

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

**ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ
ОТДЕЛЕНИЕ ЭКОНОМИКИ
Кафедра Системного анализа и информатики**

УТВЕРЖДЕНА
на заседании кафедры
Системного анализа и информатики
Протокол от «1» сентября 2017г. № 1

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Индекс Б1.В.ОД.4 «Машинное обучение»

по направлению подготовки 38.04.01 «Экономика»

направленность «Системы больших данных в экономике»

квалификация магистр

очная форма обучения

Год набора - 2018

Москва, 2017г.

Автор(ы)—составитель(и):

к.т.н., доцент кафедры Системного анализа и информатики Стефановский Д.В.

Заведующий кафедрой

Системного анализа и информатики, к.т.н., доцент, Маруев С.А.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине соотнесенных с планируемыми результатами освоения образовательной программы.....	4
2. Объем и место дисциплины в структуре образовательной программы.....	5
3. Содержание и структура дисциплины.....	5
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине.....	7
5. Методические указания для обучающихся по освоению дисциплины	15
6. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет», учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине	18
6.1. Основная литература	18
6.2. Дополнительная литература	18
6.3. Учебно-методическое обеспечение самостоятельной работы	19
6.4. Нормативные правовые документы	19
6.5. Интернет-ресурсы	19
6.6. Иные источники	19
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы	19

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы

1.1. Дисциплина Б1.В.ОД.4 «Машинное обучение» обеспечивает овладение следующими компетенциями:

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПК-9	Способность анализировать и использовать различные источники информации для проведения экономических расчетов	ПК-9.	Способен анализировать и использовать различные источники информации для проведения экономических расчетов, используя методы машинного обучения.

1.2. В результате освоения дисциплины у студентов должны быть сформированы:

ОТФ/ТФ (при наличии профстандарта)	Код этапа освоения компетенции	Результаты обучения
Проведение аналитического исследования в соответствии с согласованными требованиями	ПК-9.2	<p>Знать:</p> <ul style="list-style-type: none"> Алгоритмы машинного обучения: обучение с учителем, обучение без учителя, полууправляемое обучение, обучение с одкреплением Определять теоретические верхние оценки переобученности: сложность, разделимость, устойчивость Решать проблемы переобучения и недообучения алгоритма Формировать предложения по использованию результатов анализа Машинное обучение: классификация, кластеризация, обнаружение выбросов, фильтрация Методы и модели классификации: логистическая регрессия, деревья решений, предредукция, постредукция, модели, основанные на правилах, наивный байесовский алгоритм, теорема Байеса, усиление энтропии информации Фильтрация шумовых выбросов, виды шумовых выбросов: глобальный, контекстуальный, коллективный Анализ изображений: тепловые карты, анализ сетей, анализ пространственных данных, анализ временных рядов Методы идентификации шаблонов Методы оценки моделей: оценка качества построенной модели по тестовой выборке и анализ обобщающих способностей алгоритма Распределенный анализ данных Анализ данных в реальном времени

		Уметь: • Осуществлять поиск информации о новых и перспективных методах анализа больших данных, сравнительный анализ методов
--	--	---

2. Объем и место дисциплины в структуре ОП ВО

Объем дисциплины

53Е, 64 ак. часа на контактную работу с преподавателем, 80 ак. часов на самостоятельную работу обучающихся;

Место дисциплины в структуре ОП ВО

- Б1.В.ОД.4 «Машинное обучение», 1 курс, 2 семестр;
- дисциплина реализуется после изучения дисциплин:
 - математический анализ (в объеме бакалавриата),
 - линейная алгебра (в объеме бакалавриата),
 - экономическая информатика (в объеме бакалавриата),
 - дискретная математика (в объеме бакалавриата),
 - теория вероятности и математическая статистика (в объеме бакалавриата),
 - алгоритмизация и программирование распределенных вычислений;
- форма промежуточной аттестации – экзамен.

3. Содержание и структура дисциплины

Таблица 2.

№ п/п	Наименование тем (разделов)	Объем дисциплины, ак.час./час						Форма текущего контроля успеваемости*, промежуточной аттестации	
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР		
			Л	ЛР	ПЗ	КСР			
Очная форма обучения									
Тема 1	Введение в машинное обучение. Цели и основная проблематика машинного обучения.	18	2	6				10	Опрос
Тема 2	Методы оценки точности полученных решений, включая ROC анализ.	18	2	6				10	Опрос
Тема 3	Современные регрессионные методы, включая эластичные сети, регрессионные деревья и леса. Стандартный метод наименьших квадратов. Методы распознавания.	18	2	6				10	Опрос
Тема 4	Байесовские методы и другие статистические модели, включая логистическую регрессию и др.	18	2	6				10	Опрос
Тема 5	Нейросетевые методы. Современные подходы	18	2	6				10	Опрос

