

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. Минзарипов Р.Г.

_____ 20__ г.

Программа дисциплины
Машинные методы обучения М2.ДВ.3

Направление подготовки: 020700.68 - Геология

Профиль подготовки: Перспективные геоинформационные технологии в геологии и геофизике

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Савельев А.А.

Рецензент(ы):

Мухарамова С.С.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Зарипов Ш. Х.

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

Учебно-методическая комиссия Института геологии и нефтегазовых технологий:

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

Регистрационный No

Казань
2014

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) профессор, д.н. (профессор) Савельев А.А. кафедра моделирования экологических систем отделение экологии , Anatoly.Saveliev@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 " М2.ДВ.3 Профессиональный" основной образовательной программы 020700.68 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
ОК-1 (общекультурные компетенции)	ОК-1
ОК-6 (общекультурные компетенции)	ОК-6
ОК-7 (общекультурные компетенции)	ОК-7
ОК-8 (общекультурные компетенции)	ОК-8
ОК-10 (общекультурные компетенции)	ОК-10

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 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
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

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
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:

К. В. Воронцов. Математические методы обучения по прецедентам (теория обучения машин). 2012.-141с.

URL: <file:///M:/work/PLAN/Methodichki/MachineLearning/voron/voron-ML/index.php.htm>

Статистическая система R. R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Самоучитель по экспертным системам

<http://www.computerbooks.ru/books/Programming/Book-Introduction-to-Expert-Systems/index.html> (экз.)

Курс лекций по машинному обучению <http://www.cs.nyu.edu/~roweis/csc2515/> (экз.)

Документация по экспертным системам JFS, ARTGallery, C4_5. (экз.)

Курс лекций по машинному обучению http://logic.pdmi.ras.ru/_sergey/ (экз.)

Курс лекций по машинному обучению

[http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_\(курс_лекций%2C_К.В.Воронцов\)](http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_(курс_лекций%2C_К.В.Воронцов)) (экз.)

7.2. Дополнительная литература:

Новак В. Математические принципы нечеткой логики / Новак Вилем, Перфильева Ирина, Мочкорж Иржи; пер. с англ. под ред. А.Н. Аверкина. Москва: Физматлит, 2006. 347 с.; 22. Загл. и авт. ориг.: Mathematical principles of fuzzy logic/ Vilém Novák, Irina Perfilieva, Jiří Močkoř. Библиогр.: с. 335-343 (143 назв.). Предм. указ.: с. 345-347. (1 экз.)

Нечеткие гибридные системы: теория и практика / [Батыршин И.З., Недосекин А.О., Стецко А.А. и др.]; под ред. Н.Г. Ярушкиной. Москва: Физматлит, 2007. 207 с.: ил., табл.; 22. (Информационные и компьютерные технологии). Авт. указаны в вып. дан.. Библиогр.: с. 181-207. ISBN 978-5-9221-0786-0, 500. (1 экз.)

Степанов Р.Г. Технология Data Mining: Интеллектуальный анализ данных: учебное пособие / Р. Г. Степанов; Казан. гос. ун-т. Казань: Казанский государственный университет, 2009. 110 с.: ил.; 21. Библиогр.: с. 110 (7 назв.), 100. (48 экз.)

Хайкин С. Нейронные сети: полный курс / С. Хайкин; [пер. с англ. д.т.н. Н.Н. Куссуль, к.т.н. А.Ю. Шелестова]. Издание 2-е, испр.. Москва [и др.]: Вильямс, 2008. 1103 с.: ил.; 24. Загл. и авт. ориг.: Neural Networks / Simon Haykin. пред. изд. нет. Библиогр.: с. 996-1069 (1183 назв.). Предм. указ.: с. 1070-1103. ISBN 978-5-8459-0890-2, 1000. (5 экз.)

Барский А.Б. Нейронные сети: распознавание, управление, принятие решений / А.Б. Барский. Москва: Финансы и статистика, 2004. 174, [1] с.: ил.; 21. (Прикладные информационные технологии). Библиогр.: с. 170-173. ISBN 5-279-02757-X, 3000. (1 экз.)

Осовский С. Нейронные сети для обработки информации / С. Осовский; пер. с пол. И. Д. Рудинского. Москва: Финансы и статистика, 2004. 343 с.: ил.; 24. Пер. изд.: Sieci neuronowe do przetwarzania informacji / Stanislaw Osowski. - Warszawa, 2000. Предм. указ.: с. 340-343. Библиогр.: с. 330-339. ISBN 5-279-02567-4((в обл.)). (1 экз.)

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 согласно утвержденному учебному плану

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

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 020700.68 "Геология" и магистерской программе Перспективные геоинформационные технологии в геологии и геофизике .

Автор(ы):

Савельев А.А. _____

"__" _____ 201__ г.

Рецензент(ы):

Мухарамова С.С. _____

"__" _____ 201__ г.