МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное образовательное учреждение высшего образования**

**«Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского»**

|  |
| --- |
| **Институт информационных технологий, математики и механики** |

|  |
| --- |
|  |
| УТВЕРЖДЕНО  решением ученого совета ННГУ  протокол от  «\_\_\_» \_\_\_\_\_\_\_\_\_\_ 20\_\_ г. № \_\_\_ |

**Рабочая программа дисциплины**

**Машинное обучение**

Уровень высшего образования

**магистратура**

Направление подготовки (специальность)

**020402 Фундаментальная информатика и информационные технологии**

Профиль подготовки (специализация)

**Когитивные системы**

Форма обучения

**очная**

Нижний Новгород

2020 **1. Место дисциплины в структуре ООП**

Дисциплина Б1.В.01 «Машинное обучение» относится к Части, формируемой участниками образовательных отношений, Блока 1 «Дисциплины (модули)» направления подготовки 02.04.02 «Фундаментальная информатика и информационные технологии» профиля подготовки «Когнитивные системы». Дисциплина преподается в 1 семестре. Трудоемкость дисциплины составляет 5 зачетных единиц, 180 час., экзамен.

|  |  |  |
| --- | --- | --- |
| **№ варианта** | **Место дисциплины в учебном плане образовательной программы** | **Стандартный текст для автоматического заполнения в конструкторе РПД** |
| 2 | Блок 1. Дисциплины (модули) Часть, формируемая участниками образовательных отношений | Дисциплина Б1.В.01 «Машинное обучение» относится к части ООП направления подготовки 02.04.02 «Фундаментальная информатика и информационные технологии», формируемой участниками образовательных отношений. |

1. **Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Формируемые компетенции** (код, содержание компетенции) | **Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции** | | **Наименование оценочного средства** |
| **Индикатор достижения компетенции**\*  (код, содержание индикатора) | **Результаты обучения**  **по дисциплине\*\*** |
| ПК-10. Способен конвертировать результаты научно-исследовательских и/или опытно-конструкторских работ в требования ИТ-проекта и обратно: способен обеспечить ИТ-проект необходимым исследованием и опытно-конструкторскими работами | ПК-10.1.  Знать проблематику и методы научных исследований и опытно-конструкторских ИТ-разработок в области когнитивных систем (КС). | **Знать**алгоритмы и методы машинного обучения | собеседование |
| ПК-10.2.  Иметь навыки выполнения научных исследований и опытно-конструкторских ИТ-разработок в области КС | **Уметь**использовать методы и алгоритмы машинного обучения на практике, оценивать качество методов.  **Владеть** современными программными средствами для решения задач машинного обучения: библиотекой Scikit-Learn, или средой языка Python или средой R для статистический вычислений;опытом реализации программных систем для решения практических задач с использованием методов машинного обучения | собеседование, задания |

1. **Структура и содержание дисциплины**

**3.1. Трудоемкость дисциплины**

|  |  |
| --- | --- |
| **Общая трудоемкость** | **5 ЗЕТ** |
| **Часов по учебному плану** | **180** |
| **в том числе:** |  |
| **аудиторные занятия (контактная работа):**  **- занятия лекционного типа**  **- занятия семинарского типа**  **- занятия лабораторного типа**  **- текущий контроль (КСР)** | **66**  **32**  **32**    **2** |
| **самостоятельная работа** | **78** |
| **Промежуточная аттестация - экзамен** | **36** |

