**Приложение**

**к Рабочей программе дисциплины**

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** |

(факультет / институт / филиал)

#### Кафедра алгебры, геометрии и дискретной математики

#### (наименование кафедры)

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

**ФОНД**

**ОЦЕНОЧНЫХ СРЕДСТВ**

#### ПО УЧЕБНОЙ ДИСЦИПЛИНЕ

**Машинное обучение**

(наименование дисциплины)

**02.04.02 Фундаментальная информатика и информационные технологии**

(код и наименование направления подготовки)

**Когнитивные системы**

(наименование профиля подготовки, направленности программы)

Нижний Новгород

2020

***Цель фонда оценочных средств.*** Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «*Машинное обучение*»*.* Перечень видов оценочных средств соответствует рабочей программе дисциплины.

***Фонд оценочных средств включает*** материалы для проведения текущего контроля в форме практических заданий и задач, вопросов для собеседования и итоговой аттестации в форме вопросов к экзамену.

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. **Перечень контрольных заданий и иных материалов, необходимых для оценки знаний, умений, навыков и опыта деятельности**
   1. **Вопросы к экзамену по дисциплине «Современные проблемы анализа данных»**

|  |  |
| --- | --- |
| Вопрос | Код компетенции (согласно РПД) |
| 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 |

* 1. **Типовые задания для текущего контроля успеваемости**
     1. **Задания (задачи) для оценки 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.
2. Доказать, что в случае квадратичной функции потерь минимум среднему риску доставляет условное среднее. Чему равен при этом средний риск?
3. Доказать, что если функция потерь равна модулю разности, то минимум среднему риску доставляет условная медиана. Чему равен при этом средний риск?
4. Пусть ответ задается в виде аналитической функции **x XOR ((y XOR z) OR w)**, где w, x, y и z – принимают значение TRUE или FALSE. Постройте дерево решений, предсказывающее ответ с нулевой ошибкой.
5. Загрузите набор данных Spam ((<http://www-stat.stanford.edu/~tibs/ElemStatLearn/)>. Разделите данные на обучающую и тестовую выборку (согласно меткам в файле spam.traintest). Сравните качество обучения с использованием метода опорных векторов и *K* ближайших соседей. Параметры моделей выберите на Ваше усмотрение.
6. Загрузите набор данных Spam ((<http://www-stat.stanford.edu/~tibs/ElemStatLearn/)>. Разделите данные на обучающую и тестовую выборку (согласно меткам в файле spam.traintest). Сравните качество обучения с использованием деревьев решений и метода *K* ближайших соседей. Параметры моделей выберите на Ваше усмотрение.
7. Загрузите набор данных Spam ((<http://www-stat.stanford.edu/~tibs/ElemStatLearn/)>. Разделите данные на обучающую и тестовую выборку (согласно меткам в файле spam.traintest). Сравните качество обучения с использованием деревьев решений и метода опорных векторов. Параметры моделей выберите на Ваше усмотрение.
   1. **Задания (оценочные средства), выносимые на экзамен**

На экзамен/зачет для оценки компетенции ПК-10.1 выносятся знания по вопросам п.2.1, и для оценки ПК-10.2 - умение, владение по задачам п.2.2. Теоретический вопрос определяется содержанием билета (см. Приложение 1), а номер задачи из списка задач п.2.2, определяется номером билета N по правилу: №Задачи = mod(№Билета,11)+1. Оценка экзамена складывается из указанных двух составляющих с весовыми коэффициентами 0.5.

Составитель:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Н.Ю. Золотых

«\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020