

Санкт-Петербургский государственный университет

РАБОЧАЯ ПРОГРАММА

учебной дисциплины

Машинное обучение

Machine Learning

Язык(и) обучения

русский

Трудоёмкость (границы трудоёмкости) в зачетных единицах: 3

Регистрационный номер рабочей программы: 056748

Санкт-Петербург

2018

Раздел 1. Характеристики учебных занятий

1.1. Цели и задачи учебных занятий

Основной целью освоения дисциплины «Машинное обучение» является формирование у обучающихся устойчивого понимания основных понятий и базовых алгоритмов машинного обучения, приобретение практических навыков работы с инструментами, применяемыми в области машинного обучения.

Поставленная цель достигается путем решения следующих задач курса:

- 1) ознакомить студентов с основными задачами, решаемыми в области машинного обучения, базовыми алгоритмами машинного обучения, областями применения алгоритмов;
- 2) способствовать развитию практических навыков работы с инструментами, применяемыми в области машинного обучения;
- 3) ознакомить с основными тенденциями развития подходов в области машинного обучения.

1.2. Требования к подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)

Освоение дисциплин: дискретная математика, теория вероятностей, линейная алгебра, основы программирования.

Знание одного из высокоуровневых языков программирования (Java, Python).

1.3. Перечень результатов обучения (learning outcomes)

После изучения курса обучающиеся должны

знать: статистические (байесовские), метрические, логические, линейные методы классификации; методы регрессионного анализа; нейросетевые и композиционные методы классификации и регрессии; методы кластеризации;

уметь: осуществлять выбор между различными методами с учетом решаемой задачи; решать задачи обучения с учителем, обучения без учителя, задачи с частичным обучением и задачи обучения с подкреплением; использовать существующие инструментальные средства для решения задач машинного обучения;

владеть: средствами и приемами решения задач классификации, регрессии, кластеризации; навыками работы с существующими инструментальными средствами для решения задач машинного обучения.

1.4. Перечень активных и интерактивных форм учебных занятий

При изложении части тем применяется мультимедиа-проектор для проведения презентаций. В процессе преподавания данной дисциплины используются как классические методы обучения (семинары), так и различные виды самостоятельной работы студентов по заданию преподавателя.

В рамках данного курса используются такие *активные формы работы*, как:

- подготовка к семинарским занятиям;

- подготовка доклада по предложенной преподавателем теме;
- построение интеллект-карты по курсу;
- выполнение индивидуальных заданий по реализации изученных алгоритмов;

интерактивные формы обучения:

- метод групповой работы при решении задач во время практических занятий;
- экспертная оценка другими обучающимися результатов решения индивидуальных заданий и коллективное обсуждение полученных результатов;
- экспертная оценка другими обучающимися построенных интеллект-карт.

Общий объем активных и интерактивных форм учебных занятий составляет 32 часа.

Раздел 2. Организация, структура и содержание учебных занятий

2.1. Организация учебных занятий

2.1.1. Профиль Технологии баз данных

Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся																		
Период обучения (модуль)	Контактная работа обучающихся с преподавателем											Самостоятельная работа				Объём активных и интерактивных форм учебных занятий	Трудоёмкость	
	лекции	семинары	консультации	практические занятия	лабораторные работы	контрольные работы	коллоквиумы	текущий контроль	промежуточная аттестация	итоговая аттестация	под руководством преподавателя	в присутствии преподавателя	сам.раб. с использованием методических материалов	текущий контроль (сам.раб.)	промежуточная аттестация (сам.раб.)			итоговая аттестация (сам.раб.)
ОСНОВНАЯ ТРАЕКТОРИЯ																		
очная форма обучения																		
Семестр 3		32		32					2				35		7		32	3
ИТОГО		32		32					2				35		7		32	3

Формы текущего контроля успеваемости, виды промежуточной и итоговой аттестации			
Период обучения (модуль)	Формы текущего контроля успеваемости	Виды промежуточной аттестации	Виды итоговой аттестации (только для программ итоговой аттестации и дополнительных образовательных программ)
ОСНОВНАЯ ТРАЕКТОРИЯ			
очная форма обучения			
Семестр 3		зачет	

