

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Институт геологии и нефтегазовых технологий



УТВЕРЖДАЮ

Проректор по образовательной деятельности КФУ

Проф. Талорский Д.А.

_____ 20__ г.

подписано электронно-цифровой подписью

Программа дисциплины
Машинные методы обучения Б1.В.ДВ.4

Направление подготовки: 05.04.01 - Геология

Профиль подготовки: Перспективные геоинформационные технологии в геологии и геофизике

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Чернова И.Ю.

Рецензент(ы):

Нугманов И.И.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Нургалиев Д. К.

Протокол заседания кафедры No ____ от " ____ " _____ 201__ г

Учебно-методическая комиссия Института геологии и нефтегазовых технологий:

Протокол заседания УМК No ____ от " ____ " _____ 201__ г

Регистрационный No 337416

Казань
2016

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) доцент, к.н. (доцент) Чернова И.Ю. кафедра геофизики и геоинформационных технологий Институт геологии и нефтегазовых технологий, Inna.Chernova@kpfu.ru

1. Цели освоения дисциплины

Целями освоения дисциплины (модуля) Машинные методы обучения являются знакомство с теоретическими основами машинного обучения, и приобретение практических навыков их применения для решения прикладных задач.

2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " Б1.В.ДВ.4 Дисциплины (модули)" основной образовательной программы 05.04.01 Геология и относится к дисциплинам по выбору. Осваивается на 1 курсе, 2 семестр.

М2.ДВ3. Для изучения дисциплины "Машинные методы обучения" необходимо освоить курсы математики, информатики, теории вероятностей и математической статистики в объеме общеобразовательного и профессионального циклов бакалавриата.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОПК-1 (профессиональные компетенции)	способностью самостоятельно приобретать, осмысливать, структурировать и использовать в профессиональной деятельности новые знания и умения, развивать свои инновационные способности
ОПК-2 (профессиональные компетенции)	способностью самостоятельно формулировать цели исследований, устанавливать последовательность решения профессиональных задач
ОПК-4 (профессиональные компетенции)	способностью профессионально выбирать и творчески использовать современное научное и техническое оборудование для решения научных и практических задач
ОПК-5 (профессиональные компетенции)	способностью критически анализировать, представлять, защищать, обсуждать и распространять результаты своей профессиональной деятельности
ПК-1 (профессиональные компетенции)	способностью формировать диагностические решения профессиональных задач путем интеграции фундаментальных разделов геологических наук и специализированных знаний, полученных при освоении программы магистратуры
ОК-8 (общекультурные компетенции)	готовностью к проектированию комплексных научно-исследовательских и научно-производственных работ при решении профессиональных задач

В результате освоения дисциплины студент:

1. должен знать:

математические модели теории обучения машин,
основные алгоритмы машинного обучения.

2. должен уметь:

применять методы машинного обучения для решения практических задач.

3. должен владеть:

навыками выполнения машинного обучения в среде статистической системы R.

Курс направлен на приобретение практических навыков применения машинных методов обучения для решения прикладных задач.

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 2 зачетных(ые) единиц(ы) 72 часа(ов).

Форма промежуточного контроля дисциплины зачет во 2 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Задачи обучения по прецедентам (примерам).	2	1-2	1	0	4	
2.	Тема 2. Байесовские методы обучения.	2	2-4	1	0	4	
3.	Тема 3. Метрические методы обучения.	2	5-7	2	0	4	
4.	Тема 4. Линейные методы обучения.	2	8-10	2	0	4	
5.	Тема 5. Методы восстановления зависимостей.	2	11-13	2	0	4	
	Тема . Итоговая форма контроля	2		0	0	0	зачет
	Итого			8	0	20	

4.2 Содержание дисциплины

Тема 1. Задачи обучения по прецедентам (примерам).

лекционное занятие (1 часа(ов)):

Основные понятия и определения. Объекты и признаки, степень измеримости, кодирование (контрасты).

лабораторная работа (4 часа(ов)):

Основные понятия и определения. Объекты и признаки, степень измеримости, кодирование (контрасты).

Тема 2. Байесовские методы обучения.

лекционное занятие (1 часа(ов)):

Вероятностная постановка задачи классификации: функционал сред-него риска, оптимальное байесов-ское решающее правило.

лабораторная работа (4 часа(ов)):

Вероятностная постановка задачи классификации: функционал сред-него риска, оптимальное байесов-ское решающее правило.

Тема 3. Метрические методы обучения.

лекционное занятие (2 часа(ов)):

Метод ближайшего соседа и его обобщения: обобщённый метриче-ский классификатор, метод бли-жайших соседей

лабораторная работа (4 часа(ов)):

Метод ближайшего соседа и его обобщения: обобщённый метриче-ский классификатор, метод бли-жайших соседей

Тема 4. Линейные методы обучения.

лекционное занятие (2 часа(ов)):

Аппроксимация и регуляризация эмпирического риска. Линейная модель классификации.

лабораторная работа (4 часа(ов)):

Аппроксимация и регуляризация эмпирического риска. Линейная модель классификации.

Тема 5. Методы восстановления зависимостей.

