

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО
Директор центра

А.С. Микуленков

	Рабочая программа дисциплины (модуля)
по дисциплине:	Машинное обучение
по направлению:	Прикладные математика и физика
профиль подготовки:	Физика перспективных технологий: альтернативная энергетика, научное программирование и функциональные материалы Физтех-школа Электроники, Фотоники и Молекулярной Физики кафедра инновационной фармацевтики, медицинской техники и биотехнологии
курс:	1
квалификация:	магистр

Семестры, формы промежуточной аттестации:

- 1 (осенний) - Дифференцированный зачет
- 2 (весенний) - Экзамен

Аудиторных часов: 180 всего, в том числе:

- лекции: 30 час.
- семинары: 30 час.
- лабораторные занятия: 120 час.

Самостоятельная работа: 105 час.

Подготовка к экзамену: 30 час.

Всего часов: 315, всего зач. ед.: 7

Программу составил: Л.А. Урошлев, канд. физ.-мат. наук

Программа обсуждена на заседании кафедры инновационной фармацевтики, медицинской техники и биотехнологии 04.06.2021

Аннотация

Курс знакомит студентов с современным состоянием машинного обучения и искусственного интеллекта: от классических алгоритмов до глубокого обучения и последних достижений в области искусственного интеллекта. В результате студенты формируют устойчивую теоретическую базу и практические навыки для дальнейшего развития в области ИИ.

Методы машинного обучения нашли применение во многих областях и оказали значительное влияние на различные предметные области. В данном курсе рассматриваются основные результаты, достигнутые в обработке естественного языка, обучении с подкреплением и компьютерном зрении. В ходе анализа подходов, позволивших добиться значимых успехов в различных областях, обнаруживается, что они сильно взаимосвязаны и, более того, зачастую могут дополнить друг друга. В данном курсе рассматриваются как теоретические основы существующих подходов к различным задачам, так и их техническая реализация и перспективы для дальнейшего развития.

1. Цели и задачи

Цель дисциплины

сформировать теоретические и практические знания в области обучения машин, современных методов восстановления зависимостей по эмпирическим данным, включая дискриминантный, кластерный и регрессионный анализ, частичное обучение.

Задачи дисциплины

правильно формулировать задачу в терминах машинного обучения, овладеть навыками практического решения задач интеллектуального анализа данных.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные научные знания в области физико-математических наук
	ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности
	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач профессиональной деятельности
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.2 Способен применять знания в области физико-математических наук для решения поставленной задачи, формулирования выводов и оценки полученных результатов
	ОПК-4.3 Способен аргументировано выбирать способ проведения научного исследования
ПК-3 Способен профессионально работать с исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

основные принципы и проблематику теории обучения машин, основные современные методы обучения по прецедентам — классификации, кластеризации и регрессии.

уметь:

формализовать постановки прикладных задач анализа данных, использовать современные методы обучения по прецедентам для решения практических задач, оценивать точность и эффективность полученных решений.

владеть:

основными понятиями теории машинного обучения.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Методы обработки естественного языка		10	20	15
2	Введение в обучение с подкреплением		10	20	15
3	Глубокое обучение в компьютерном зрении		10	20	15
4	Введение в машинное обучение. Метрические алгоритмы, оценка качества моделей	5		10	10
5	Линейные модели	5		10	10
6	Деревья и ансамбли моделей	5		10	10
7	Работа с признаками. Ограничения машинного обучения	5		10	10
8	Введение в глубокое обучение	5		10	10
9	Обучение без учителя	5		10	10
Итого часов		30	30	120	105
Подготовка к экзамену		30 час.			
Общая трудоёмкость		315 час., 7 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 1 (Осенний)

1. Методы обработки естественного языка

Представление текстов в векторной форме. Классические подходы и подходы, основанные на глубоком обучении. Word2vec, GloVe. Латентное (скрытое) представление последовательностей с помощью рекуррентных нейронных сетей.

Проблема затухающего и градиента (Vanishing gradient). Проблема всплеска амплитуды градиента (Exploding gradient). Сверточные сети в анализе текстов.

Машинный перевод. Исторический экскурс. Статистический машинный перевод. Оценка качества перевода. Нейронные сети в машинном переводе. Понятие кодировщика (encoder) и декодировщика (decoder). Лучевой поиск (beam search).

Механизм внимания (attention) в искусственных нейронных сетях. Механизм внимания в машинном переводе.

Архитектура Transformer (Attention Is All You Need). Механизм self-attention.

Предобученные представления (embeddings). Архитектуры и приемы, используемые в ELMo, GPT (1, 2, 3), BERT, RoBERTa.

