|  |
| --- |
| утверждаю  Руководитель Департамента  анализа данных, принятия решений и финансовых технологий  \_\_\_\_\_\_\_\_\_\_В.И. Соловьев  «15» октября 2019 г. |
|  |

ПРИЛОЖЕНИЕ К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ

**Прикладные модели и методы факторного, дискриминантного и кластерного анализа**

НАПРАВЛЕНИЕ ПОДГОТОВКИ: **01.04.02 «Прикладная математика и информатика»**

НАПРАВЛЕННОСТИ ПРОГРАММЫ МАГИСТРАТУРЫ: «**Анализ больших данных и машинное обучение в экономике и финансах», «Технологии блокчейн и криптовалюты»**

ФОРМА ОБУЧЕНИЯ: **Очная**

ГОД ПРИЕМА: **2019, 2020**

Год утверждения программы: 2017 год

*Одобрено департаментом анализа данных, принятия*

*решений и финансовых технологий*

*Протокол № 3 от «15» октября 2019 г.*

**Содержание Приложения**

|  |  |
| --- | --- |
| **Наименование разделов РПД** | **стр.** |
| Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине | 2 |
| Место дисциплины в структуре образовательной программы | 5 |
| Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся | 5 |
| Учебно-тематический план | 6 |
| Содержание семинаров, практических занятий | 7 |
| Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы | 9 |
| Перечень вопросов, заданий, тем для подготовки к текущему контролю | 10 |
| Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине | 14 |
| Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем | 21 |

**2. Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине**

***НПМ - Анализ больших данных и машинное обучение в экономике и финансах***

|  |  |  |  |
| --- | --- | --- | --- |
| **Код компетенции** | **Наименование**  **компетенции** | **Индикаторы достижения компетенции** | **Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции** |
| **ДКН-1** | Способность проводить поиск источников данных, подготавливать данные для анализа, визуализировать данные | 1. Демонстрирует знание общих принципов визуального представления абстрактных понятий. | **Знать** основные базы данных, используемых в экономических расчетах, методы и алгоритмы подготовки и визуализации данных.  **Уметь** решать задачи поиска, подготовки и визуализации данных с использованием MS Excel, PYTHON и R. |
| 2. Проводит поиск источников данных, подготавливает данные для анализа, визуализирует данные. | **Знать** методы поиска источников данных, подготовки данных для анализа, визуализации данных.  **Уметь** Проводить поиск источников данных, подготовить данные для анализа, визуализировать данные |
| 3. Владеет профессиональной терминологией в области инфографики, техникой создания 2D и 3D образов цифровых данных и их использования для информационной поддержки при принятии решений | **Знать** профессиональную терминологию в области инфографики, техникой создания 2D и 3D образов цифровых данных  **Уметь** использовать профессиональную терминологию в области инфографики, техникой создания 2D и 3D образов цифровых данных для информационной поддержки при принятии решений |
| **ДКН-2** | Способность применять методы регрессионного, факторного, кластерного и дискриминантного анализа, анализа нечисловой информации и теории сложных сетей к решению прикладных задач в экономике и финансах | 1.Владеет методами регрессионного, факторного, кластерного и дискриминантного анализа. | **Знать** методы и алгоритмы решения задач регрессионного, факторного, кластерного и дискриминантного анализа.  **Уметь** решать задачи регрессионного, кластерного и  дискриминантного анализа с использованием MS  Excel, в PYTHON и R |
| 2.Анализирует нечисловую информацию и владеет методами теории сложных сетей | **Знать** методы анализа нечисловой информации и теории сложных сетей  **Уметь** анализироватьнечисловую информацию |
| 3.Решает прикладные экономические и финансовые задачи методами регрессивного, факторного, кластерного и дискриминантного анализа, анализа нечисловой информации и теории сложных сетей. | **Знать** прикладные экономические и финансовые задачи методами регрессивного, факторного, кластерного и дискриминантного анализа, анализа нечисловой информации и теории сложных сетей  **Уметь** Решать прикладные экономические и финансовые задачи методами регрессивного, факторного, кластерного и дискриминантного анализа, анализа нечисловой информации и теории сложных сетей |
| **ДКН-4** | Способность обосновывать и принимать решения с помощью технологий интеллектуального анализа данных и машинного обучения | 1.Владеет методикой принятия решений, основанной на технологиях интеллектуального анализа данных и машинного обучения. | **Знать** методы принятия решений, основанной на технологиях интеллектуального анализа данных и машинного обучения.  **Уметь** применять методы принятия решений, основанной на технологиях интеллектуального анализа данных и машинного обучения. |
| 2.Обосновывает принимаемые решения с помощью технологий интеллектуального анализа данных и машинного обучения. | **Знать** методы принятия решения с помощью технологий интеллектуального анализа данных и машинного обучения.  **Уметь** Обосновывать принимаемые решения с помощью технологий интеллектуального анализа данных и машинного обучения. |

