашних заданий и контрольных. Темы задач фиксированы, количество и форма выдачи остается на усмотрение преподавателя практических занятий. Возможна выдача задач повышенной сложности, решение которых засчитывается в качестве индивидуальных достижений студента (при подаче заявок на именные стипендии, конкурсы и т.п.); сдача таких заданий проводится в устной форме.

**Методика проведения экзамена**

Экзамен проводится в устной форме. Билет состоит из двух вопросов. Время подготовки ответа на вопросы билета составляет 60 минут.

Использование конспектов и учебников, а также электронных устройств хранения, обработки или передачи информации при подготовке и ответе на вопросы экзамена категорически запрещено. В случае обнаружения факта использования недозволенных материалов (устройств) составляется акт и студент удаляется с экзамена. После ответа на вопросы билета преподаватель задает несколько дополнительных вопросов, на основании оценки ответов на которые итоговая оценка по предмету может быть повышена или понижена.

Критерии выставления оценок

Оценка «отлично» ставится за полностью раскрытый теоретический материал и правильные ответы на дополнительные вопросы преподавателя. В болонской шкале оценка может быть скорректирована в ту или иную сторону с учетом малозначительных погрешностей изложения или, напротив, углубленного изложения материала.

Оценка «хорошо» ставится за изложенный теоретический материал билета (возможно с помощью наводящих подсказок преподавателя).

Оценка «удовлетворительно» ставится за знание основных вопросов по каждой теме.

Оценка «неудовлетворительно» выставляется, если не выполняются условия для получения оценок «отлично», «хорошо» и «удовлетворительно».

Соответствие оценки СПбГУ и оценки ECTS (Европейской системы переноса и накопления зачётных единиц):

|  |  |  |  |
| --- | --- | --- | --- |
| Итоговый процент выполнения, % | Оценка СПбГУ при  проведении зачёта | Оценка ECTS | Оценка СПбГУ при  проведении экзамена |
| 90-100 | зачтено | A | отлично |
| 80-89 | зачтено | B | хорошо |
| 70-79 | зачтено | C | хорошо |
| 60-69 | зачтено | D | удовлетворительно |
| 50-59 | зачтено | E | удовлетворительно |
| менее 50 | не зачтено | F | неудовлетворительно |

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

Период обучения (модуль): **Семестр 1**

Темы задач для практических занятий:

1. Вычисление линейной регрессии для данной модели.

2. Вычисление логистической регрессии для данной модели.

3. Программирование линейной регрессии.

4. Программирование логистической регрессии.

5. Программирование метода опорных векторов.

6. Исследование данных скрытых марковских моделей.

7. Программирование скрытых марковских моделей.

8. Программирование рекомендательных систем.

**Список вопросов к экзамену**:

1. Теорема Байеса.

2. Правило Лапласа. Априорные распределения. Сопряжённые априорные распределения.

3. Наименьшие квадраты и ближайшие соседи. Линейная регрессия. Логистическая регрессия.

4. Разложение bias-variance-noise. Оверфиттинг.

5. Регуляризация. Линейная регрессия по-байесовски.

6. Линейная регрессия. Лассо-регрессия.

7. Задачи классификации. Линейный дискриминант Фишера

8. Мультиномиальный и многомерный байесовский классификатор.

9. Мультиклассовая логистическая регрессия. Аппроксимация по Лапласу. Пробит.

10. Логистическая регрессия по-байесовски.

11. Метод опорных векторов (SVM). Трюк с ядрами.

12. Варианты SVM, SVM по-байесовски.

13. Кластеризация: иерархическая, методами теории графов. Алгоритм EM для кластеризации.

14. Скрытые марковские модели.

15. Комбинация моделей: усреднение, бутстрап, бэггинг. Бустинг: AdaBoost.

16. Обучение ранжированию: постановка задачи, RankBoost. LambdaRank.

17. Рекомендательные системы: метод ближайших соседей.

18. Рекомендательные системы: сингулярное разложение матриц.

Период обучения (модуль): **Семестр 2**

Темы задач для практических занятий:

1. Задачи на свойства графических моделей.

2. Программирование графических моделей.

3. Программирование методов сэмплирования.

4. Разработка перцептронов для данных задач.

5. Программирование градиентного спуска.

6. Программирование рекуррентных сетей.

7. Разработка и программирование систем обработки текстов.

8. Применение модели AlphaGo.

**Список вопросов к экзамену**:

1. Графические модели. Примеры.

2. Маргинализация. Вывод на графе без циклов.

3. Вывод на графе с циклами. Алгоритм EM.

4. Сэмплирование. Методы сэмплирования.

5. Тематическое моделирование. Модель LDA.

6. Вывод в моделях со сложными факторами.

7. Байесовские рейтинг-системы.