	и идеи.							
Тема 6	Метод опорных векторов.	18	2	6			10	Опрос
Тема 7	Решающие деревья и леса.	18	2	6			10	Опрос
Тема 8	Комбинаторно-логические методы, АВО. Представление о графических моделях (Байесовские сети)	18	2	6			10	Опрос, ДЗ (1-8), КР (1-8)
Промежуточная аттестация				-			-	Экзамен
Всего:		180/ 135	16/ 12	48/ 36			80/ 60	36

Примечание – формы текущего контроля успеваемости: контрольная работа (КР), домашнее задание (ДЗ)

Содержание дисциплины

Тема 1. Введение в машинное обучение. Цели и основная проблематика машинного обучения.

Существующие, наборы данных, визуализация модели классификации. Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные. Типы задач: классификация, регрессия, прогнозирование, ранжирование.

Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль.

Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия.

Тема 2. Методы оценки точности полученных решений, включая ROC анализ.

Линейный регрессионный анализ, чувствительность, специфичность и точность. Корреляционный анализ. Анализ выживаемости и многомерная статистика. Таблицы дожития (mortality table) и метод Каплана-Мейера (Kaplan-Meier method). Лог-ранк тест. Модель Кокса.

Тема 3. Современные регрессионные методы, включая эластичные сети, регрессионные деревья и леса. Стандартный метод наименьших квадратов. Методы распознавания.

Логистическая регрессия. Автокорреляционная функция. Алгоритм Левенберга-Марквардта. Алгоритмы выбора линейных регрессионных моделей. Вспомогательные функции. Анализ регрессионных остатков. Аппроксимация Лапласа.

Регрессионные деревья и леса. Методы распознавания.

Тема 4. Байесовские методы и другие статистические модели, включая логистическую регрессию и др.

Понятие о случайном процессе. Байесовский подход к статистическому оцениванию. Априорные распределения, сопряженные с наблюдаемой генеральной совокупностью. Байесовский прогноз зависимой переменной, основанный на нормальной линейной модели множественной регрессии. Проверка статистических гипотез при байесовском подходе.

Тема 5. Нейросетевые методы. Современные подходы и идеи.

Биологический нейрон, модель МакКаллока-Питтса как линейный классификатор. Функции активации. Проблема полноты. Задача исключющего или. Полнота двухслойных сетей в пространстве булевых функций. Теоремы Колмогорова, Стоуна, Горбаня (без доказательства). Алгоритм обратного распространения ошибок. Эвристики: формирование начального приближения, ускорение сходимости, диагональный метод Левенберга-Марквардта. Проблема «паралича» сети. Метод послойной настройки сети. Подбор структуры сети: методы постепенного усложнения сети, оптимальное прореживание нейронных сетей (optimal brain damage). Нейронная сеть Кохонена. Конкурентное обучение, стратегии WTA и WTM.

Самоорганизующаяся карта Кохонена. Применение для визуального анализа данных. Искусство интерпретации карт Кохонена.

Тема 6. Метод опорных векторов.

Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin).

Случаи линейной разделимости и отсутствия линейной разделимости. Связь с минимизацией регуляризованного эмпирического риска. Кусочно-линейная функция потерь. Задача квадратичного программирования и двойственная задача. Понятие опорных векторов. Функция ядра (kernel functions), спрямляющее пространство, теорема Мерсера.

Способы конструктивного построения ядер. Примеры ядер.

SVM-регрессия.

Регуляризации для отбора признаков: LASSO SVM, Elastic Net SVM, SFM, RFM.

Метод релевантных векторов RVM.

Тема 7. Решающие деревья и леса.

Понятие логической закономерности.

Параметрические семейства закономерностей: конъюнкции пороговых правил, синдромные правила, шары, гиперплоскости.

Переборные алгоритмы синтеза конъюнкций: стохастический локальный поиск, стабилизация, редукция. Двухкритериальный отбор информативных закономерностей, парето-оптимальный фронт в (p, n) -пространстве. Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения. Вывод критериев ветвления. Мера нечистоты (impurity) распределения. Энтропийный критерий, критерий Джини. Редукция решающих деревьев: предредукция и постредукция. Алгоритм C4.5. Деревья регрессии. Алгоритм CART. Небрежные решающие деревья (oblivious decision tree). Решающий лес. Случайный лес (Random Forest).

Тема 8. Комбинаторно-логические методы, АВО. Представление о графических моделях (Байесовские сети)

Аппарат графических моделей (байесовские и марковские сети). Аппарат байесовского вывода. Некоторые методы дискретной оптимизации. Методы структурного обучения. Факторизация байесовских сетей. Потенциалы и энергия клика, связь с байесовскими сетями.