***НПМ - Технологии блокчейн и криптовалюты***

|  |  |  |  |
| --- | --- | --- | --- |
| **Код компетенции** | **Наименование**  **компетенции** | **Индикаторы достижения компетенции** | **Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции** |
| **ДКН-1** | Способность применять технологии блокчейн и смарт- контракты к решению прикладных задач в экономике и финансах | 1. Решает вопрос о целесообразности применения технологии блокчейн для решения прикладной задачи. | **Знать** принципы организации технологии блокчейн.  **Уметь** разрабатывать смарт-контракты на одной из блокчейн платформ. |
| 2.Формулирует предложения по созданию новых подходов для решения имеющихся задач | **Знать** технологии блокчейн и смарт- контракты к решению прикладных задач в экономике и финансах  **Уметь** формироватьпредложения по созданию новых подходов для решения имеющихся задач |
| 3.Формулирует новые задачи, способствующие развитию отрасли, и предлагает пути их решения с использованием технологии блокчейн. | **Знать** технологии блокчейн и смарт- контракты к решению прикладных задач в экономике и финансах  **Уметь** Формулировать задачи, способствующие развитию отрасли, и предлагает пути их решения с использованием технологии блокчейн |
| **ОПК-1** | Способен решать актуальные задачи фундаментальной и прикладной математики | 1. Демонстрирует знания в области фундаментальной и прикладной математики. | **Знать** актуальные задачи фундаментальной и прикладной математики  **Уметь** решать актуальные задачи фундаментальной и прикладной математики |
| 2. Системно выбирает математические методы для решения прикладных задач. | **Знать** математические методы для решения прикладных задач  **Уметь** Системно выбирать математические методы для решения прикладных задач |
| 3. Осуществляет решение актуальных задач с использованием математических методов, в том числе методов обработки больших данных и машинного обучения. | **Знать** актуальные задачи и их решение с использованием математических методов, в том числе методов обработки больших данных и машинного обучения  **Уметь** решать актуальные задачи с использованием математических методов, в том числе методов обработки больших данных и машинного обучения. |
| **ОПК-3** | Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности | 1. Демонстрирует знания в области теории и методологии математического моделирования. | **Знать основы** теории и методологии математического моделирования  **Уметь** применять теории и методологии математического моделирования |
| 2. Демонстрирует умение разрабатывать математические модели в области экономики и финансов с использованием методов обработки больших данных и машинного обучения. | **Знать** математические модели в области экономики и финансов с использованием методов обработки больших данных и машинного обучения  **Уметь** разрабатывать математические модели в области экономики и финансов с использованием методов обработки больших данных и машинного обучения |
| 3. Анализирует математические модели, осуществляя интерпретацию результатов моделирования и оценивая степень достоверности найденных решений. | **Знать** математические модели, осуществляя интерпретацию результатов моделирования и оценивая степень достоверности найденных решений  **Уметь** Анализировать математические модели, осуществляя интерпретацию результатов моделирования и оценивая степень достоверности найденных решений |

**3. Место дисциплины в структуре образовательной программы**

Дисциплина **Прикладные модели и методы факторного, дискриминантного и кластерного анализа** относится к Модулю дисциплин по выбору, углубляющих освоение программы магистратуры «Технологии блокчейн и криптовалюты» направления подготовки 01.04.02 «Прикладная математика и информатика».

# **4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся**

***НПМ - Анализ больших данных и машинное обучение в экономике и финансах***

Общая трудоемкость дисциплины составляет 3 зачетные единицы.

Вид промежуточной аттестации - экзамен.

Вид текущего контроля – контрольная работа.

*Очная форма обучения – 2019, 2020 г.г.*

|  |  |  |
| --- | --- | --- |
| **Вид учебной работы по дисциплине** | **Всего**  **(в з/е и часах)** | **Модуль 3**  **(в часах)** |
| **Общая трудоемкость дисциплины** | **3/108** | **108** |
| ***Контактная работа-Аудиторные занятия*** | **40** | **40** |
| *Лекции* | 16 | 16 |
| *Семинары, практические занятия* | 24 | 24 |
| **Самостоятельная работа** | **68** | **68** |
| Вид текущего контроля | Контрольная работа | Контрольная работа |
| Вид промежуточной аттестации | Экзамен | Экзамен |

***НПМ - Технологии блокчейн и криптовалюты***

Общая трудоемкость дисциплины составляет 3 зачетные единицы.

Вид промежуточной аттестации - зачет.

Вид текущего контроля – контрольная работа.