2. Введение в обучение с подкреплением

Исторический экскурс. Основные понятия обучения с подкреплением. Стратегия. Агент. Среда. Метод перекрестной энтропии. Генетический алгоритм.

Уравнения Беллмана. Методы value iteration, policy iteration. Value-функция, q-функция.

Обучение без модели среды. Q-learning. Проблема автокорреляции. Double Q-learning. Experience replay. Обзор достижений последних лет.

Оценка градиента для случаев, когда найти градиент аналитически не представляется возможным. Log-derivative trick. Градиент по политике (policy gradient). Алгоритм REINFORCE.

Получение более устойчивых оценок на градиент. Baselines. Метод Advantage Actor Critic (A2C).

Методы обучения с подкреплением в прикладных задачах. Self-critical Sequence Training (в задаче генерации текста).

3. Глубокое обучение в компьютерном зрении

Исторический экскурс. Методы, широко использовавшиеся до популяризации нейронных сетей. Предобученные модели. Дообучение моделей под конкретную задачу.

Задача распознавания и обнаружения объектов на изображении. Object detection. Обзор используемых подходов на примере архитектур: R-CNN, Fast R-CNN, Faster R-CNN, YOLO (1, 2, 3, 4).

Сегментация изображений. Обзор используемых подходов на примере архитектур U-Net, Mask R-CNN.

Перенос стилей между изображениями с помощью нейронных сетей.

Генеративные сети. Вариационный автокодировщик (VAE). Генеративные состязательные сети (GAN). Понятие генератора и дискриминатора. Многокритериальная оптимизация.

Обзор актуальных задач компьютерного зрения: биометрия, трекинг объектов в кадре, анализ поведения, оценка дорожной ситуации в автопилотах, повышение разрешения (super-resolution imaging) и пр.

Семестр: 2 (Весенний)

4. Введение в машинное обучение. Метрические алгоритмы, оценка качества моделей

Основные понятия в машинном обучении. Обзор приложений машинного обучения. Обучение с учителем и без учителя. Задачи: классификация, регрессия, кластеризация, снижение размерности.

Метрические алгоритмы. Метод ближайших соседей (kNN) в задаче классификации и регрессии. Кластеризация и алгоритм k средних (k means).

Байесовский подход. Понятие правдоподобия. Наивный байесовский классификатор.

Отложенная выборка. Кросс-валидация. Переобучение и недообучение. Гиперпараметры.

5. Линейные модели

Линейная регрессия. Метод наименьших квадратов. Градиентный спуск и стохастический градиентный спуск. Переобучение моделей. Регуляризация Тихонова. Теорема Гаусса-Маркова. Функции потерь в задаче регрессии.

Линейная классификация. Понятие отступа. Функции потерь в задаче классификации. Логистическая регрессия. Метод наибольшего правдоподобия. Логистическая функция потерь. Функции Softmax, Sigmoid. Многоклассовая классификация. Регуляризация линейных классификаторов.

Методы оценки качества классификации. Accuracy, Precision, Recall, ROC-AUC, PR-curve, Confusion matrix.

Метод опорных векторов (SVM). Теорема Каруша-Куна-Такера. Двойственная задача. Понятие опорных векторов. Kernel trick (подмена ядра). Регуляризация в SVM.

Метод главных компонент (PCA). Теорема Эккарта-Янга. SVD-разложение. Зависимость объясненной дисперсии от числа компонент.

6. Деревья и ансамбли моделей

Смещение и разброс. Bias-Variance decomposition. Неустойчивость моделей машинного обучения.

Решающее дерево. Рекурсивная процедура построения решающего дерева. Критерии информативности в задаче классификации: энтропийный, Джини; в задаче регрессии. Переобучение решающих деревьев. Прунинг. Регуляризация решающих деревьев. Алгоритмы построения: ID3, C4.5, C5, CART. Небинарные решающие деревья. Связь решающих деревьев и линейных моделей.

Бутстрап. Бэггинг. Out-of-bag error. Метод случайных подпространств (RSM). Случайный лес (Random Forest). Развитие идеи: Extremely Randomized Trees. Сравнение Random Forest и метрических алгоритмов (kNN). Isolation Forest.

Стекинг и блендинг моделей машинного обучения.

Бустинг. Историческая справка, алгоритм AdaBoost. Градиентный бустинг (GBM).

7. Работа с признаками. Ограничения машинного обучения

Проклятие размерности. No Free Lunch Theorem, Wolpert (Теорема о бесплатных обедах). Принцип “Garbage in – garbage out”.