лекционное занятие (2 часа(ов)):

Метод наименьших квадратов. Непараметрическая регрессия: ядерное сглаживание: формула Надарая?Ватсона, выбор ядра и ширины окна, проблема выбросов: робастная непараметрическая ре-грессия

лабораторная работа (4 часа(ов)):

Метод наименьших квадратов. Непараметрическая регрессия: ядерное сглаживание: формула Надарая?Ватсона, выбор ядра и ширины окна, проблема выбросов: робастная непараметрическая ре-грессия

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Задачи обучения по прецедентам (примерам).	2	1-2	Изучение литературы, выполнение упражнений из учебника	8	Обсуждение
2.	Тема 2. Байесовские методы обучения.	2	2-4	Изучение литературы, выполнение упражнений из учебника	8	Обсуждение

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
3.	Тема 3. Метрические методы обучения.	2	5-7	Изучение литературы, выполнение упражнений из учебника	8	Обсуждение
4.	Тема 4. Линейные методы обучения.	2	8-10	Изучение литературы, выполнение упражнений из учебника	8	Обсуждение
5.	Тема 5. Методы восстановления зависимостей.	2	11-13	Изучение литературы, выполнение упражнений из учебника	12	Обсуждение
	Итого				44	

5. Образовательные технологии, включая интерактивные формы обучения

Проводятся лекции и лабораторные занятия с использованием компьютеров с применением специализированного программного обеспечения. Часть материала изучается самостоятельно.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

Тема 1. Задачи обучения по прецедентам (примерам).

Обсуждение , примерные вопросы:

1. Записать общую формулу байесовского классификатора. 2. Какие вы знаете три подхода к восстановлению плотности распределения по выборке? 3. Что такое наивный байесовский классификатор? 4. Что такое оценка плотности Парзена-Розенблатта.

Тема 2. Байесовские методы обучения.

Обсуждение , примерные вопросы:

5. На что влияет ширина окна, а на что вид ядра в методе парзеновского окна? 6. Многомерное нормальное распределение. Формула квадратичного дискриминанта. При каком условии он становится линейным? 7. На каких предположениях основан линейный дискриминант Фишера? 8. Что такое "проблема мультиколлинеарности", в каких задачах и при использовании каких алгоритмов она возникает? Какие есть подходы к её решению?

Тема 3. Метрические методы обучения.

Обсуждение , примерные вопросы:

9. Что такое "смесь распределений"? 10. Что такое EM-алгоритм, какова его основная идея? Какая задача решается на E-шаге, на M-шаге? Каков вероятностный смысл скрытых переменных?

Тема 4. Линейные методы обучения.

Обсуждение , примерные вопросы:

11. Что такое стохастический EM-алгоритм, какова основная идея? В чём его преимущество (какой недостаток стандартного EM-алгоритма он устраняет)? 12. Что такое сеть радиальных базисных функций? 13. Что такое "выбросы"? Как осуществляется фильтрация выбросов? 14. Что такое метод потенциальных функций? Идея алгоритма настройки. Сравните с методом радиальных базисных функций. 15. Зачем нужен отбор опорных объектов в метрических алгоритмах классификации? 16. Метод стохастического градиента. 17. Обоснование логистической регрессии, основные посылки и следствия. Как выражается апостериорная вероятность классов. 18. Две мотивации и постановка задачи метода опорных векторов. Постановка задачи SVM. 19. Что такое ядро в SVM? Зачем вводятся ядра? Любая ли функция может быть ядром? 20. Что такое ROC-кривая, как она определяется? Как она эффективно вычисляется? 21. В каких алгоритмах классификации можно узнать не только классовую принадлежность классифицируемого объекта, но и вероятность того, что данный объект принадлежит каждому из классов? 22. Каков вероятностный смысл регуляризации? Какие типы регуляризаторов Вы знаете?

Тема 5. Методы восстановления зависимостей.

Обсуждение, примерные вопросы:

23. Что есть общего между ядром в непараметрической регрессии и ядром SVM? 24. На что влияет ширина окна, а на что вид ядра в непараметрической регрессии? 25. Постановка задачи многомерной линейной регрессии. 26. Что такое сингулярное разложение? Как оно используется для решения задачи наименьших квадратов? 27. Что такое "проблема мультиколлинеарности" в задачах многомерной линейной регрессии? Какие есть три подхода к её устранению? 28. Сравнить гребневую регрессию и лассо. В каких задачах предпочтительнее использовать лассо? 29. Какую проблему решает метод главных компонент в многомерной линейной регрессии?

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

Максимальный суммарный балл по результатам выполнения контрольных работ - 15.

Оценка активности студентов во время лабораторных занятий - до 35 баллов.

Максимальный балл на зачете - 50 .