*Очная форма обучения – 2019 г.*

|  |  |  |
| --- | --- | --- |
| **Вид учебной работы по дисциплине** | **Всего**  **(в з/е и часах)** | **Модуль 6**  **(в часах)** |
| **Общая трудоемкость дисциплины** | **3/108** | **108** |
| ***Контактная работа-Аудиторные занятия*** | **20** | **20** |
| *Лекции* | 10 | 10 |
| *Семинары, практические занятия* | 10 | 10 |
| **Самостоятельная работа** | **88** | **88** |
| Вид текущего контроля | Контрольная работа | Контрольная работа |
| Вид промежуточной аттестации | Зачет | Зачет |

**5.2. Учебно-тематический план**

***НПМ - Анализ больших данных и машинное обучение в экономике и финансах***

*Очная форма обучения – 2019, 2020 г.г.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| №  п/п | **Наименование тем (разделов) дисциплины** | **Трудоемкость в часах** | | | | | | **Формы текущего контроля успеваемости** |
| **Все**  **го** | **Аудиторная работа** | | | | **Самостоятельная работа** |
| Общ  ая, в т.ч.: | Лекции | Семинары, практические занятия | Занятия в интерактивных формах |
| 1 | |  | | --- | | Корреляционный | | анализ многомерно | | генеральной сово-  купности | | 27 | 10 | 4 | 6 | 2 | 17 | Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям. |
| 2 | |  | | --- | | Регрессионный анализ | | 27 | 10 | 4 | 6 | 2 | 17 |
| 3 | Методы снижения размерности многомерных признаков | 27 | 10 | 4 | 6 | 2 | 17 |
| 4 | Методы многомерной классификации | 27 | 10 | 4 | 6 | 2 | 17 |
|  | В целом по дисциплине | 108 | 40 | 16 | 24 | 8 | 68 | Контрольная работа |
|  | Итого в % |  |  |  |  | 20% |  |  |

***НПМ - Технологии блокчейн и криптовалюты***

*Очная форма обучения – 2019 г.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| №  п/п | **Наименование тем (разделов) дисциплины** | **Трудоемкость в часах** | | | | | | **Формы текущего контроля успеваемости** |
| **Все**  **го** | **Аудиторная работа** | | | | **Самостоятельная работа** |
| Общ  ая, в т.ч.: | Лекции | Семина  ры, практические занятия | Занятия в интерактивных формах |
| 1 | |  | | --- | | Корреляционный | | анализ многомерно | | генеральной сово-  купности | | 28 | 6 | 2 | 4 | 2 | 22 | Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям. |
| 2 | |  | | --- | | Регрессионный анализ | | 26 | 4 | 2 | 2 | 2 | 22 |
| 3 | Методы снижения размерности многомерных признаков | 27 | 5 | 3 | 2 | 2 | 22 |
| 4 | Методы многомерной классификации | 27 | 5 | 3 | 2 | 2 | 22 |
|  | В целом по дисциплине | 108 | 20 | 10 | 10 | 8 | 88 | Контрольная работа |
|  | Итого в % |  |  |  |  | 40% |  |  |

**5.3. Содержание семинаров, практических занятий**

***НПМ - Анализ больших данных и машинное обучение в экономике и финансах***

|  |  |  |
| --- | --- | --- |
| **Наименование тем (разделов) дисциплины** | **Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)** | **Формы проведения занятий** |
| 1. Корреляционный анализ многомерной генеральной совокупности | Статистические оценки многомерной генеральной совокупности. Корреляционная матрица и методы ее оценивания. Проверка статистических гипотез о параметрах. Модели случайных векторов.  *Рекомендуемые источники: п.8, [1]; п.9, [3], [4]* | Решение задач в интерактивной форме, проверка самостоятельной  работы и разбор ошибок |
| 1. Регрессионный анализ | |  | | --- | | Регрессионная модель и методы ее реализации. | | Основной алгоритм оценивания качества регрессионной модели | | Методы регуляризации Тихонова, LASSO.  *Рекомендуемые источники: п.8, [1]; п.9, [3], [4]* | | Решение задач в интерактивной форме, проверка самостоятельной  работы и разбор ошибок |
| 1. Методы снижения размерности многомерных признаков | |  | | --- | | Метод главных компонент. Многомерное шкали-  рование. Построение устойчивого SVD разложе-  ния. Нелинейные методы PCA.  *Рекомендуемые источники: п.8, [1]; п.9, [3], [4]* | | Решение задач в интерактивной форме, проверка самостоятельной  работы и разбор ошибок |
| 1. Методы многомерной классификации | Наивный Байесовский классификатор, логистическая модель. Линейные методы классификации. Ядровые методы классификации, оценка Надарайи.  *Рекомендуемые источники: п.8, [1]; п.9, [3], [4]* | Решение задач в интерактивной форме, проверка самостоятельной  работы и разбор ошибок |

