

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное учреждение науки
Институт системного программирования им. В.П. Иванникова
Российской академии наук
(ИСП РАН)

Одобрено решением учёного совета ИСП РАН.

Протокол № 2023-3 - от 07 июля 2023 года.



«УТВЕРЖДАЮ»

Директор ИСП РАН

д.ф.-м.н., академик РАН

А.И. Аветисян

«07» июля 2023 г.

ПРОГРАММА

вступительного экзамена в аспирантуру

по специальности

1.2.1 – Искусственный интеллект и машинное обучение

по физико-математическим наукам.

Москва 2023

Научная специальность 1.2.1. «Искусственный интеллект и машинное обучение»

Базовое машинное обучение

1. Постановка задачи машинного обучения. Регрессия, классификация
2. Измерение качества классификации: accuracy, precision, recall, f1-score, ROC-AUC.
3. Измерение качества регрессии: MSE, MAE, R2
4. Оценка максимального правдоподобия. Ее связь с регрессией и классификацией
5. Наивный байесовский классификатор
6. Метод k-ближайших соседей (KNN)
7. Линейная регрессия. Постановка задачи для случая функции потерь MSE. Аналитическое решение. Теорема Гаусса-Маркова. Градиентный подход в линейной регрессии
8. Регуляризация в линейных моделях: L1 и L2, их свойства. Вероятностная интерпретация
9. Логистическая регрессия. Эквивалентность подхода MLE и минимизации логистических потерь
10. Мультиклассовая классификация. Один против одного, один против всех, их свойства
11. Метод опорных векторов (SVM). Проблема оптимизации для SVM. Ядерный метод. Свойства ядра.
12. Метод главных компонент (PCA). Связь с сингулярным разложением.
13. Этапы разработки модели: обучение, валидация и тестирование. Проблема переобучения, способы ее обнаружения. Стратегия валидации, метод перекрестной проверки, утечка данных
14. Процедура построения дерева решений. Информационные критерии. Энтропийные критерии, примесь Джини. Бутстрэп, бэггинг, метод случайного леса.
15. Бустинг и градиентный бустинг основные идеи
16. Искусственные нейронные сети. Обучение нейронной сети. Метод обратного распространения ошибки, цепное правило. Полносвязный слой (FC). Проблема XOR

17. Функции потерь для нейронной сети: логистическая функция потерь, кросс-энтропия. Функции активации, их влияние на нейронную сеть, вычислительная сложность. Многопеременная логистическая функция (Softmax)
18. Регуляризация в глубоком обучении: Dropout, Batch Normalization. Различия в этапах обучения и оценки

Прикладные задачи (Обработка текста, компьютерное зрение, обучение с подкреплением)

19. Задача классификации последовательностей. Методы: HMM, MEMM и CRF
20. Задача извлечения информации. Распознавание именованных сущностей, разметка последовательностей
21. Постановка задачи RL. Состояние, действие, вознаграждение, окружающая среда, действие
22. Q-обучение, приближенное Q-обучение, DQN, experience replay, двойной DQN, проблема автокорреляции
23. Постановки задач компьютерного зрения: классификация, обнаружение, сегментация. Метрики в CV: IoU, mAP
24. Матричная свертка. Сверточный слой, обратное распространение через него. Свертки 1x1, сравнение со слоями FC. Max/Average Pooling

Математика машинного обучения

25. Центральная предельная теорема. Закон больших чисел
26. Нулевая гипотеза. Уровень значимости. P-value. Ошибки 1-го и 2-рода
27. Выбор статистических гипотез. Простые и сложные статистические гипотезы. Мощность статистического критерия. Наиболее мощный и равномерно наиболее мощный критерий
28. T-критерий Стьюдента. U-критерий Манна-Уитни.
29. Интервальные оценки и доверительные интервалы. Понятия интервальной оценки и доверительного интервала.
30. Сравнение классификаторов. Критерий Уилкоксона. Проблемы использования статистической проверки гипотез
31. Критерий Неймана-Пирсона для двух простых гипотез. Решающее правило, оптимальное по критерию Неймана-Пирсона

Технологии программирования

32. Системы управления базами данных. Понятия базы данных, системы баз данных и систем управления базами данных (СУБД). Требования к СУБД. Характеристики, функции СУБД
33. Модели данных. Реляционная модель. Иерархическая модель. Сетевая модель.
34. Реляционная модель данных. Тип данных, домен, отношение, кортеж. Первичный ключ, внешний ключ, ссылочная целостность. Фундаментальные свойства отношений. Основные операции над отношениями. Реляционная алгебра и реляционное исчисление.
35. Проектирование баз данных. Декомпозиция без потерь. Теорема Хита. Нормальные формы. Аномалии обновления.
36. Понятие виртуальной машины. Основные компоненты виртуальной машины. Гипервизоры первого и второго рода: функции и отличия. Контейнеризация и виртуализация.
37. Принципы построения распределенных СУБД. CAP-"теорема". Принципы BASE. Примеры распределенных СУБД каждого класса в соответствии с CAP- " теоремой