Вопросы на зачет:

1. Записать общую формулу байесовского классификатора.
2. Какие вы знаете три подхода к восстановлению плотности распределения по выборке?
3. Что такое наивный байесовский классификатор?
4. Что такое оценка плотности Парзена-Розенблатта.
5. На что влияет ширина окна, а на что вид ядра в методе парзеновского окна?
6. Многомерное нормальное распределение. Формула квадратичного дискриминанта. При каком условии он становится линейным?
7. На каких предположениях основан линейный дискриминант Фишера?
8. Что такое "проблема мультиколлинеарности", в каких задачах и при использовании каких алгоритмов она возникает? Какие есть подходы к её решению?
9. Что такое "смесь распределений"?
10. Что такое EM-алгоритм, какова его основная идея? Какая задача решается на E-шаге, на M-шаге? Каков вероятностный смысл скрытых переменных?
11. Что такое стохастический EM-алгоритм, какова основная идея? В чём его преимущество (какой недостаток стандартного EM-алгоритма он устраняет)?
12. Что такое сеть радиальных базисных функций?
13. Что такое "выбросы"? Как осуществляется фильтрация выбросов?
14. Что такое метод потенциальных функций? Идея алгоритма настройки. Сравните с методом радиальных базисных функций.
15. Зачем нужен отбор опорных объектов в метрических алгоритмах классификации?
16. Метод стохастического градиента.

17. Обоснование логистической регрессии, основные посылки и следствия. Как выражается апостериорная вероятность классов.
18. Две мотивации и постановка задачи метода опорных векторов. Постановка задачи SVM.
19. Что такое ядро в SVM? Зачем вводятся ядра? Любая ли функция может быть ядром?
20. Что такое ROC-кривая, как она определяется? Как она эффективно вычисляется?
21. В каких алгоритмах классификации можно узнать не только классовую принадлежность классифицируемого объекта, но и вероятность того, что данный объект принадлежит каждому из классов?
22. Каков вероятностный смысл регуляризации? Какие типы регуляризаторов Вы знаете?
23. Что есть общего между ядром в непараметрической регрессии и ядром SVM?
24. На что влияет ширина окна, а на что вид ядра в непараметрической регрессии?
25. Постановка задачи многомерной линейной регрессии.
26. Что такое сингулярное разложение? Как оно используется для решения задачи наименьших квадратов?
27. Что такое "проблема мультиколлинеарности" в задачах многомерной линейной регрессии? Какие есть три подхода к её устранению?
28. Сравнить гребневую регрессию и лассо. В каких задачах предпочтительнее использовать лассо?
29. Какую проблему решает метод главных компонент в многомерной линейной регрессии?
30. Метод настройки с возвращениями (backfitting): постановка задачи и основная идея метода.
31. Какие методы построения логистической регрессии Вы знаете?

7.1. Основная литература:

- Калиткин, Н. Н. Численные методы: учеб. пособие / Н. Н. Калиткин. ? 2-е изд., исправленное. ? СПб.: БХВ-Петербург, 2011. ? 586 с. URL: <http://znanium.com/bookread.php?book=350803>
- Комаров, Е. И. Методологический инструментарий современного социального управления [Электронный ресурс] : Учебно-практическое пособие / Е. И. Комаров; Под ред. профессора Е. И. Комарова. - М.: Издательско-торговая корпорация "Дашков и К-", 2013. - 316 с. URL: <http://znanium.com/bookread.php?book=414987>
- Гафурова, Н. В. Методика обучения информационным технологиям. Практиум [Электронный ресурс] : учеб. пособие / Н. В. Гафурова, Е. Ю. Чурилова. - Красноярск: Сибирский федеральный университет, 2011. - 181 с. URL: <http://znanium.com/bookread.php?book=441409>

7.2. Дополнительная литература:

- Павлюковец М.А. Пантюхова П.В. Использование синквейна и ПОПС-формулы как интерактивных методов обучения английскому языку при формировании учебно-познавательной компетенции у студентов-лингвистов / Интернет-журнал "Науковедение", Вып. 1, 2014 <http://znanium.com/bookread.php?book=475484>
- Студент вуза: технологии обучения и профессиональной карьеры.: Учебное пособие / Под ред. С.Д. Резника - 3-е изд., перераб. и доп. - М.: НИЦ Инфра-М, 2013. - 509 с. URL: <http://znanium.com/bookread.php?book=373095>
- Методы менеджмента качества. Методология управления риском стандартизации / П.С. Серенков, В.Л. Гуревич и др. - М.: НИЦ ИНФРА-М; Мн.: Нов. знание, 2014 - 256 с.: ил.; 60x90 1/16. - (Высшее образование: Магистр.). (п) ISBN 978-5-16-009427-4, 450 экз. <http://znanium.com/bookread2.php?book=440747>

7.3. Интернет-ресурсы:

Департамент машинного обучения CMU - <http://www.ml.cmu.edu/>

Журнал "Машинное обучение", MIT - <http://jmlr.csail.mit.edu/>

Курс машинного обучения в Стэнфордском университете - <http://cs229.stanford.edu/>

Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. - <http://www.machinelearning.ru/>

Статистическая система R - <http://cran.r-project.org/>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Машинные методы обучения" предполагает использование следующего материально-технического обеспечения:

Компьютерный класс.

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 05.04.01 "Геология" и магистерской программе Перспективные геоинформационные технологии в геологии и геофизике .

Автор(ы):

Чернова И.Ю. _____

"__" _____ 201__ г.

Рецензент(ы):

Нугманов И.И. _____

"__" _____ 201__ г.