**3.2 Содержание дисциплины**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| №  п/п | Наименование и краткое содержание разделов и тем дисциплины (модуля),  форма промежуточной аттестации по дисциплине (модулю) | Часов | | | | | |
| Всего | В том числе | | | | |
| Контактная работа (работа во взаимодействии с преподавателем), часы  из них | | | | Самостоятельная работа обучающегося |
| Занятия лекционного типа | Занятия семинарского типа | Занятия лабораторного типа | Всего |
| 1. | Содержательная постановка задачи машинного обучения. Дедуктивное и индуктивное обучения. Признаковое описание объектов. Обучения с учителем (обучение по прецедентам). Решающая функция (решающее правило). Обобщающая способность решающей функции (проблема качества обучения). Задачи классификации и задачи восстановления регрессии. Обучение без учителя. Примеры практических задач. | 8 | 2 | 2 |  | 4 | 4 |
| 2. | Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Регрессионная функция. Байесов классификатор. Метод ближайшего соседа в задачах классификации и восстановления регрессии. | 9 | 2 | 2 |  | 4 | 5 |
| 3. | Экспериментальные методы оценки качества обучения. Разделение данных на обучающую и тестовую выборки. Метод скользящего контроля. | 9 | 2 | 2 |  | 4 | 5 |
| 4. | Метод наименьших квадратов для решения задачи восстановления регрессии. Его вывод на основе метода максимального правдоподобия. Линейная регрессионная модель. Система нормальных уравнений. Основы регрессионного анализа (проверка значимости коэффициентов, коэффициент детерминации Пирсона, доверительные интервалы, анализ остатков). | 9 | 2 | 2 |  | 4 | 5 |
| 5. | Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов. | 9 | 2 | 2 |  | 4 | 5 |
| 6. | Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа. | 9 | 2 | 2 |  | 4 | 5 |
| 7. | Наивный байесовский классификатор. | 9 | 2 | 2 |  | 4 | 5 |
| 8. | Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия. | 8 | 2 | 2 |  | 4 | 4 |
| 9. | Нейронные сети. Персептрон Розенблатта. Алгоритм обучения персептрона как метод стохастического градиентного спуска. Нейронные сети для решения задач классификации и восстановления регрессии. Обучение сети. Регуляризация как метод борьбы с переобучением. Понятие о глубоких нейронных сетях. | 8 | 2 | 2 |  | 4 | 4 |
| 10. | Машина опорных векторов. Ядра и спрямляющие пространства. | 10 | 2 | 2 |  | 4 | 6 |
| 11. | Деревья решений. Метод CART (classification and regression trees) для решения задач классификации и восстановления регрессии. Отсечения ветвей и выбор финального дерева. Методы обработки пропущенных значений. | 10 | 2 | 2 |  | 4 | 6 |
| 12. | Ансамбли решающих правил (классификаторов). Простое и взвешенное голосования. Бустинг. Алгоритм AdaBoost. Оценка ошибки предсказания. Бустинг и аддитивные модели. Градиентный бустинг. Алгоритм градиентного бустинга деревьев решений (MART). Баггинг. Алгоритм случайных деревьев (.случайный лес.). | 10 | 2 | 2 |  | 4 | 6 |
| 13. | Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Метод медиан. Метод нечетких множеств. EM-алгоритм. | 10 | 2 | 2 |  | 4 | 6 |
| 14. | Иерархическая кластеризация. Агломеративные и разделяющие методы. | 10 | 2 | 2 |  | 4 | 6 |
| 15 | Основы теории Вапника–Червоненкиса. Лемма Бернштейна. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечного класса решающих правил. Обоснование принципа минимизации эмпирического риска. Размерность Вапника– Червоненкиса. Лемма Зауэра. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечной размерности Вапника–Червоненкиса. Принцип структурной минимизации риска. | 14 | 4 | 4 |  | 8 | 6 |
|  | Текущий контроль | 2 |  | 2 |  |  |  |
|  | Промежуточная аттестация: экзамен | 36 |  |  |  |  |  |
|  | Итого | 144 | 32 | 34 |  | 66 | 78 |

Текущий контроль успеваемости реализуется в формах опросов на занятиях семинарского типа. Промежуточная аттестация проходит в традиционных формах (экзамен).

1. **Учебно-методическое обеспечение самостоятельной работы обучающихся**

Самостоятельная работа обучающихся осуществляется в виде работы с рекомендованной обязательной и дополнительной литературой (приведена в разделе 6), подготовке к лекциям, подготовке к экзамену и выполнения лабораторных работ. Контрольные вопросы и задания для проведения текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведены в п. 5.2.