***НПМ - Технологии блокчейн и криптовалюты***

|  |  |  |
| --- | --- | --- |
| **Наименование тем (разделов) дисциплины** | **Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)** | **Формы проведения занятий** |
| 1. Корреляционный анализ многомерной генеральной совокупности | Статистические оценки многомерной генеральной совокупности. Корреляционная матрица и методы ее оценивания. Проверка статистических гипотез о параметрах. Модели случайных векторов.  *Рекомендуемые источники: п.8, [1]; п.9, [3], [4]* | Решение задач в интерактивной форме, проверка самостоятельной  работы и разбор ошибок |
| 1. Регрессионный анализ | |  | | --- | | Регрессионная модель и методы ее реализации. | | Основной алгоритм оценивания качества регрессионной модели | | Методы регуляризации Тихонова, LASSO.  *Рекомендуемые источники: п.8, [1]; п.9, [3], [4]* | | Решение задач в интерактивной форме, проверка самостоятельной  работы и разбор ошибок |
| 1. Методы снижения размерности многомерных признаков | |  | | --- | | Метод главных компонент. Многомерное шкали-  рование. Построение устойчивого SVD разложе-  ния. Нелинейные методы PCA.  *Рекомендуемые источники: п.8, [1]; п.9, [3], [4]* | | Решение задач в интерактивной форме, проверка самостоятельной  работы и разбор ошибок |
| 1. Методы многомерной классификации | Наивный Байесовский классификатор, логистическая модель. Линейные методы классификации. Ядровые методы классификации, оценка Надарайи.  *Рекомендуемые источники: п.8, [1]; п.9, [3], [4]* | Решение задач в интерактивной форме, проверка самостоятельной  работы и разбор ошибок |

**6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы**

***НПМ - Анализ больших данных и машинное обучение в экономике и финансах***

|  |  |  |  |
| --- | --- | --- | --- |
| **Наименование тем** | **Перечень вопросов, отводимых на** | **Формы внеаудиторной** |  |
| **(разделов) дисциплины** | **самостоятельное освоение** | **самостоятельной работы** |  |
| 1. Корреляционный  анализ многомерной  генеральной совокупности | Понятие копулы, модели копул, методы оценивания параметрических моделей копул. Понятие эмпирической многомерной плостности распределения и методы ее оценивания, ядерные оценки. | Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию. |  |
| 2. Регрессионный анализ | Непараметрическая регрессия, методы построения. Ядерная оценка непараметрической регрессии. Байесовский подход к построению регрессии, байесовская линейная регрессия. | Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию. |  |
| 3. Методы снижения размерности многомерных  признаков | Неметрическое шкалировование  Применение регуляризации  к снижению размерности. | Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию |  |
| 4. Методы многомерной  классификации | Самоорганизующие карты Кохонена | Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию |  |

***НПМ - Технологии блокчейн и криптовалюты***

|  |  |  |  |
| --- | --- | --- | --- |
| **Наименование тем** | **Перечень вопросов, отводимых на** | **Формы внеаудиторной** |  |
| **(разделов) дисциплины** | **самостоятельное освоение** | **самостоятельной работы** |  |
| 1. Корреляционный анализ многомерной генеральной совокупности | Непараметрические методы оценивания ногомерной плотности. Ядерные оценки. Метод ближайших соседей. | Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию |  |
| 2. Регрессионный анализ | Непараметрические методы регрессионного анализа. Ядерная регрессия, квантильная регрессия. | Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию |  |
| 3. Методы снижения размерности многомер-  ных признаков | Алгоритм t-SNE (t-Distributed Stochastic Neighbor Embedding). | Работа с учебной литературой. Решение типовых задач.  Разбор вопросов по теме   |  | | --- | | занятия. Выполнение домашних заданий | | к каждому занятию. | |  |
| 4. Методы многомерной  классификации | Iterative Self-Organizing Data Analysis  (ISODATA) | Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию |  |