4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине

4.1. Формы и методы текущего контроля успеваемости.

4.1.1. В ходе реализации дисциплины Б1.В.ОД.4 «Машинное обучение» используются следующие методы текущего контроля успеваемости обучающихся: домашние задания и контрольные работы.

Тема	Методы текущего контроля успеваемости
Тема 1	Опрос
Тема 2	Опрос
Тема 3	Опрос
Тема 4	Опрос
Тема 5	Опрос
Тема 6	Опрос
Тема 7	Опрос
Тема 8	Опрос, домашнее задание по темам 1-8, контрольная работа по темам 1-8

4.1.2. Экзамен проводится с применением следующих методов (средств): в письменной форме в виде контрольной работы.

4.2. Материалы текущего контроля успеваемости обучающихся

Типовые оценочные материалы по теме 1

Опрос:

- Опишите существующие, наборы данных и модели классификации. Поясните на примере.
- Опишите постановку задач обучения по прецедентам. Объекты и признаки.
- Перечислите основные типы шкал: бинарные, номинальные, порядковые, количественные. Поясните на пример использования.
- Типы задач: классификация, регрессия, прогнозирование, ранжирование. Поясните на пример использования.
- Опишите следующие понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества. Поясните на примерах их использования.
- Опишите следующие понятия: принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль. Поясните на примерах их использования.
- Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия.

Типовые оценочные материалы по теме 2

Опрос:

- Линейный регрессионный анализ. Поясните на примерах суть и особенности.
- Опишите следующие понятия: чувствительность, специфичность и точность. Поясните на примерах их использования.
- Корреляционный анализ. Поясните на примерах суть и особенности.
- Анализ выживаемости и многомерная статистика. Поясните на примерах суть и особенности.
- Опишите способы использования таблицы дожития (mortality table) и применение метод Каплана-Мейера (Kaplan-Meier method). Поясните на примерах.
- Опишите Лог-ранк тест и модель Кокса. Поясните на примерах суть и особенности.

Типовые оценочные материалы по теме 3

Опрос:

- Опишите понятие "Логистической регрессии". Поясните на примерах суть и особенности.
- Опишите понятие "Автокорреляционная функция". Поясните на примерах суть и особенности.

- Опишите основные этапы и суть алгоритма Левенберга-Марквардта. Поясните на примере его использование.
- Опишите основные этапы и суть алгоритмов выбора линейных регрессионных моделей. Поясните на примере их использование.
- Опишите понятие "Регрессионные деревья и леса". Поясните на примерах суть и особенности.

Типовые оценочные материалы по теме 4

Опрос:

- Опишите понятие "случайный процесс". Поясните на примерах
- Опишите особенности и суть Байесовского подхода к статистическому оцениванию. Поясните на примерах.
- Опишите особенности и суть Априорных распределений, сопряженных с наблюдаемой генеральной совокупностью. Поясните на примере
- Опишите этапы и особенности построения Байесовского прогноз зависимой переменной, основанной на нормальной линейной модели множественной регрессии. Поясните на примере
- Опишите алгоритм проверки статистических гипотез при байесовском подходе. Поясните на примере

Типовые оценочные материалы по теме 5

Опрос:

- Опишите понятие Биологический нейрон. Поясните на примере особенности его использования.
- Опишите модель МакКаллока-Питтса как линейный классификатор.
- Приведите теоремы: Колмогорова, Стоуна, Горбаня (без доказательства). Поясните на примерах их использование
- Опишите алгоритм обратного распространения ошибок. Поясните на примере
- Приведите описания следующих эвристик: формирование начального приближения, ускорение сходимости, диагональный метод Левенберга-Марквардта. Поясните на примерах
- Опишите следующие подходы для подбора структуры сети: методы постепенного усложнения сети, оптимальное прореживание нейронных сетей (optimal brain damage). Поясните на примерах.
- Опишите нейронную сеть Кохонена. Поясните на примерах
- Опишите конкурентное обучение, стратегии WTA и WTM. Поясните на примерах.

Типовые оценочные материалы по теме 6

Опрос:

- Опишите понятия оптимальной разделяющей гиперплоскости и зазора между классами (margin). Поясните на примерах.
- Опишите связь с минимизацией регуляризованного эмпирического риска. Поясните на примере.
- Опишите суть задач квадратичного программирования и двойственной задачи. Поясните на примере.
- Опишите понятие опорных векторов. Поясните на примере

Типовые оценочные материалы по теме 7

Опрос:

- Понятие логической закономерности. Поясните на примере
- Опишите следующие параметрические семейства закономерностей: конъюнкции пороговых правил, синдромные правила, шары, гиперплоскости. Поясните на примерах
- Опишите следующие переборные алгоритмы синтеза конъюнкций: стохастический локальный поиск, стабилизация, редукция. Поясните на примере.