Типы признаков: континуальные, бинарные, категориальные. Работа с разреженными признаками. Работа с пропусками.

Работа с текстовыми данными. Мешок слов (bag of words), TF-IDF.

Оценка значимости признаков. Permutation importance, Partial-dependence plots, shap. Recursive Feature Elimination. LARS.

8. Введение в глубокое обучение

Исторический экскурс. Искусственные нейронные сети. Математическая модель нейрона Маккалока-Питтса. Персептрон Розенблатта. Проблема исключаящего или (XOR problem).

Основные понятия в глубоком обучении (Deep Learning). Метод обратного распространения ошибки (backpropagation). Функции активации: Sigmoid, Tanh, ReLU, Leaky ReLU, ELU, Softmax. Полносвязный слой.

Градиентная оптимизация в глубоком обучении. Методы, основанные на градиентном спуске: Momentum, Nesterov Momentum, Adagrad, Adadelta, RMSprop, Adam, AdamW. Learning rate decay. Начальная инициализация параметров нейронной сети.

Регуляризация в нейронных сетях. Batch normalization. Instance and layer normalization. Dropout. Weight decay. Аугментация данных.

Рекуррентные нейронные сети. RNN. Проблема затухающего градиента (Vanishing gradient). Механизм памяти в LSTM и GRU. Рекуррентные нейронные сети в анализе текстов и последовательностей.

Сверточные нейронные сети. Операция свертки. Сверточный слой (convolutional layer). Нормализация данных. Pooling layer. Пропуск градиента с помощью skip connections. Исторический обзор архитектур и их основных свойств: LeNet, AlexNet, VGGNet, GoogLeNet, ResNet.

Классические подходы к векторизации текстов. Векторное представление слов с помощью нейронных сетей. Word2Vec, GloVe.

Снижение размерности с помощью нейронных сетей. Автоэнкодеры в различных задачах (снижение размерности, фильтрация шумов, поиск аномалий).

9. Обучение без учителя

Кластеризация. Метрический подход, алгоритм k-means. Иерархическая кластеризация. Алгоритм DBSCAN.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

6.Перечень рекомендуемой литературы

Основная литература

Предоставляется на базовой кафедре:

1. Нейронные сети [Текст] : Обучение, организация и применение : учеб. пособие для вузов / В. А. Головкин .— М. : ИПРЖР, 2001 .— 256 с.
2. Математическая статистика [Текст] : [учебник для вузов] / А. А. Боровков .— [3-е изд., испр.] .— М. : Физматлит, 2007 .— 704 с.
3. Математическая статистика [Текст] : оценка параметров, проверка гипотез: учеб. пособие для вузов: доп. М-вом образования СССР / А. А. Боровков .— М. : Наука, 1984 .— 472 с.
4. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. Deep Learning. MIT Press, 2016.
5. Pattern Recognition and Machine Learning / Christopher M. Bishop. Springer, 2011
6. Reinforcement Learning: An Introduction / Richard S. Sutton and Andrew G. Barto. MIT Press, Cambridge, MA, 2018
7. Математическая статистика [Текст] : учеб. пособие для вузов / А. А. Натан, О. Г. Горбачев, С. А. Гуз ; Моск. физико-техн. ин-т (гос. ун-т) .— М. : МЗ Пресс, 2004, 2005 .— 160 с.
8. Математическая статистика [Текст] : [учебник для вузов] / А. А. Боровков .— [3-е изд., испр.] .— М. : Физматлит, 2007 .— 704 с.

Дополнительная литература

Предоставляется на базовой кафедре:

1. Математическая статистика [Текст] : учеб. пособие для вузов / А. А. Натан, О. Г. Горбачев, С. А. Гуз ; Моск. физико-техн. ин-т (гос. ун-т) .— М. : МЗ Пресс, 2004, 2005 .— 160 с.
2. Прикладная статистика. Принципы и примеры [Текст] : [учеб. пособие для вузов] / Д. Кокс, Э. Снелл ; пер. с англ. Е. В. Чепурина ; под ред. Ю. К. Беляева .— М. : Мир, 1984 .— 200 с.
3. Достоверный и правдоподобный вывод в интеллектуальных системах [Текст] / В. Н. Вагин [и др.] .— М. : Физматлит, 2004
4. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. Deep Learning. MIT Press, 2016.
5. Pattern Recognition and Machine Learning / Christopher M. Bishop. Springer, 2011
6. Прикладная непараметрическая регрессия [Текст] = Applied nonparametric regression / В. Хардле; пер. с англ. А. В. Назина, под ред. М. Б. Малютова .— [Научное изд.] .— М. : Мир, 1993 .— 349 с. : ил.
7. Обучение с подкреплением / Р. С. Саттон, Э. Г. Барто ; пер. с англ. Е. О. Романова под ред. Ю. В. Тюменцева .— 2-е изд. — М. : Бином. Лаборатория знаний, 2014 .
8. Задачи классификации и их программное обеспечение (пакет квазар) [Текст] / В. С. Казанцев ; отв. ред. В. Д. Мазуров ; Акад. наук СССР, Уральское отд-ние, ин-т математики и механики .— М. : Наука, 1990 .— 136 с. : ил.
9. Моделирование сложных систем по экспериментальным данным [Текст] / А. Г. Ивахненко, Ю. П. Юрачковский .— М. : Радио и связь, 1987 .— 120 с.
10. Алгоритмы обнаружения эмпирических закономерностей [Текст] / Н. Г. Загоруйко, В. Н. Илкин, Г. С. Лбов ; отв. ред. В. А. Скоробогатов ; Акад. наук СССР, Сибирское отд-ние, Ин-т математики .— [Научное изд.] .— Новосибирск : Наука, 1985 .— 110 с. : ил.
11. Восстановление зависимостей по эмпирическим данным [Текст] / В. Н. Вапник .— М. : Наука, 1979 .— 447 с. : ил.

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. <https://ml-mipt.github.io> – сайт с описанием курса и ссылками на дополнительные ресурсы

2. <http://www.machinelearning.ru> – профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.
3. <http://shad.yandex.ru> – сайт школы анализа данных Яндекса.
4. [http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_\(курс_лекций,_К.В.Воронцов\)](http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_(курс_лекций,_К.В.Воронцов))

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

На лекционных занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

В процессе самостоятельной работы обучающихся предполагается использование таких программных средств, как git, CLI (command-line interface), Jupyter Notebook и др.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладные математика и физика
профиль подготовки:	Физика перспективных технологий: альтернативная энергетика, научное программирование и функциональные материалы Физтех-школа Электроники, Фотоники и Молекулярной Физики кафедра инновационной фармацевтики, медицинской техники и биотехнологии
курс:	1
квалификация:	магистр
Семестры, формы промежуточной аттестации:	
1 (осенний) - Дифференцированный зачет	
2 (весенний) - Экзамен	
Разработчик:	Л.А. Урошлев, канд. физ.-мат. наук

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные научные знания в области физико-математических наук
	ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности
	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач профессиональной деятельности
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.2 Способен применять знания в области физико-математических наук для решения поставленной задачи, формулирования выводов и оценки полученных результатов
	ОПК-4.3 Способен аргументировано выбирать способ проведения научного исследования
ПК-3 Способен профессионально работать с исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)

2. Показатели оценивания компетенций

В результате изучения дисциплины «Машинное обучение» обучающийся должен:

знать:

основные принципы и проблематику теории обучения машин, основные современные методы обучения по прецедентам — классификации, кластеризации и регрессии.

уметь:

формализовать постановки прикладных задач анализа данных, использовать современные методы обучения по прецедентам для решения практических задач, оценивать точность и эффективность полученных решений.

владеть:

основными понятиями теории машинного обучения.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Во время текущего контроля студент должен уметь ответить на следующие вопросы:

Постановка задач обучения с учителем (supervised learning).

Задачи обучения без учителя. Назвать хотя бы две.

Что означает свойство i.i.d.?

Основная идея наивного Байесовского классификатора. В чём его наивность?

Запишите формулы для модели линейной регрессии и для среднеквадратичной ошибки.

Запишите формулу для одного шага градиентного спуска. Как модифицировать градиентный спуск для очень большой выборки?

Что такое правдоподобие, метод максимального правдоподобия? Является ли правдоподобие вероятностью?

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

Что такое переобучение и недообучение? Как их можно детектировать?

Чем гиперпараметры отличаются от параметров? Что является параметрами и гиперпараметрами в линейных моделях и в решающих деревьях?

Что такое регуляризация? Чем на практике отличается L1-регуляризация от L2?

Учитывается ли коэффициент сдвига w_0 в регуляризаторе? Почему?

Почему линейные модели рекомендуется применять к выборке с нормированными значениями признаков?

Запишите формулу для линейной модели классификации. Что такое отступ?

Что такое точность и полнота? Почему нужно учитывать их вместе?

В задаче бинарной классификации доля одного класса составляют 95% выборки. Какие метрики разумно использовать для оценки работы модели? почему?