2.2. Структура и содержание учебных занятий

Период обучения (модуль): **Семестр 3**

№ п/п	Наименование темы (раздела, части)	Вид учебных занятий	Количество часов
1	Основные методы классификации и регрессии.	семинары	10
		практические занятия	12
		по методическим материалам	7
2	Работа с данными. Другие задачи машинного обучения.	семинары	12
		практические занятия	10
		по методическим материалам	14
3	Нейросетевые и композиционные методы. Графические модели.	семинары	10
		практические занятия	10
		по методическим материалам	14

Содержание учебных занятий**Раздел 1: Основные методы классификации и регрессии.***Семинары:*

- 1.1 Введение в машинное обучение. Основные понятия и примеры решаемых прикладных задач. Оценка качества классификации. Обучающая, проверочная и тестовая выборки, принципы разделения. Метод перекрестного контроля.
- 1.2 Метрические методы классификации. Метод ближайших соседей и его обобщения. Проклятие размерности.
- 1.3 Статистические методы классификации. Байесовские методы классификации. Непараметрические и параметрические оценки плотности.
- 1.4 Линейные методы классификации. Логистическая регрессия. Метод опорных векторов. Нелинейные обобщения SVM, идея Kernel Trick.
- 1.5 Линейные методы регрессии. Принципы регуляризации. Непараметрическая регрессия. Многомерная линейная регрессия и метод главных компонент.
- 1.6 Логические методы классификации и регрессии. Построение деревьев решений, алгоритмы ID3, C4.5, CART.

Практические занятия:

- 1.1 Обзор существующих инструментальных средств разработчика для решения задач машинного обучения: Octave, PyBrain, Weka, R.
- 1.2 Практическое занятие по методу ближайших соседей: распознавание рукописных цифр.
- 1.3 Практическое занятие по использованию байесовского классификатора: задача классификации текстовых документов.
- 1.4 Практическое занятие по линейным методам классификации: задача о кредите.
- 1.5 Практическое занятие по линейным методам регрессии: задача о пожарах.

Раздел 2: Работа с данными. Другие задачи машинного обучения.*Семинары:*

- 2.1 Основная схема работы с данными. Этапы подготовки данных.
- 2.2 Работа с пропущенными значениями и метод максимизации ожидания.

- 2.3 Обучение без учителя: кластеризация. Графовые, статистические и иерархические методы решения задачи.
- 2.4 Обучение без учителя: поиск ассоциативных правил. Алгоритм APriori. Алгоритм построения префиксного FP-дерева. Алгоритм выделения частых наборов по FP-дереву.
- 2.5 Задача частичного обучения. Эвристические методы. Порождающие модели. Разделение по низкой плотности. Методы на основе графов. Решение с предварительным обучением без учителя.
- 2.6 Задача обучения с подкреплением. Задача о многоруком бандите. Методы «действие-ценность». Конечные марковские процессы принятия решений. Методы динамического программирования. Методы Монте-Карло. Методы временных различий.
- 2.7 Задача коллаборативной фильтрации. Корреляционные модели. Латентные модели.
- 2.8 Построение различных моделей, настройка параметров и выбор модели. Оценка качества для различных задач. Проблема переобучения.

Практические занятия:

- 2.1 Практическое занятие по кластеризации: определение спама.
- 2.2 Практическое занятие по выделению ассоциативных правил: супермаркет.
- 2.3 Практическое занятие по частичному обучению: определение пола клиента.

Раздел 3: Нейросетевые и композиционные методы. Графические модели.

Семинары:

- 3.1 Нейросетевые методы классификации и регрессии. Многослойные нейронные сети. Различные виды нейронных сетей. Алгоритм обратного распространения.
- 3.2 Нейронные сети: современные тенденции. Ограниченная машина Больцмана. Сверточная нейронная сеть. Long-Short Term Memory.
- 3.3 Композиционные методы классификации и регрессии. Линейные композиции, бустинг. Бэггинг и метод случайных подпространств. Смеси алгоритмов.
- 3.4 Байесовские сети доверия. Обучение сетей. Основы выбора оптимальной структуры: детерминированные и недетерминированные алгоритмы. Инструменты для работы с байесовскими сетями.
- 3.5 Задача статистического анализа последовательностей. Вероятностные модели последовательностей. Скрытые марковские модели.
- 3.6 Задачи статистического анализ изображений. Марковские и условные случайные поля. Подход статистической физики. Подход теории вероятностей.