- Опишите понятие - Решающее дерево. Поясните на примере.
- Опишите алгоритм ID3. Поясните на примере.
- Перечислите недостатки жадной стратегии и способы их устранения. Поясните на примерах.
- Опишите алгоритм C4.5. Поясните на примерах.
- Опишите алгоритм CART. Поясните на примерах.
- Опишите алгоритм "Случайный лес (Random Forest)". Поясните на примерах.

Типовые оценочные материалы по теме 8

Опрос:

- Дайте понятие об аппарате графических моделей (байесовские и марковские сети). Поясните на примерах.
- Дайте понятие об аппарате Байесовского вывода. Поясните на примерах.
- Опишите основные методы структурного обучения. Поясните на примерах.

Домашнее задание.

При выполнении домашнего задания студентам предлагается разработать программу, решающую одну из задач использования одного из методов, предложенных на лекциях.

Студенты самостоятельно выбирают источники данных для разрабатываемой программы. Если преподаватель считает выбранные источники достаточными для успешного выполнения задания, то студент может приступить к выполнению задания. В противном случае студент должен устранить все замечания преподавателя и повторно представить тему на согласование.

После согласования темы, студентом проводится анализ с разработкой необходимых программ и алгоритмов.

Например, домашняя работа по теме *«Решающие деревья и леса»* представляется в письменной форме и содержит следующие разделы:

1. Для набора объектов производится расчет основных статистических показателей динамики.
2. Подбирается и, если необходимо, модифицируется алгоритм(ы) рассматриваемый(ые) в рамках темы.
3. Представляются в письменном виде результаты машинного обучения, модифицированные алгоритмы и программы, также указываются параметры качества параметров ошибок.

Контрольная работа:

Вариант 1.

Критерий AIC для модели регрессии построенной для некоторого набора объектов показывает насколько хорошо модель предсказывает целевой показатель. Для данного набора Выберите целевую переменную и подберите модель на основе критерия AIC (50 баллов из 100)

Вариант 2.

Выделить на числовой оси области значений показателя X с отнесением к классам $K1$ и $K2$. Показатель X в классах $K1$ и $K2$ распределён нормально с параметрами:

$K1$: математическое ожидание 0, стандартное отклонение 4;

$K2$: математическое ожидание 1, стандартное отклонение 1.

Выделить на числовой оси области значений показателя X с отнесением к классам $K1$ и $K2$ байесовским классификатором. Априорные вероятности классов $K1$ -0.6 и $K2$ -0.4.

Вариант 3.

Класс 1					Класс 1				
	X1	X2	X3	X4		X1	X2	X3	X4
Об. 1	1	1	1	0	Об. 1	0	1	0	0
Об. 2	0	0	1	1	Об. 2	0	1	1	0
Об. 3	1	0	0	1	Об. 3	1	0	0	0
Об. 4	1	0	1	1	Об. 4	0	1	1	0

Указать один тупиковый тест и один из представительных наборов

Задача 2. (50 баллов из 100) Ответьте на вопросы:

1. Что такое решающее дерево? Как по построенному дереву найти прогноз для объекта?
2. Зачем в вершинах нужны предикаты? Какие типы предикатов вы знаете? Приведите примеры.
3. Почему для любой выборки можно построить решающее дерево, имеющее нулевую ошибку на ней?
4. Почему не рекомендуется строить небинарные деревья (т.е. имеющие больше двух потомков у каждой вершины)?
5. Как задается критерий ошибки классификации? Критерий Джини? Энтропийный критерий? Какой у них смысл?

4.3. Оценочные средства для промежуточной аттестации.

4.3.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы. Показатели и критерии оценивания компетенций с учетом этапа их формирования.

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПК-9	Способность анализировать и использовать различные источники информации для проведения экономических расчетов	ПК-9.2	Способен анализировать и использовать различные источники информации для проведения экономических расчетов, используя методы машинного обучения.

Этап освоения компетенции	Показатель оценивания	Критерий оценивания
---------------------------	-----------------------	---------------------

ПК-9.2	Способен анализировать и использовать различные источники информации для проведения экономических расчетов, используя методы машинного обучения.	Понимает и может обосновать целесообразность выбранных источники информации для проведения экономических расчетов, используя методы машинного обучения.
--------	--	---

4.3.2. Типовые оценочные средства.