Что такое ROC-AUC? Как построить ROC-кривую?

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

Идея метода опорных векторов (в случае разделимой выборки).

Опишите жадный алгоритм обучения решающего дерева.

Почему с помощью решающего дерева можно достичь нулевой ошибки на обучающей выборке без повторяющихся объектов?

Если в лист дерева попали объекты разных классов, то какие предсказания нужно выдавать в этом листе? Почему?

Какое предсказание нужно выдавать в листе дерева в задаче регрессии если мы минимизируем MSE? а в случае MAE?

Что такое bagging?

Что такое случайный лес? Чем он отличается от бэггинга над решающими деревьями?

Как в градиентном бустинге обучаются базовые алгоритмы?

Зачем нужен backprop, что такое производная вектора по вектору?

Опишите принцип работы свёрточного слоя (CNN).

В чем недостатки полносвязных нейронных сетей какая мотивация к использованию свёрточных?

Опишите принцип работы базового рекуррентного слоя (RNN).

Что такое dropout?

Как dropout и batch normalization меняют свое поведение при эксплуатации модели (в режиме inference)?

Запишите постановку задачи в методе главных компонент.

Как работает метод k-means?

Во время занятий могут проходить интерактивные обсуждения в чатах курса, что будет являться домашним заданием. Возможно выполнение патентного поиска в качестве самостоятельной задачи. Успешное выполнение всех заданий по курсу и выполнение контрольных срезов знаний дает преимущество на дифференцированном зачете и экзамене.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Вопросы для дифференцированного зачета:

- 1.Постановка задач машинного обучения, их классификация, примеры.
- 2.Метрики качества классификации: accuracy, balanced accuracy, precision, recall, f1-score, ROC-AUC, multiclass extensions.
- 3.Метрики качества регрессии: MSE, MAE, R2, другие варианты.
- 4.Метод максимального правдоподобия (MLE).
- 5.Наивный байесовский классификатор.
- 6.Метод ближайшего соседа (kNN).
- 7.Линейная регрессия.Постановка задачи линейной регрессии. Аналитическое решение МНК, оптимальность оценки, 8.теорема Гаусса-Маркова (формулировка).
- 9.Градиентное решение задачи линейной регрессии.
- 10.Регуляризация: L1 и L2. Свойства, вероятностная интерпретация.
- 11.Логистическая регрессия.

12. Эквивалентность решений, полученных методом максимального правдоподобия и минимизации логистической функции потерь.

Вопросы для экзамена:

1. Стратегии мультиклассовой классификации, их свойства.
2. Метод опорных векторов (SVM), ядра (kernel trick).
3. Алгоритм PCA. Связь с SVD, теорема Эккарта-Янга (формулировка), связь с дисперсией. Практическое применение.
4. Этапы построения модели: тренировка, валидация, тестирование. Роль каждого этапа.
5. Переобучение.
6. Стратегии валидации. Утечки (leaks).
7. Bias-variance tradeoff.
8. Понятие информации, информационной энтропии.
9. Критерии информативности: энтропийный, Джини.
10. Жадный алгоритм построения дерева. Стандарты для построения деревьев.
11. Ансамблирование. Процедура bootstrap. Бэггинг. Метод случайных подпространств.
12. Random Forest, другие леса.
13. Бустинг. Градиентный бустинг.
14. Матричные вычисления. Матричное дифференцирование.
15. Производные основных функций: σ , \tanh .
16. Backpropagation, chain rule.
17. Neural network concept. Fully-Connected layer (FC).
18. Logistic regression as simple NN.
19. XOR problem.
20. Losses for NNs: logistic loss, cross-entropy.

Пример для экзаменационного билета из трех вопросов:

1. Этапы построения модели: тренировка, валидация, тестирование. Роль каждого этапа. Переобучение.
2. Регуляризация нейронных сетей: Dropout, Batch Normalization. Разница в поведении при обучении и предсказании
3. Логистическая регрессия. Эквивалентность решений, полученных методом максимального правдоподобия и минимизации логистической функции потерь

Критерии оценивания

Оценка отлично (10 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (9 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (8 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо (7 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо (6 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо (5 баллов) - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно (4 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно (3 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно (2 балла) - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно (1 балл) - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Аттестация по дисциплине осуществляется в форме дифференцированного зачета. При проведении устного дифференцированного зачета обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету при устном ответе не должен превышать одного астрономического часа. При проведении устного экзамена обучающемуся предоставляется 60 минут на подготовку. Опрос обучающегося по билету на устном экзамене не должен превышать одного астрономического часа.