Практические занятия:

- 3.1 Практическое занятие по нейронным сетям: прогнозирование болезни Паркинсона.
- 3.2 Практическое занятие по композиционным методам: задача о депозите.
- 3.3 Практическое занятие по марковским моделям: задача о погоде.

Раздел 3. Обеспечение учебных занятий

3.1. Методическое обеспечение

3.1.1. Методические указания по освоению дисциплины

Самостоятельная работа студентов включает в себя самостоятельную проработку рассмотренных на аудиторных занятиях материалов; выполнение практических заданий по изучаемым разделам; построение интеллект-карты по курсу, выступление с презентацией перед группой; изучение рекомендованной литературы и ресурсов в сети интернет. Время и место самостоятельной работы выбираются студентами по своему усмотрению с учетом рекомендаций преподавателя.

Подготовка к семинару заключается в следующем:

- внимательно изучить материал предыдущего семинара;
- целесообразно составить краткий конспект или схему, отображающую смысл и связи основных понятий данного раздела, включенных в него тем, а затем, полезно изучить выдержки из литературы;
- узнать тему предстоящего семинара (по тематическому плану, по материалам, размещенным в системе дистанционного обучения Blackboard);
- ознакомиться с учебным материалом по учебным пособиям;
- записать возможные вопросы, которые вы зададите преподавателю в ходе аудиторного занятия.

Изучение дисциплины заканчивается зачетом. При непосредственной подготовке к зачету рекомендуется тщательно разобрать все основные определения и алгоритмы. Отметить пробелы в знаниях, которые следует ликвидировать в ходе консультации.

3.1.2. Методическое обеспечение самостоятельной работы

Самостоятельная работа студентов в рамках дисциплины «Машинное обучение» является важным компонентом обучения. Данной программой предусмотрены виды деятельности студента, которые направляются и корректируются преподавателем, и виды учебной деятельности, которые осуществляются студентом самостоятельно в рамках плана изучения данной учебной дисциплины.

К группе видов и форм самостоятельной работы студентов относятся:

- подготовка к семинарским занятиям;
- подготовка доклада по предложенной преподавателем теме;
- построение интеллект-карты по курсу;
- выполнение индивидуальных заданий по реализации изученных алгоритмов.

Для организации самостоятельной работы студентов рекомендуется предоставить презентации с изучаемым материалом, проводить консультации во время аудиторных занятий.

3.1.3. Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания

По результатам освоения дисциплины студентам предлагается выполнить доклад по предложенной преподавателем теме или построить интеллект-карту по изученному материалу.

Промежуточная аттестация проходит в форме собеседования по пройденному материалу, с учетом результатов выполнения индивидуальных заданий, доклада или интеллект-карты.

Оценка «зачтено» ставится при условии успешного выполнения не менее 60% практических заданий, а также выполнения доклада или интеллект-карты по курсу, знания основных определений и алгоритмов.

Оценка «зачтено» также может быть поставлена при отсутствии результатов практических заданий в случае ответа на два теоретических вопроса.

Преподаватель имеет право предоставить информацию о задолженностях студента в аттестационную комиссию.

3.1.4. Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)

Примерный перечень тем докладов:

1. Самоорганизующиеся карты Кохонена и их применение к задачам машинного обучения.
2. Генетические алгоритмы и их применение к задачам машинного обучения.
3. Text Mining. Особенности машинного обучения при работе с текстовыми коллекциями.
4. Современные алгоритмы кластеризации: DBScan и Affinity Propagation.
5. Нечеткие деревья решений. Сравнение с классическими (Модификация на основе C4.5).

Примерный список обязательных компонентов интеллект-карты:

1. Данные, особенности работы.
2. Задачи машинного обучения.
3. Цикл решения задачи (выборки, оценка качества, критерии выбора модели).
4. Методы машинного обучения.
5. Литература.
6. Творческие элементы.

Для оценки интеллект-карт применяется трансфертный лист:

Я считаю, что... (общая оценка)	Особенно удачным является... (достоинства, интересные моменты)
В то же время, я бы посоветовал... (пожелания)	Не кажется ли Вам, что... (вопросы)

Примерный перечень теоретических вопросов по курсу:

1. Основные понятия и задачи машинного обучения. Постановка задачи обучения с учителем. Примеры решаемых прикладных задач.
2. Основные понятия и задачи машинного обучения. Постановка задачи обучения без учителя. Примеры решаемых прикладных задач.