Вопросы к экзамену:

- Опишите существующие, наборы данных и модели классификации. Поясните на примере.
- Опишите постановку задач обучения по прецедентам. Объекты и признаки.
- Перечислите основные типы шкал: бинарные, номинальные, порядковые, количественные. Поясните на пример использования.
- Типы задач: классификация, регрессия, прогнозирование, ранжирование. Поясните на пример использования.
- Опишите следующие понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества. Поясните на примерах их использования.
- Опишите следующие понятия: принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль. Поясните на примерах их использования.
- Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия.
- Линейный регрессионный анализ. Поясните на примерах суть и особенности.
- Опишите следующие понятия: чувствительность, специфичность и точность. Поясните на примерах их использования.
- Корреляционный анализ. Поясните на примерах суть и особенности.
- Анализ выживаемости и многомерная статистика. Поясните на примерах суть и особенности.
- Опишите способы использования таблицы дожития (mortality table) и применение метод Каплана-Мейера (Kaplan-Meier method). Поясните на примерах.
- Опишите Лог-ранк тест и модель Кокса. Поясните на примерах суть и особенности.
- Опишите понятие "Логистической регрессии". Поясните на примерах суть и особенности.
- Опишите понятие "Автокорреляционная функция". Поясните на примерах суть и особенности.
- Опишите основные этапы и суть алгоритма Левенберга-Марквардта. Поясните на примере его использования.
- Опишите основные этапы и суть алгоритмов выбора линейных регрессионных моделей. Поясните на примере их использования.
- Опишите понятие "Регрессионные деревья и леса". Поясните на примерах суть и особенности.
- Опишите понятие "случайный процесс". Поясните на примерах
- Опишите особенности и суть Байесовского подхода к статистическому оцениванию. Поясните на примерах.
- Опишите особенности и суть Априорных распределений, сопряженных с наблюдаемой генеральной совокупностью. Поясните на примере

- Опишите этапы и особенности построения Байесовского прогноз зависимой переменной, основанной на нормальной линейной модели множественной регрессии. Поясните на примере
- Опишите алгоритм проверки статистических гипотез при байесовском подходе. Поясните на примере
- Опишите понятие Биологический нейрон. Поясните на примере особенности его использования.
- Опишите модель МакКаллока-Питтса как линейный классификатор.
- Приведите теоремы: Колмогорова, Стоуна, Горбаня (без доказательства). Поясните на примерах их использование
- Опишите алгоритм обратного распространения ошибок. Поясните на примере
- Приведите описания следующих эвристик: формирование начального приближения, ускорение сходимости, диагональный метод Левенберга-Марквардта. Поясните на примерах
- Опишите следующие подходы для подбора структуры сети: методы постепенного усложнения сети, оптимальное прореживание нейронных сетей (optimal brain damage). Поясните на примерах.
- Опишите нейронную сеть Кохонена. Поясните на примерах
- Опишите конкурентное обучение, стратегии WTA и WTM. Поясните на примерах.
- Опишите понятия оптимальной разделяющей гиперплоскости и зазора между классами (margin). Поясните на примерах.
- Опишите связь с минимизацией регуляризованного эмпирического риска. Поясните на примере.
- Опишите суть задач квадратичного программирования и двойственной задачи. Поясните на примере.
- Опишите понятие опорных векторов. Поясните на примере
- Понятие логической закономерности. Поясните на примере
- Опишите следующие параметрические семейства закономерностей: конъюнкции пороговых правил, синдромные правила, шары, гиперплоскости. Поясните на примерах
- Опишите следующие переборные алгоритмы синтеза конъюнкций: стохастический локальный поиск, стабилизация, редукция. Поясните на примере.
- Опишите понятие - Решающее дерево. Поясните на примере.
- Опишите алгоритм ID3. Поясните на примере.
- Перечислите недостатки жадной стратегии и способы их устранения. Поясните на примерах.
- Опишите алгоритм C4.5. Поясните на примерах.
- Опишите алгоритм CART. Поясните на примерах.
- Опишите алгоритм "Случайный лес (Random Forest)". Поясните на примерах.
- Дайте понятие об аппарате графических моделей (байесовские и марковские сети). Поясните на примерах.
- Дайте понятие об аппарате Байесовского вывода. Поясните на примерах.
- Опишите основные методы структурного обучения. Поясните на примерах.

Шкала оценивания.