**6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю**

***Примерные задания к контрольной работе***

1. Объясните, с какой целью в метрических классификаторах используются ядерные функции.
2. [Модель Бернулли](http://ru.learnmachinelearning.wikia.com/wiki/%D0%9C%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C_%D0%91%D0%B5%D1%80%D0%BD%D1%83%D0%BB%D0%BB%D0%B8) и [мультиномиальная модель](http://ru.learnmachinelearning.wikia.com/wiki/%D0%9C%D1%83%D0%BB%D1%8C%D1%82%D0%B8%D0%BD%D0%BE%D0%BC%D0%B8%D0%B0%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C) классификации. Являются ли они линейными классификаторами? [Предположение "наивного Байеса"](http://ru.learnmachinelearning.wikia.com/wiki/%D0%9F%D1%80%D0%B5%D0%B4%D0%BF%D0%BE%D0%BB%D0%BE%D0%B6%D0%B5%D0%BD%D0%B8%D0%B5_%22%D0%BD%D0%B0%D0%B8%D0%B2%D0%BD%D0%BE%D0%B3%D0%BE_%D0%91%D0%B0%D0%B9%D0%B5%D1%81%D0%B0%22). Как оно работает в случае моделей Бернулли, мультиномиальной и Гауссова классификатора?
3. Сингулярное разложение. Пусть u1, u2,…,ud – собственные векторы матрицы XTX, а λ1, λ2,…, λd – соответствующие им собственные значения. Выразите ненулевые собственные векторы и собственные значения матрицы XXT через u1, u2,…,ud, λ1, λ2,…, λd. Выразите собственные векторы и собственные значения матрицы XTX через собственные векторы и собственные значения матрицы XXT. Какой вывод можно сделать относительно алгоритма вычисления сингулярного разложения матрицы X?
4. Что такое «проблема мультиколлинеарности», в каких задачах и при использовании каких алгоритмов она возникает? Какие есть подходы к её решению?
5. Какие объекты в SVM называются опорными?
6. В каких алгоритмах классификации можно узнать не только классовую принадлежность классифицируемого объекта, но и вероятность того, что данный объект принадлежит каждому из классов?
7. Сравнение [методов классификации](http://ru.learnmachinelearning.wikia.com/wiki/%D0%97%D0%B0%D0%B4%D0%B0%D1%87%D0%B0_%D0%BA%D0%BB%D0%B0%D1%81%D1%81%D0%B8%D1%84%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D0%B8): *k* ближайших соседей, логистическая регрессия с *L*1/*L*2 регуляризацией, дерево решений, метод опорных векторов с линейным и RBF-ядром, байесовское правило классификации? Какие методы всегда дают линейные границы между классами?
8. Дана обучающая выборка:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
|  | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |

1. С помощью наивного байесова классификатора оценить вероятности ;
2. Дана обучающая выборка



1. Найти уравнение оптимальной разделяющей гиперплоскости, указать опорные точки
2. Дана выборка



Найти главные направления и объясненные дисперсии по главным компонентам.

1. Покажите, что гауссовское ядро является таковым .
2. Какую проблему решает метод главных компонент в многомерной линейной регрессии? Записать матричную постановку задачи для метода главных компонент.
3. Для какого из этих методов обучения классификаторов изменение масштаба признаков (например, нормализация) может привести к изменению ответа, а для каких не может? Ответ обосновать. k-ближайших соседей со стандартной евклидовой метрикой; решающее дерево; логистическая регрессия без регуляризатора; логистическая регрессия с регуляризатором?
4. Найти главные направления и объясненные дисперсии по главным компонентам.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X1 | 4 | -1 | 3 | 3 | 2 | 1 |
| X2 | 4 | -1 | 0 | 2 | 0 | 2 |

1. Записать матричную постановку задачи для метода главных компонент. Какую проблему решает метод главных компонент в многомерной линейной регрессии?
2. Рассмотрим следующую функцию на пространстве вещественных чисел:

. Покажите, что она не является ядром.

1. Дана выборка:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 4 | 0 | -2 | 2 |
|  | 3 | 1 | -3 | -1 |

Найти главные направления и дисперсии по главным компонентам. Изобразить точки и главные направления

1. Задача снижения размерности. [Метод главных компонент](http://ru.learnmachinelearning.wikia.com/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%B3%D0%BB%D0%B0%D0%B2%D0%BD%D1%8B%D1%85_%D0%BA%D0%BE%D0%BC%D0%BF%D0%BE%D0%BD%D0%B5%D0%BD%D1%82_(Principal_component_analysis)) - 2 определения (через проекции и отклонения), их эквивалентность. Оценка качества аппроксимации отдельной компонентой и первыми *k* компонентами
2. Дана выборка:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 4 | 0 | -1 | 3 | 4 |
|  | 2 | -3 | -2 | 1 | 2 |
|  | 3 | 2 | 2 | 1 | -3 |

Постройте линейную дискриминантную функцию, квадратичную дискриминантную функцию.

***Пример контрольной работы***

ВАРАНТ 1.

1. Объясните, с какой целью в метрических классификаторах используются ядерные функции.
2. Дана обучающая выборка:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
|  | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |

С помощью наивного байесовского классификатора оценить вероятности ;

1. Дана обучающая выборка



Найти уравнение оптимальной разделяющей гиперплоскости, указать опорные точки

1. В каких алгоритмах классификации можно узнать не только классовую принадлежность классифицируемого объекта, но и вероятность того, что данный объект принадлежит каждому из классов?
2. Как вычислить норму вектора , зная лишь ядро ?