3. Основные понятия и задачи машинного обучения. Постановка задачи частичного обучения. Примеры решаемых прикладных задач.
4. Основные понятия и задачи машинного обучения. Постановка задачи обучения с подкреплением. Примеры решаемых прикладных задач.
5. Основные понятия и задачи машинного обучения. Постановка задачи коллаборативной фильтрации. Примеры решаемых прикладных задач.
6. Основная схема работы с данными. Этапы подготовки данных.
7. Работа с пропущенными значениями. Метод максимизации ожидания.
8. Оценка качества классификации и кластеризации. Обучающая, проверочная и тестовая выборки, принципы разделения. Метод перекрестного контроля.
9. Проблема переобучения. Принцип минимизации среднего риска. Принцип минимизации эмпирического риска.
10. Построение различных моделей обучения, настройка параметров и критерии выбора модели.
11. Метрические методы классификации. Метод ближайших соседей и его обобщения.
12. Статистические методы классификации. Наивный байесовский классификатор.
13. Линейные методы классификации. Логистическая регрессия.
14. Линейные методы классификации. Метод опорных векторов. Нелинейные обобщения метода опорных векторов.
15. Задача восстановления регрессии. Случай линейно разделимой выборки. Принципы регуляризации.
16. Задача восстановления регрессии. Многомерная линейная регрессия. Метод главных компонент.
17. Логические методы классификации. Построение деревьев решений, алгоритм ID3.
18. Логические методы классификации. Построение деревьев решений, алгоритм C4.5.
19. Нейросетевые методы классификации и регрессии. Многослойные нейронные сети и их виды.
20. Обучение нейронных сетей. Алгоритм обратного распространения ошибки.
21. Обучение нейронных сетей. Градиентный спуск и его вариации.
22. Нейронные сети: современные тенденции. Ограниченная машина Больцмана.
23. Нейронные сети: современные тенденции. Сверточная нейронная сеть.
24. Нейронные сети: современные тенденции. Long-Short Term Memory.
25. Композиционные методы классификации. Бустинг и алгоритм AdaBoost.
26. Стохастические методы построения композиций: бэггинг и метод случайных подпространств.
27. Обучение без учителя: кластеризация. Графовые методы решения задачи.
28. Обучение без учителя: кластеризация. Статистические методы решения задачи.
29. Обучение без учителя: кластеризация. Иерархические методы решения задачи.
30. Обучение без учителя: поиск ассоциативных правил. Алгоритм APriori.
31. Задача частичного обучения. Эвристические методы: метод самообучения, метод co-training, метод co-learning, метод активного обучения.
32. Задача частичного обучения. Решение с помощью порождающих моделей.
33. Задача частичного обучения. Low-Density Separation. T3SVM.
34. Задача частичного обучения. Методы на основе графов. Решение с предварительным обучением без учителя.
35. Задача обучения с подкреплением. Задача о многоруком бандите. Жадная стратегия. ϵ -жадная стратегия. Стратегия softmax.

36. Задача обучения с подкреплением. Фундаментальные методы решения: метод итераций, методы временных различий (TD), метод SARSA, метод Q-обучения.
37. Задача обучения с подкреплением. Многошаговые обобщения методов: TD(λ), SARSA(λ), Q(λ).
38. Задача коллаборативной фильтрации. Корреляционные модели.
39. Задача коллаборативной фильтрации. Латентные модели.
40. Байесовские сети доверия. Понятие условной независимости. D-разделение. Обучение сетей.
41. Байесовские сети доверия. Основы выбора оптимальной структуры: детерминированные и недетерминированные алгоритмы.
42. Задача статистического анализа последовательностей. Скрытые марковские модели. Вычисление вероятностей и алгоритм Forward-Backward.
43. Задача статистического анализа последовательностей. Скрытые марковские модели и их применение. Алгоритм Витерби.
44. Задача статистического анализа последовательностей. Обучение скрытых марковских моделей. Обобщения.
45. Задачи статистического анализа изображений. Марковские и условные случайные поля: подход статистической физики.
46. Задачи статистического анализа изображений. Марковские и условные случайные поля: подход теории вероятностей.

3.1.5. Методические материалы для оценки обучающимися содержания и качества учебного процесса

Для оценки обучающимися содержания и качества учебного процесса используется анкета-отзыв установленная локальными актами СПбГУ.