Оценка определяется по формуле:

$$\frac{1}{4} \text{ опрос} + \frac{1}{4} \text{ д.з.} + \frac{1}{4} \text{ к.р.} + \frac{1}{4} \text{ сдача зачета.}$$

10- бальная шкала	Традиционн ая шкала	«Зачтено»/ «Не зачтено»	Определение
10	Отлично	Зачтено	Полные, глубокие и систематические знания, полный и правильный ответ на теоретический вопрос, полное и правильное решение задачи.
9	Отлично	Зачтено	Глубокие и систематические знания, правильный ответ на теоретический вопрос, правильное решение задачи.
8	Отлично	Зачтено	Систематические знания, правильный ответ на теоретический вопрос, правильное решение задачи.
7	Хорошо	Зачтено	Систематические знания, правильный ответ на теоретический вопрос с незначительными неточностями, правильное решение задачи.
6	Хорошо	Зачтено	Систематические знания, правильный ответ на теоретический вопрос с незначительными неточностями, правильное решение задачи с незначительными неточностями.
5	Удовлетвори тельно	Зачтено	Ответ на теоретический вопрос неполный, правильное решение задачи с незначительными неточностями.
4	Удовлетвори тельно	Зачтено	Ответ на теоретический вопрос неполный, решение задачи содержит арифметические ошибки, не влияющие на правильность хода решения задачи.
3	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос неполный, решение задачи содержит идеологические ошибки.
2	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос неверный и/или решение задачи содержит идеологические ошибки.
1	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос неверный и решение задачи отсутствует.
0	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос отсутствует и решение задачи отсутствует.

4.4. Методические материалы по проведению промежуточной аттестации

Зачет проводится в аудитории. Отсчет времени, отведенного на письменную работу, идет по завершении процедуры размещения студентов и раздачи заданий.

Студент обязан являться на письменный контроль в указанное в расписании время. В случае опоздания время, отведенное на письменный контроль знаний, не продлевается.

При себе студенты могут иметь только письменные принадлежности. Необходимую для выполнения работы бумагу выдает преподаватель.

Преподаватель раздает варианты работы, содержащий 2 вопроса. Листы с заданиями должны быть повернуты текстом вниз, чтобы студенты до окончания процедуры раздачи не могли начать выполнение работы. По окончании раздачи вариантов студентам разрешается перевернуть текст задания и одновременно приступить к выполнению работы. По окончании отведенного времени студенты одновременно заканчивают выполнение работы. Если работа завершена существенно раньше срока, то по разрешению преподавателя студент может покинуть аудиторию досрочно.

Мобильные телефоны должны быть выключены и убраны со столов, допускается использование калькуляторов, выполняющих только простые арифметические вычисления.

Во время проведения письменного контроля знаний студентам не разрешается пользоваться учебными программами, справочниками и прочими источниками информации.

Использование материалов, а также попытка общения с другими студентами или иными лицами, в том числе с применением электронных средств связи, несанкционированные перемещения и т.п. являются основанием для удаления студента из аудитории и последующего проставления в ведомость оценки «неудовлетворительно».

Во время проведения письменного контроля знаний студентам разрешается покинуть аудиторию только при условии сдачи работы в объеме, выполненном к моменту выхода из аудитории. Дальнейшее продолжение работы запрещается.

Ответы в работе без объяснений не засчитываются. Рисунки должны быть четкими, все линии графиков, используемых при ответах на вопросы задач, должны быть подписаны.

Продолжительность экзаменационной письменной работы 120 минут.

5. Методические указания для обучающихся по освоению дисциплины

Любой вид занятий, создающий условия для зарождения самостоятельной мысли, познавательной и творческой активности студента связан с самостоятельной работой. В широком смысле под самостоятельной работой понимают совокупность всей самостоятельной деятельности студентов как в учебной аудитории, так и вне ее, в контакте с преподавателем и в его отсутствие. Самостоятельная работа может реализовываться: непосредственно в процессе аудиторных занятий – на лекциях, практических и семинарских занятиях, при выполнении контрольных и лабораторных работ и др.; в контакте с преподавателем вне рамок аудиторных занятий – на консультациях по учебным вопросам, в ходе творческих контактов, при ликвидации задолженностей, при выполнении индивидуальных заданий и т.д.; в библиотеке, дома, в общежитии, на кафедре и других местах при выполнении студентом учебных и творческих заданий.

Лекции

Главное в период подготовки к лекционным занятиям – научиться методам самостоятельного умственного труда, сознательно развивать свои творческие способности и овладевать навыками творческой работы. Для этого необходимо строго соблюдать дисциплину учебы и поведения. Четкое планирование своего рабочего времени и отдыха является необходимым условием для успешной самостоятельной работы. В основу его нужно положить рабочие программы изучаемых в семестре дисциплин. Каждому студенту следует составлять еженедельный и семестровый планы работы, а также план на каждый рабочий день. С вечера всегда надо распределять работу на завтрашний день. В конце каждого дня целесообразно подводить итог работы: тщательно проверить, все ли выполнено по намеченному плану, не было ли каких-либо отступлений, а если были, по какой причине это произошло. Нужно осуществлять самоконтроль, который является необходимым условием успешной учебы. Если что-то осталось невыполненным, необходимо изыскать время для завершения этой части работы, не уменьшая объема недельного плана.