**7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине**

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2 «Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

**Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний**

***НПМ - Анализ больших данных и машинное обучение в экономике и финансах***

|  |  |  |
| --- | --- | --- |
| **Код компетенций** | **Наименование компетенций** | **Примеры заданий для оценки индикаторов достижения компетенций** |
| ДКН-1 | Способность проводить поиск источников данных, подготавливать данные для анализа, визуализировать данные | **1. Демонстрирует знание общих принципов визуального представления абстрактных понятий.**  **Задание 1.**  Постройте график квадратичной функции с использованием Matplotlib Python.  **2. Проводит поиск источников данных, подготавливает данные для анализа, визуализирует данные.**  **Задание 2.**  Напишите код программы на Python, загружающей значения акций Tesla и строящей их график с использованием библиотеки Seaborn.  **3. Владеет профессиональной терминологией в области инфографики, техникой создания 2D и 3D образов цифровых данных и их использования для информационной поддержки при принятии решений**  **Задание 3.**  Напишите код программы на языке Python для распознавания рукописных цифр с датасета MNIST |
| ДКН-2 | Способность применять методы регрессионного, факторного, кластерного и дискриминантного анализа, анализа нечисловой информации и теории сложных сетей к решению прикладных задач в экономике и финансах | **1.Владеет методами регрессионного, факторного, кластерного и дискриминантного анализа.**  **Задание 1.**  Опишите общий алгоритм построения и оценки модели регрессии.  **2.Анализирует нечисловую информацию и владеет методами теории сложных сетей**  **Задание 2.**  Сформулируйте и обоснуйте выбор основных методов решения задачи кластеризации данных  **3.Решает прикладные экономические и финансовые задачи методами регрессивного, факторного, кластерного и дискриминантного анализа, анализа нечисловой информации и теории сложных сетей.**  **Задание 3.**  Предложите метод решения задачи классификации эмоций человека из речевых сигналов. |
| ДКН-4 | Способность обосновывать и принимать решения с помощью технологий интеллектуального анализа данных и машинного обучения | **1.Владеет методикой принятия решений, основанной на технологиях интеллектуального анализа данных и машинного обучения.**  **Задание 1.**  Сформулируйте методику оптимального выбора гиперпараметров модели  **2.Обосновывает принимаемые решения с помощью технологий интеллектуального анализа данных и машинного обучения.**  **Задание 2.**  Обоснуйте метод оптимального выбора параметров регуляризации в условиях переобучения |

***НПМ - Технологии блокчейн и криптовалюты***

|  |  |  |
| --- | --- | --- |
| **Код компетенций** | **Наименование компетенций** | **Примеры заданий для оценки индикаторов достижения компетенций** |
| ДКН-1 | Способность применять технологии блокчейн и смарт- контракты к решению прикладных задач в экономике и финансах | **1. Решает вопрос о целесообразности применения технологии блокчейн для решения прикладной задачи.**  **Задание 1.**  Предложите метод решения задачи классификации смарт-контрактов  **2.Формулирует предложения по созданию новых подходов для решения имеющихся задач**  **Задание 2.**  Предложите метод кластеризации данных с несбалансированными классами  **3.Формулирует новые задачи, способствующие развитию отрасли, и предлагает пути их решения с использованием технологии блокчейн.**  **Задание 3.**  Построите P2P-модель энергетической торговой платформы для торговли энергией и использованием технологий блокчейн |
| ОПК-1 | Способен решать актуальные задачи фундаментальной и прикладной математики | **1.Демонстрирует знания в области фундаментальной и прикладной математики.**  **Задание 1.**  Назовите ключевые отличия классификации от кластеризации  **2. Системно выбирает математические методы для решения прикладных задач.**  **Задание 2.**  Перечислите основные методы минимизации функции потерь  **3. Осуществляет решение актуальных задач с использованием математических методов, в том числе методов обработки больших данных и машинного обучения.**  **Задание 3.**  Предложите метод восстановления многомерной зависимости в условиях неизвестной модели распределения |
| ОПК-3 | Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности | **1. Демонстрирует знания в области теории и методологии математического моделирования.**  **Задание 1.**  Сформулируйте основные положения имитационного моделирования  **2. Демонстрирует умение разрабатывать математические модели в области экономики и финансов с использованием методов обработки больших данных и машинного обучения.**  **Задание 2.**  Перечислите основные этапы построения модели финансового временного ряда  **3. Анализирует математические модели, осуществляя интерпретацию результатов моделирования и оценивая степень достоверности найденных решений.**  **Задание 3.**  Назовите основные метрики качества, применяемые в процессе построения математической модели |