8. Перцептрон. Функции активации. Обучение перцептрона.

9. Градиентный спуск. Обратное распространение градиента.

10. Методы градиентного спуска.

11. Регуляризация в нейронных сетях.

12. Рекуррентные сети.

13. Свёрточные сети.

14. Глубокие сети для обработки текстов: распределённые представления слов.

15. Глубокие сети для обработки текстов: рекурсивные нейронные сети, сети со стеком, сети с памятью.

16. Обучение с подкреплением.

17. Система AlphaGo.

18. Нейробайесовские методы.

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

Анкета для студентов для оценки качества преподавания курса.

Просим Вас заполнить анкету-отзыв по прочитанной дисциплине. Обобщенные данные анкет будут использованы для ее совершенствования. По каждому вопросу проставьте соответствующие оценки по шкале от 1 до 10 баллов (обведите выбранный Вами балл). В

случае необходимости впишите свои комментарии.

1. Насколько Вы удовлетворены содержанием дисциплины в

целом?

1 2 3 4 5 6 7 8 9 10

Комментарий\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. Насколько Вы удовлетворены общим стилем преподавания?

1 2 3 4 5 6 7 8 9 10

Комментарий\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. Как Вы оцениваете качество подготовки предложенных

методических материалов?

1 2 3 4 5 6 7 8 9 10

Комментарий\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Насколько Вы удовлетворены использованием

преподавателями активных методов обучения?

1 2 3 4 5 6 7 8 9 10

Комментарий\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

5. Какой из модулей (разделов) дисциплины Вы считаете наиболее полезным, ценным с точки зрения дальнейшего обучения и/или

применения в последующей практической деятельности?

Комментарий\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

6. Что бы Вы предложили изменить в методическом и

содержательном плане для совершенствования преподавания данной

дисциплины?

Комментарий\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

СПАСИБО!

**3.2. Кадровое обеспечение**

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

К чтению лекций должны привлекаться преподаватели, имеющие ученую степень доктора или кандидата наук (в том числе степень PhD, прошедшую установленную процедуру признания и установления эквивалентности) и/или ученое звание профессора или доцента.

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

не требуется

**3.3. Материально-техническое обеспечение**

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

Учебные аудитории для проведения учебных занятий, оснащенные стандартным оборудованием, используемым для обучения в СПбГУ в соответствии с требованиями материально-технического обеспечения.

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

Стандартное оборудование, используемое для обучения в СПбГУ. MS Windows, MS Office, Mozilla FireFox, Google Chrome, Acrobat Reader DC, WinZip, Антивирус Касперского.

**3.3.3 Характеристики специализированного оборудования**

не требуется

**3.3.4 Характеристики специализированного программного обеспечения**

не требуется

**3.3.5 Перечень и объёмы требуемых расходных материалов**

Мел — не менее 1 куска на час лекционных занятий, фломастеры для доски, губка.

**3.4. Информационное обеспечение**

**3.4.1 Список обязательной литературы**

1. Николенко, С., Кадурин, А., и Архангельская, Е. Глубокое обучение: Погружение в мир нейронных сетей. — Издательский дом «Питер», 2018. — 480 с.

2. Goodfellow, I., Bengio, Y., and Courville, A. Deep Learning. — MIT Press, 2016. — 800 p.

3. Bishop, C.M. Pattern Recognition and Machine Learning. — Springer, 2006. — xx + 738 p.

**3.4.2 Список дополнительной литературы**

1. Murphy, K.P. Machine Learning: A Probabilistic Perspective. — MIT Press, 2012. — 1104 p.

2. MacKay, D.J.C. Information Theory, Inference and Learning Algorithms. — Cambridge University Press, 2003. — 640 p.

3. Hastie, T., Tibshirani, R., and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer, 2009. — xxii + 745 p.

**3.4.3 Перечень иных информационных источников**

1. Сайт Научной библиотеки им. М. Горького СПбГУ: http://www.librarv.spbu.ru/

2. Электронный каталог Научной библиотеки им. М. Горького СПбГУ: http://www.librarv.spbu.ru/cgibin/irbis64r/cgiirbis 64.ехе?С21 COM=F&I21 DBN=IBIS&P21 DBN=IBIS

3. Перечень электронных ресурсов, находящихся в доступе СПбГУ: http://cufts.librarv.spbu.ru/CRDB/SPBGU/

4. Перечень ЭБС, на платформах которых представлены российские учебники, находящиеся в доступе СПбГУ: http://cufts.library.spbu.ru/CRDB/SPBGU/browse?name=rures&resource tvpe=8

**Раздел 4. Разработчики программы**

Николенко Сергей Игоревич, кандидат физико-математических наук, доцент СПбГУ, sergey@logic.pdmi.ras.ru