Самостоятельная работа на лекции. Слушание и запись лекций – сложный вид вузовской аудиторной работы. Внимательное слушание и конспектирование лекций предполагает интенсивную умственную деятельность студента. Краткие записи лекций, их конспектирование помогает усвоить учебный материал. Конспект является полезным

тогда, когда записано самое существенное, основное и сделано это самим студентом. Не надо стремиться записать дословно всю лекцию. Такое «конспектирование» приносит больше вреда, чем пользы. Запись лекций рекомендуется вести по возможности собственными формулировками. Желательно запись осуществлять на одной странице, а следующую оставлять для проработки учебного материала самостоятельно в домашних условиях. Конспект лекции лучше подразделять на пункты, параграфы, соблюдая красную строку. Этому в большой степени будут способствовать пункты плана лекции, предложенные преподавателям. Принципиальные места, определения, формулы и другое следует сопровождать замечаниями «важно», «особо важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек. Лучше если они будут собственными, чтобы не приходилось присить их у однокурсников и тем самым не отвлекать их во время лекции. Целесообразно разработать собственную «маркографию» (значки, символы), сокращения слов. Не лишним будет и изучение основ стенографии. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть знаниями.

Семинар и проведение опроса

Каждый студент должен начать с ознакомления с планом занятия, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованную к данной теме. На основе индивидуальных предпочтений студенту необходимо самостоятельно выбрать тему доклада по проблеме семинара и по возможности подготовить по нему презентацию. Если программой дисциплины предусмотрено выполнение практического задания, то его необходимо выполнить с учетом предложенной инструкции (устно или письменно). Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности студента свободно ответить на теоретические вопросы семинара, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильном выполнении практических заданий и контрольных работ.

Работа с литературными источниками.

В процессе подготовки к семинарским занятиям, студентам необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также научной и популярной) литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у студентов свое отношение к конкретной проблеме. Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем по каждой теме семинарского или практического занятия, что позволяет студентам проявить свою индивидуальность в рамках выступления на данных занятиях, выявить широкий спектр мнений по изучаемой проблеме.

Методические указания по выполнению самостоятельной работы:

Тема 1. Введение в машинное обучение. Цели и основная проблематика машинного обучения.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Существующие наборы данных, визуализация модели классификации. Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные. Типы задач: классификация, регрессия, прогнозирование, ранжирование.

Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль.

Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия.

Тема 2. Методы оценки точности полученных решений, включая ROC анализ.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Линейный регрессионный анализ, чувствительность, специфичность и точность. Корреляционный анализ. Анализ выживаемости и многомерная статистика. Таблицы дожития (mortality table) и метод Каплана-Мейера (Kaplan-Meier method). Лог-ранк тест. Модель Кокса.

Тема 3. Современные регрессионные методы, включая эластичные сети, регрессионные деревья и леса. Стандартный метод наименьших квадратов. Методы распознавания.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Логистическая регрессия. Автокорреляционная функция. Алгоритм Левенберга-Марквардта. Алгоритмы выбора линейных регрессионных моделей. Вспомогательные функции. Анализ регрессионных остатков. Аппроксимация Лапласа.

Регрессионные деревья и леса. Методы распознавания.

Тема 4. Байесовские методы и другие статистические модели, включая логистическую регрессию и др.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Понятие о случайном процессе. Байесовский подход к статистическому оцениванию. Априорные распределения, сопряженные с наблюдаемой генеральной совокупностью. Байесовский прогноз зависимой переменной, основанный на нормальной линейной модели множественной регрессии. Проверка статистических гипотез при байесовском подходе.

Тема 5. Нейросетевые методы. Современные подходы и идеи.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Биологический нейрон, модель МакКаллока-Питтса как линейный классификатор. Функции активации. Проблема полноты. Задача исключающего или. Полнота двухслойных сетей в пространстве булевых функций. Теоремы Колмогорова, Стоуна, Горбаня (без доказательства). Алгоритм обратного распространения ошибок. Эвристики: формирование начального приближения, ускорение сходимости, диагональный метод Левенберга-Марквардта. Проблема «паралича» сети. Метод послойной настройки сети. Подбор структуры сети: методы постепенного усложнения сети, оптимальное прореживание нейронных сетей (optimal brain damage). Нейронная сеть Кохонена. Конкурентное обучение, стратегии WTA и WTM.