3.2. Кадровое обеспечение

3.2.1. Образование и (или) квалификация преподавателей и иных лиц, допущенных к проведению учебных занятий

К чтению лекций должны привлекаться преподаватели, имеющие ученую степень и/или ученое звание, опыт планирования и организации учебного процесса, или специалисты в этой области.

3.2.2. Обеспечение учебно-вспомогательным и (или) иным персоналом

Для технического обеспечения учебного процесса необходима возможность прибегать к помощи специалистов, ответственных за надлежащее функционирование компьютеров и программного обеспечения, а также за своевременное поддержание в рабочем состоянии другой используемой техники.

3.3. Материально-техническое обеспечение

3.3.1. Характеристики аудиторий (помещений, мест) для проведения занятий

Аудитории и помещения, предназначенные для проведения занятий по данной дисциплине должны отвечать санитарным нормам, предусмотренным Образовательным стандартом реализации программ высшего профессионального образования Санкт-Петербургского государственного университета.

В аудиториях требуется наличие компьютеризированных рабочих мест для проведения совместных практических работ и демонстрации материалов курса.

3.3.2. Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования

Для реализации программы необходим доступ преподавателей к офисной технике (персональный компьютер, копировальный аппарат, принтер), а также достаточное количество расходных материалов к ней, выделенных для использования в учебном процессе.

Минимально необходимый для реализации курса перечень материально-технического обеспечения включает: мультимедийный проектор для презентаций и демонстраций, компьютеры для проведения практических работ.

3.3.3. Характеристики специализированного оборудования

Нет специальных требований.

3.3.4. Характеристики специализированного программного обеспечения

При практической работе каждый обучающийся во время занятий и самостоятельной подготовки должен быть обеспечен рабочим местом в компьютерном классе с выходом в Интернет.

Необходим доступ к инструментам и библиотекам для разработки:

Eclipse или NetBeans с установленной Java / MS Visual Studio 2015 Ultimate Edition / Python; Matlab или Octave; PyBrain; Weka; R и RStudio.

3.3.5. Перечень и объёмы требуемых расходных материалов

Фломастеры цветные, губки, бумага формата A4, канцелярские товары, картриджи принтеров, диски, флеш-накопители и др. в объёме, необходимом для организации и проведения занятий, по заявкам преподавателей, подаваемым в установленные сроки

3.4. Информационное обеспечение

3.4.1. Список обязательной литературы

1. C. Sammut, G.I. Webb (Ed.) "Encyclopedia of Machine Learning" (2011)
(электронный доступ через библиотеку университета).
2. Hastie, T., Tibshirani, R., Friedman, J. "Elements of Statistical Learning: Data Mining, Inference and Prediction" (2nd ed., 2009)
(электронный доступ через библиотеку университета).
3. Galushkin, A. I. "Neural Networks Theory" (2007)
(электронный доступ через библиотеку университета).

3.4.2. Список дополнительной литературы

1. Christmann, A., Steinwart, I. "Support Vector Machines" (2008)
(электронный доступ через библиотеку университета).
2. Sivanandam, S.N., Deepa, S.N. "Introduction to Genetic Algorithms" (2008)
(электронный доступ через библиотеку университета).

3. Cios, K.J., Swiniarski, R.W., Pedrycz, W., Kurgan L.A. “Data Mining: A Knowledge Discovery Approach” (2007)
(электронный доступ через библиотеку университета).
4. Bramer, M. “Principles of Data Mining” (2nd ed., 2013)
(электронный доступ через библиотеку университета).

3.4.3. Перечень иных информационных источников

1. Сайт проекта PyBrain: <http://www.pybrain.org>
2. Машинное обучение, курс лекций: <http://www.machinelearning.ru/>
3. ШАД Yandex: http://shad.yandex.ru/lectures/machine_learning.xml
4. Официальный сайт Weka: <http://www.cs.waikato.ac.nz/ml/weka/>
5. Официальный сайт Encog: <http://www.heatonresearch.com/encog>
6. Сайт компании BaseGroup: <http://www.basegroup.ru>
7. Официальный сайт FANN: <http://leenissen.dk/fann/wp/>

Раздел 4. Разработчики программы

Фамилия, имя, отчество	Учёная степень	Учёное звание	Должность	Контактная информация (служебный адрес электронной почты, служебный телефон)
Романенко Елена Станиславовна	--	--	ст. преп.	e.s.romanenko@spbu.ru