Литература

- [1] Understanding and visualizing convolutional neural networks. URL: <http://cs231n.github.io/understanding-cnn/>.
- [2] Курс лекций ИСП РАН основы обработки текстов. URL: <https://www.youtube.com/watch?v=QnuLkiXKjc8&list=PL5cBzMoPJgCXFdSvWaun0y4cILirWHMD>.
- [3] Daniel Jurafsky, James H. Martin. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Second Edition. Prentice Hall.
- [4] Christopher D. Manning, Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.
- [5] Fadi AlMahamid, Katarina Grolinger. Reinforcement learning algorithms: An overview and classification. In *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1-7. IEEE, 2021.
- [6] Amir Dembo. Probability theory. URL: <https://web.stanford.edu/~montanar/TEACHING/Stat310A/lnotes.pdf>.
- [7] Ian Goodfellow, Yoshua Bengio, Aaron Courville. *Deep Learning*. MIT Press, 2016. <https://www.deeplearningbook.org>.
- [8] Erik Learned-Miller. Vector, matrix, and tensor derivatives. URL: <http://cs231n.stanford.edu/vecDerivs.pdf>.
- [9] Kevin P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. URL <https://probml.github.io/pml-book/book1.html>.
- [10] Г Андерсон. Введение в многомерный статистический анализ. М.: Физматгиз, 500: 1, 1963.
- [11] Александр Алексеевич Боровков. Математическая статистика. 2007.
- [12] Леман Э. и Прохоров Ю. В. Проверка статистических гипотез. 1964.
- [13] В.Н. Вапник, А.Я. Червоненкис. О равномерной сходимости частот появления событий к их вероятностям. *Теория вероятностей и ее применения*, 16(2), 1971.
- [14] Д.П. Ветров, Д.А. Кропотов. Байесовские методы машинного обучения. *Учебное пособие*. М, 2007.
- [15] КВ Воронцов. Математические методы обучения по прецедентам (теория обучения машин). Москва, 2011.
- [16] Кузнецов, С. Д. *Основы современных баз данных*. М: Центр Информационных

Технологий, 1998.

- [17] Гектор Гарсиа-Молина, Джеффри Д Ульман, Дженнифер Уидом. *Системы баз данных: Полный курс*. Вильямс, 2003.
- [18] К Дж Дейт. *Введение в системы баз данных*. Вильямс, 2008.
- [19] Соколов Евгений. Лекции Евгения Соколова. URL <https://github.com/esokolov/ml-course-hse/tree/master/2021-fall/lecture-notes>.
- [20] И.С. Енюков, Л.Д. Мешалкин. *Прикладная статистика: Исследование зависимостей*. 1985.
- [21] В А Епанечников. *Непараметрическая оценка многомерной плотности вероятности. Теория вероятностей и ее применения*, 14(1), 1969.
- [22] Сергей Михайлович Ермаков, Геннадий Алексеевич Михайлов. *Курс статистического моделирования*. 1976.
- [23] Николай Григорьевич Загоруйко. *Прикладные методы анализа данных и знаний*. 1999.
- [24] Гаральд Крамер. *Математические методы статистики*. Регулярная и хаотическая динамика, 2003.
- [25] Игорь Давидович Мандель. *Кластерный анализ*. Finansy i statistika, 1988.
- [26] АВ Мерков. Основные методы, применяемые для распознавания рукописного текста. *доступно по адресу <http://www.recognition.msc.ru/pub/RecognitionLab.html/methods.html>*, 2004.
- [27] Нейчев. Лекции Нейчева, базовое машинное обучение. URL: https://www.youtube.com/watch?v=8s9073kNXgY&list=PL4_hYwCyhAvZyW6qS58x4uElZgAkMVUvj.
- [28] Нейчев. Лекции Нейчева, прикладное машинное. URL: https://www.youtube.com/watch?v=I9FF-3UAm4o&list=PL4_hYwCyhAvZeq93ssEUaR47xhvs7IhJM.
- [29] Синицин Ф Федотов С. Учебник по машинному обучению, школа анализа данных. URL: <https://academy.yandex.ru/handbook/ml>.
- [30] АМ Шурыгин. *Прикладная стохастика: робастность, оценивание, прогноз*. М.: *Финансы и статистика*, 223, 2000.