***Примеры типовых заданий***

1. Ядерная функция имеет вид Как может выглядеть функция переводящие исходные векторы в расширенное пространство?
2. Рассмотрим задачу классификации с двумя классами 0 и 1. Пусть пространство признаков двумерное. Объекты каждого класса имеют нормальное распределение с математическим ожиданием и соответственно матрицей ковариации . Априорные вероятности классов равны и . Найти уравнение разделяющей поверхности байесовского классификатора. Перейти к новым координатам, оси которых совпадают с собственными векторами матрицы . Выписать уравнение разделяющей поверхности байесовского классификатора в новых координатах.
3. Рассмотрим двухмерную задачу классификации с двумя классами Y = { —1, +1}. В обучающей выборке 5 точек: {(1,1), (1,2), (2, 3)} класса +1 и {(3,1), (4, 2)} класса —1. Постройте оптимальную разделяющую гиперплоскость a(x) = sgn(w1x1+ w2x2 — *w*0). Какие из точек обучающей выборки получились опорными? Какова ширина получившейся разделяющей полосы?
4. Пять производственных объектов характеризуются двумя признаками: объемом продаж и среднегодовой стоимостью основных производственных фондов.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Объект | 1 | 2 | 3 | 4 | 5 |
| Объем продаж | 2 | 5 | 7 | 12 | 13 |
| Среднегодовая стоимость основных производственных фондов | 7 | 9 | 10 | 8 | 5 |

Провести классификацию этих объектов с помощью метода «ближайшего соседа».

1. Рассмотрим двухмерную задачу классификации с двумя классами Y = { —1, +1}. В обучающей выборке 5 точек: {(1,1), (1,2), (2, 3)} класса +1 и {(3,1), (4, 2)} класса —1. Постройте оптимальную разделяющую гиперплоскость a(x) = sign(*w1x1*+ *w2x2 —* *w*0). Какие из точек обучающей выборки получились опорными? Какова ширина получившейся разделяющей полосы?

***Теоретические вопросы для подготовки к зачету***

1.Что такое объект, признак, модель, функционал ошибки и обучение?

2.Линейные методы классификации и кластеризации: функционалы качества, методы настройки, особенности применения.

3.Сформулируйте метрики качества классификации.

4.Что такое кросс-валидация? На что влияет количество блоков в кросс-валидации?

5.Что такое точность, полнота?

6.Запишите выражение для линейной модели классификации. Что такое отступ? Как обучаются линейные классификаторы и для чего нужны верхние оценки пороговой функции потерь?

7.Что такое меры AUC-ROC? Опишите алгоритм построения ROC-кривой.

8.Запишите функционал логистической регрессии. Как он связан с методом максимума правдоподобия?

9.Запишите задачу метода опорных векторов для линейно неразделимого случая.

10.Опишите алгоритм K-средних. Каковы его ограничения в применении?

11.Опишите достоинства и недостатки ЕМ-алгоритма.

12.Опишите алгоритм DBSCAN и сформулируете его основные преимущества и недостатки.

13. Сравните результаты применения линейного дискриминантного анализа и наивного байесовского алгоритма для задачи разделения гауссовой смеси.

14.Опишите основные этапы факторного анализа.

15.Как связаны дисперсии главных компонент и собственные значения матрицы признаков?

16.Дендрограмма как графическое представление процесса кластеризации. Дендрограмма как метод кластеризации, различные условия остановки, основные этапы.

17.Иерархический кластерный анализ на основе матрицы расстояний. Обзор методов кластеризации. Условия остановки кластеризации.

18.Метод k-средних. Стартовые догадки для центров кластеров и зависимость результатов от их положения. Распределение данных по кластерам, корректировка центра.

19.Плотностные алгоритмы. DBSCAN.

20.Основные метрики качества кластеризации. Индекс Рэнда, силуэта.