1. **Фонд оценочных средств для промежуточной аттестации по дисциплине**, включающий:
   1. **Описание шкал оценивания результатов обучения по дисциплине**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Уровень сформированности компетенций (индикатора достижения компетенций)** | **Шкала оценивания сформированности компетенций** | | | | | | |
| **плохо** | **неудовлетворительно** | **удовлетворительно** | **хорошо** | **очень хорошо** | **отлично** | **превосходно** |
| Не зачтено | | зачтено | | | | |
| Знания | Отсутствие знаний теоретического материала.  Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа | Уровень знаний ниже минимальных требований. Имели место грубые ошибки. | Минимально допустимый уровень знаний. Допущено много негрубых ошибки. | Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок | Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок | Уровень знаний в объеме, соответствующем программе подготовки, без ошибок. | Уровень знаний в объеме, превышающем программу подготовки. |
| Умения | Отсутствие минимальных умений . Невозможность оценить наличие умений вследствие отказа обучающегося от ответа | При решении стандартных задач не продемонстрированы основные умения.  Имели место грубые ошибки. | Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания но не в полном объеме. | Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами. | Продемонстрированы все основные умения. Решены все основные задачи . Выполнены все задания, в полном объеме, но некоторые с недочетами. | Продемонстрированы все основные умения,решены все основные задачи с отдельными несущественным недочетами, выполнены все задания в полном объеме. | Продемонстрированы все основные умения,. Решены все основные задачи. Выполнены все задания, в полном  объеме без недочетов |
| Навыки | Отсутствие владения материалом. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа | При решении стандартных задач не продемонстрированы базовые навыки.  Имели место грубые ошибки. | Имеется минимальный  набор навыков для решения стандартных задач с некоторыми недочетами | Продемонстрированы базовые навыки  при решении стандартных задач с некоторыми недочетами | Продемонстрированы базовые навыки  при решении стандартных задач без ошибок и недочетов. | Продемонстрированы навыки  при решении нестандартных задач без ошибок и недочетов. | Продемонстрирован творческий подход к решению нестандартных задач |

**Шкала оценки при промежуточной аттестации**

|  |  |  |
| --- | --- | --- |
| **Оценка** | | **Уровень подготовки** |
| зачтено | Превосходно | Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно» |
| Отлично | Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично», при этом хотя бы одна компетенция сформирована на уровне «отлично» |
| Очень хорошо | Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо», при этом хотя бы одна компетенция сформирована на уровне «очень хорошо» |
| Хорошо | Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо», при этом хотя бы одна компетенция сформирована на уровне «хорошо» |
| Удовлетворительно | Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно» |
| не зачтено | Неудовлетворитель-  но | Хотя бы одна компетенция сформирована на уровне «неудовлетворительно», ни одна из компетенций не сформирована на уровне «плохо» |
| Плохо | Хотя бы одна компетенция сформирована на уровне «плохо» |

* 1. **Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения**

**5.2.1 Контрольные вопросы**

|  |  |
| --- | --- |
| Вопрос | Код формируемой компетенции |
| 1. Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Байесов классификатор. | ПК-10.1 |
| 1. Экспериментальные методы оценки качества обучения. | ПК-10.1 |
| 1. Метод наименьших квадратов для решения задачи восстановления регрессии. Его вывод на основе метода максимального правдоподобия. Линейная регрессионная модель. Система нормальных уравнений. Основы регрессионного анализа (проверка значимости коэффициентов, коэффициент детерминации Пирсона, доверительные интервалы, анализ остатков). | ПК-10.1 |
| 1. Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов. | ПК-10.1 |
| 1. Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа. | ПК-10.1 |
| 1. Наивный байесовский классификатор. | ПК-10.1 |
| 1. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия. | ПК-10.1 |
| 1. Нейронные сети. Персептрон Розенблатта. Алгоритм обучения персептрона как метод стохастического градиентного спуска. Нейронные сети для решения задач классификации и восстановления регрессии. Обучение сети. Регуляризация как метод борьбы с переобучением. Понятие о глубоких нейронных сетях. | ПК-10.1 |
| 1. Машина опорных векторов. Ядра и спрямляющие пространства. | ПК-10.1 |
| 1. Деревья решений. Метод CART (classification and regression trees) для решения задач классификации и восстановления регрессии. Отсечения ветвей и выбор финального дерева. Методы обработки пропущенных значений. | ПК-10.1 |
| 1. Ансамбли решающих правил (классификаторов). Простое и взвешенное голосования. Бустинг. Алгоритм AdaBoost. Оценка ошибки предсказания. Бустинг и аддитивные модели. Градиентный бустинг. Алгоритм градиентного бустинга деревьев решений (MART). Баггинг. Алгоритм случайных деревьев (.случайный лес.). | ПК-10.1 |
| 1. Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Метод медиан. Метод нечетких множеств. EM-алгоритм. | ПК-10.1 |
| 1. Иерархическая кластеризация. Агломеративные и разделяющие методы. | ПК-10.1 |
| 1. Основы теории Вапника–Червоненкиса. Лемма Бернштейна. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечного класса решающих правил. Обоснование принципа минимизации эмпирического риска. | ПК-10.1 |
| 1. Размерность Вапника– Червоненкиса. Лемма Зауэра. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечной размерности Вапника–Червоненкиса. Принцип структурной минимизации риска. | ПК-10.1 |