Самоорганизующаяся карта Кохонена. Применение для визуального анализа данных. Искусство интерпретации карт Кохонена.

Тема 6. Метод опорных векторов.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin).

Случаи линейной разделимости и отсутствия линейной разделимости. Связь с минимизацией регуляризованного эмпирического риска. Кусочно-линейная функция потерь. Задача квадратичного программирования и двойственная задача. Понятие опорных векторов. Функция ядра (kernel functions), спрямляющее пространство, теорема Мерсера.

Способы конструктивного построения ядер. Примеры ядер.

SVM-регрессия.

Регуляризации для отбора признаков: LASSO SVM, Elastic Net SVM, SFM, RFM.

Метод релевантных векторов RVM.

Тема 7. Решающие деревья и леса.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Понятие логической закономерности.

Параметрические семейства закономерностей: конъюнкции пороговых правил, синдромные правила, шары, гиперплоскости.

Переборные алгоритмы синтеза конъюнкций: стохастический локальный поиск, стабилизация, редукция. Двухкритериальный отбор информативных закономерностей, парето-оптимальный фронт в (p, n) -пространстве. Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения. Вывод критериев ветвления. Мера нечистоты (impurity) распределения. Энтропийный критерий, критерий Джини. Редукция решающих деревьев: предредукция и постредукция. Алгоритм C4.5. Деревья регрессии. Алгоритм CART. Небрежные решающие деревья (oblivious decision tree). Решающий лес. Случайный лес (Random Forest).

Тема 8. Комбинаторно-логические методы, АВО. Представление о графических моделях (Байесовские сети)

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Аппарат графических моделей (байесовские и марковские сети). Аппарат байесовского вывода. Некоторые методы дискретной оптимизации. Методы структурного обучения. Факторизация байесовских сетей. Потенциалы и энергия клика, связь с байесовскими сетями.

6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

6.1. Основная литература.

1. Бринк Хенрик, Джозеф Ричардс, Феверолф Марк, Машинное обучение. – СПб.: Питер, 2017. -336 с.: ил. – (Серия «Библиотека программиста»). ISBN 978-5-496-02989-6
2. Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. 2016. 302 с.

6.2. Дополнительная литература.

1. Уэс Маккинли Python и анализ данных / Пер. с англ. Слинкин А. А. – М.: ДМК Пресс, 2015. – 482 с.: ил."
2. Райан Митчелл: Скрапинг веб-сайтов с помощью Python. Сбор данных из современного интернета
3. Spark для профессионалов: современные паттерны обработки больших данных. Риза С., Лезерсон У., Оуэн Ш., Уиллс Д.

6.3. Учебно-методическое обеспечение самостоятельной работы.

Отдельное обеспечение не предусмотрено.

6.4. Нормативные правовые документы.

Не предусмотрены.

6.5. Интернет-ресурсы.

1. <http://citeseer.ist.psu.edu/> основной источник знаний по Computer Science, по многим статьям есть полные тексты
2. <http://citeseer.cs.msu.su/>— русскоязычная электронная библиотека научных статей
3. <http://arxiv.org/> — библиотека электронных публикаций, в основном по физике, но доля «Computer Science» в последнее время стремительно увеличивается
4. <http://rexa.info/> библиографическая поисковая система по статьям, авторам и грантам
5. <http://elibrary.ru/> - российская научная электронная библиотека
6. <http://iinwww.ira.uka.de/bibliography/index.html>
7. библиографическая база данных для работы с BibTeX
8. <http://www.gotai.net/> -- русскоязычный сайт об искусственном интеллекте
9. Math-Net.ru -- общероссийский математический портал

6.6. Иные источники.

Не предусмотрены.

7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Для лекций:

1. Персональный компьютер
2. Мультимедийный проектор
3. Доска, мел или маркеры

Для лабораторных занятий:

1. Компьютерный класс,
2. Виртуальная машина Ubuntu 15.04 b выше с установленным Postgresql и MongoDB
3. Мультимедийный проектор
4. Доска, маркеры
5. Компилятор R-2.15.1 – GNU - <http://www.r-project.org/> либо интегрированная среда разработки RStudio – GNU AGP - <http://www.rstudio.com/ide/>.

6. Jupyter Notebook - бесплатная интерактивная оболочка для языка программирования Python, позволяющая объединить код, текст и диаграммы.
7. Компилятор Scala – <http://www.scala-lang.org/>
8. Программный комплекс анализа новостного сайта - "Crawler-Persona"
9. База данных "Централизация государственных закупок в 2014 г".
10. База данных учебно-методических материалов по дисциплине "Макроэкономика".
11. База данных Бюджетная и социально-экономическая статистика субъектов Российской Федерации.