***Теоретические вопросы для подготовки к экзамену***

1. Объясните, с какой целью в метрических классификаторах используются ядерные функции
2. Деревья решений. Построение и оценка качества. Основные метрики
3. Постановка задачи дискриминантного анализа для бинарной классификации
4. Какие объекты в SVM называются опорными?
5. Основные недостатки наивного байесовского классификатора.
6. Постановка задачи классификации. Типы классификаторов
7. Какую проблему решает метод главных компонент в многомерной линейной регрессии? Записать матричную постановку задачи для метода главных компонент.
8. Для какого из этих методов обучения классификаторов изменение масштаба признаков (например, нормализация) может привести к изменению ответа, а для каких не может? Ответ обосновать.
9. Перечислите известные вам метрики точности (качества) классификации
10. Прогноз в задаче классификации наивного байеса.
11. Что такое ядро в SVM? Зачем вводятся ядра? Любая ли функция может быть ядром?
12. Влияние параметра гауссовского ядра на вид разделяющей поверхности
13. Как определяется качество классификации с использованием дискриминантных функций?
14. Вывести формулу наивного байесовского классификатора для случая бинарных признаков (доказать, что он линеен).
15. Задачи классификации для категориальных данных. Приведите примеры.
16. Обобщённая модель линейного классификатора. Определение отступа. Минимизация эмпирического риска.
17. Основные шаги в построении главных компонент
18. Оценки параметров многомерного нормального распределения и их свойства.
19. Перечислите известные функции потерь в задаче классификации и кратко охарактеризуйте их.
20. Сформулируйте главные отличия наивного байеса от линейного дискриминантного анализа.
21. Геометрическая интерпретация метода главных компонент
22. Применение машины опорных векторов для многоклассовой классификации
23. Какие объекты в SVM называются опорными?
24. Кратко опишите байесовский подход к задаче классификации.
25. Перечислите основные методы отбора признаков по их важности, дайте их краткую характеристику. Приведите пример.
26. Кластерный анализ. Алгоритмы кластеризации. К-средних. Алгоритмы качества кластеризации.
27. Метрики качества в задачах классификации.
28. Обобщённая модель линейного классификатора. Определение отступа. Минимизация эмпирического риска.
29. Мультиколлинеарность. Основные характеристики, методы диагностики и устранения
30. Алгоритм t-SNE. Свойства. Недостатки
31. Как вычислить норму вектора ϕ(x), зная лишь ядро K(x, z)?
32. Метод линейного дискриминантного анализа. Его применение к задаче классификации. Ограничения метода.
33. Построение прогноза в задаче регрессии. Свойства оценок прогноза.
34. Сформулировать метод наивного байесовского классификатора для случая бинарных признаков.
35. Обобщённая модель линейного классификатора. Определение отступа. Минимизация эмпирического риска.
36. Задача снижения размерности. [Метод главных компонент](http://ru.learnmachinelearning.wikia.com/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%B3%D0%BB%D0%B0%D0%B2%D0%BD%D1%8B%D1%85_%D0%BA%D0%BE%D0%BC%D0%BF%D0%BE%D0%BD%D0%B5%D0%BD%D1%82_(Principal_component_analysis)). Оценка качества аппроксимации отдельной компонентой и первыми *k* компонентами
37. Сформулируйте главные отличия метода наивного байеса от линейного дискриминантного анализа.
38. Геометрическая интерпретация метода главных компонент

***Пример экзаменационного билета***

**ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**Департамент анализа данных, принятия решений и финансовых технологий**

***Дисциплина: «Прикладные модели и методы факторного, дискриминантного и кластерного анализа»***

*Направление подготовки: «Прикладная математика и информатика»*

*Направленность программы магистратуры: «Анализ больших данных и машинное обучение в экономике и финансах»*

Экзаменационный билет №

*Учебный 20\_\_\_/20\_\_\_ год \_\_\_\_\_\_\_модуль*

1. Объясните, с какой целью в метрических классификаторах используются ядерные функции. (10 баллов)
2. Дана обучающая выборка:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
|  | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |

С помощью наивного байесовского классификатора оценить вероятности ; (15 баллов)

1. Дана обучающая выборка



Найти уравнение оптимальной разделяющей гиперплоскости, указать опорные точки (15 баллов)

1. В каких алгоритмах классификации можно узнать не только классовую принадлежность классифицируемого объекта, но и вероятность того, что данный объект принадлежит каждому из классов? (10 баллов)
2. Как вычислить норму вектора , зная лишь ядро ? (10 баллов)

Подготовил Щетинин Е.Ю.

Утверждаю

Зам. Руководителя Департамента Зададаев С.А.

**11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.**

**11.1. Комплект лицензионного программного обеспечения:**

1. Windows, Microsoft Office.

2. Антивирус ESET Endpoint Security

**11.2. Современные профессиональные базы данных и информационные справочные системы:**

1. Информационно-правовая система «Гарант»

2. Информационно-правовая система «Консультант Плюс»

3. Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>

4.Cистема комплексного раскрытия информации «СКРИН» -http://www.skrin.ru

**11.3. Сертифицированные программные и аппаратные средства защиты информации**

– не предусмотрены

11.4. Microsoft Azure

11.5. Hadoop

11.6. HDInsight

11.7. Spark

11.8. Дистрибутив языка Python 3.4 (или более поздней версии) Anaconda 3

11.9. Программа Deductor Academic версии 5.3 или выше.