**5.2.2. Типовые задания/задачи для оценки сформированности компетенции ПК-10.2**

1. Дана обучающая выборка

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *x*1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 2 | 6 |
| *x2* | 3 | 3 | 1 | 0 | 1 | 1 | 2 | 3 | 1 |
| *y* | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |

Методом линейного дискриминантного анализа для каждого класса построить дискриминантную функцию и записать уравнение разделяющей поверхности.

1. Дана обучающая выборка (см. таблицу выше). Методом квадратичного дискриминантного анализа построить дискриминантные функции.
2. Дана обучающая выборка (см. таблицу выше). С помощью наивного байесова классификатора оценить вероятности P(*Y* = 1 | *x*1= 1, *x*1 = 2)
3. Дана обучающая выборка:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *x* | –1 | 0 | 0 | 1 | 2 |
| *y* | 1 | –2 | 1 | 7 | 8 |

Методом наименьших квадратов построить полиномиальную модель вида *f*(*x*) = β0 +  β1*x* + β2*x*2.

1. Дана обучающая выборка (см. выше). Методом ридж-регрессии построить полиномиальную модель вида *f*(*x*) = β0 +  β1*x* + β2*x*2 , если параметр регуляризации λ = 2.

**Критерии оценок выполнения заданий/задач**

(каждая задача оценивается в 2 балла)

|  |  |
| --- | --- |
| Решена полностью | 2 |
| Решена основная часть задачи, или задача решена с недочетами | 1,5 |
| Решена задача наполовину | 1 |
| Сделан первый этап в решении задачи | 0,5 |
| Нет решения | 0 |

**Суммарная оценка выполнения заданий/задач**

|  |  |
| --- | --- |
| **Количество баллов** | **Оценка** |
| 4 | Отлично |
| 3,5 | Очень хорошо |
| 3 | Хорошо |
| 2-2,5 | Удовлетворительно |
| 0,5-1,5 | Неудовлетворительно |
| 0 | Плохо |

**6. Учебно-методическое и информационное обеспечение дисциплины (модуля)**

а) Основная литература:

1. Воронцов К.В. Машинное обучение. Курс лекций. <http://www.intuit.ru/studies/courses/13844/1241/info> .

б) Дополнительная литература

1. Золотых Н.Ю.Машинное обучение. Курс лекций. Нижний Новгород: ННГУ, 2007. <http://www.uic.nnov.ru/~zny/ml>

**в) программное обеспечение и Интернет-ресурсы**

1. Международный открытый ресурс и программное обеспечение для машинного обучения Scikit-Learn: Machine Learning in Python: [www.scikit-learn.org](http://www.scikit-learn.org)
2. The R Project for Statistical Computing <https://www.r-project>
3. Welcome to Python.org <https://www.python.org/>

**7. Материально-техническое обеспечение дисциплины (модуля)**

Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой, оснащенные оборудованием и техническими средствами обучения: компьютерный класс, проектор, экран.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Учебная и научная литература, учебно-методические материалы, представленные в библиотечном фонде, в электронных библиотеках и на кафедре Алгебры, геометрии и дискретной математики.

Программа составлена в соответствии с требованиями ОС ВО ННГУ с учетом рекомендаций ФГОС ВО по направлению 02.04.02. – Фундаментальная информатика и информационные технологии.

Автор д.ф.-м.н., доц. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Золотых Н.Ю.

Заведующий кафедрой\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Кузнецов М.И.

Программа одобрена на заседании методической комиссии института информационных технологий, математики и механики

от 24.02.2021 года, протокол № 5