

1. Дайте определение задаче обучения по примерам. Опишите основные варианты постановки задачи обучения по примерам. Опишите постановку задачи обучения по признакам. Опишите основные типы алгоритмов классификации(параметрические и непараметрические) и их особенности.
2. Опишите основные этапы модели обработки информации CRISP. Определите место методов машинного обучения в рамках подхода Data Driven.
3. Опишите основные подходы к оценке состава и качества признаков. Приведите примеры применения методов статистического анализа к исследованию качества данных.
4. Опишите основные подходы к решению вопросов генерации признаков. Оцените влияние методов преобразований признаков на решение обучения. Приведите примеры эффективных преобразований признаков.
5. Опишите модели преобразования пространства входных признаков (виды, их свойства и особенности). Опишите модель линейного преобразования пространства признаков на примере Метода Главных Компонент (PCA показать на примере `sklearn.decomposition.PCA`).
6. Опишите алгоритмы Заполнения пробелов (регрессия, классификация, ZET)
7. Опишите алгоритм селекции признаков на основе фильтрации (показать на примере `sklearn.feature_selection.SelectKBest`),на основе обертывания (показать на примере `sklearn.feature_selection.SelectFromModel`), на основе регуляризации (показать на примере `sklearn.linear_model.Lasso`)
8. Опишите основные положения теории Статистического обучения. Определите регрессионную модель обучения и опишите ее основные свойства.
9. Опишите алгоритм линейной регрессии, опишите его ограничения и методы формирования решений для задач нелинейного типа. Определите функцию потерь и метрики для оценки моделей.
10. Опишите алгоритм Стохастический Градиентный спуск и влияние методов регуляризации на решение задач обучения с учителем. Опишите способ формирования текущей оценки потерь для Стохастического Градиентного Спуска.
11. Опишите алгоритм Логистическая регрессия, определите функцию потерь для него и тип решаемых задач `sklearn.linear_model.LinearRegression`.
12. Опишите модель классификации на основе SVM (Машины Опорных Векторов). Определите функцию потерь. Перечислите основные особенности метода опорных векторов для линейно-разделимого пространства покажите пример на основе `sklearn.svm`.
13. Опишите алгоритмы генерации деревьев решений на примере CART. Определите основные варианты критериев информативности признаков в алгоритмах генерации деревьев. Опишите основные особенности и проблемы алгоритмов на основе деревьев
14. Опишите особенности метрических алгоритмов . Опишите особенности алгоритма KNN и его модификации (возвещенный KNN, Парзеневские окна ). Проблемы KNN и применение техник эффективной реализации KNN (вложения, проекции)
15. Опишите работу EM алгоритма.
16. Опишите типы ошибок задачи классификации и методы их оценки.

17. Приведите определение сложности классификатора по Вапнику-Червоненкису. Оцените сложность линейного классификатора. Опишите явление переобучения и перечислите методы борьбы с ним.
18. Опишите особенности вероятностного подхода к решению задачи обучения с учителем. Опишите алгоритм Наивный Байесов Классификатор.
19. Опишите особенности взаимодействия простых классификаторов на основе ансамбля. Поясните особенности дилеммы дисперсии-смещения для ансамбля решателей.
20. Опишите алгоритм формирования Случайного леса и особенности его работы (покажите пример на основе `sklearn.ensemble.RandomForestClassifier`)
21. Опишите алгоритм Градиентного бустинга, особенности его работы, схематический пример и покажите пример на основе `sklearn.ensemble.GradientBoostingClassifier`
22. Приведите постановку задачи кластеризации. Опишите методы вычисления расстояния между множествами. Опишите Алгоритм KMeans, определите границы его применения и приведите схематический пример его работы и работы в `sklearn.cluster.KMeans`.
23. Опишите Физиологическую формальную модель нейрона. Приведите примеры и охарактеризуйте свойства активационных функций. Опишите формальную модель однослойной нейронной сети. Опишите формальную модель многослойной нейронной сети. Опишите Задачи, решаемые в НС. Опишите режимы работы нейронных сетей (общая характеристика).
24. Опишите алгоритм обучения нелинейного нейрона в модели МакКалок-Питца. Дельта правило.
25. Опишите общую схему алгоритма обратного распространения ошибки. Ошибка скрытого слоя в алгоритме обратного распространения ошибки. Градиентный метод изменения весов. Перечислите основные проблемы обучения многослойных нелинейных нейронных сетей.
26. Модификации обратного распространения ошибки. Алгоритмы на основе метода моментов. Алгоритмы адаптации шага обучения. Квази-Ньютоновское обучение.
27. Перечислите принципы и особенности функционирования самоорганизующихся карт. Модель нейронной сети «Самоорганизующаяся карта Кохонена». Нейрон Кохонена (WTA). Свойства нейронной сети «Самоорганизующаяся карта Кохонена». Алгоритм обучения нейронной сети «Самоорганизующаяся карта Кохонена» и его связь с алгоритмом kmeans.
28. Модель нейронной сети Хопфилда. Условие устойчивости сети Хопфилда. Свойства сетей Хопфилда. Функционирование сети Хопфилда при решении задачи.
29. Обучение с использованием отжига (Стохастическое обучение).
30. Сверточные сети. Основные архитектурные особенности. Алгоритм обучения. Основные проблемы
31. Опишите подходы к решению проблемы исчезновения градиента в глубоких нейронных сетях. Резидуальные сети.
32. Реализация полностью сверточных сетей и их применение для задач семантической сегментации и шумоподавления для изображений.

33. Опишите реализацию сетей рекуррентного типа в форме RNN и LSTM. Опишите основные этапы алгоритма BPTT. Опишите особенности функционирования сетей LSTM.
34. Современные подходы к формированию обучения с «Вниманием». Сети архитектуры Трансформер. (самостоятельно разобрать этот вопрос : на Хабр есть неплохой цикл статей : )
35. Опишите особенности формирования вероятностных моделей в нейросетевом базисе. Генерация Вариационных моделей и особенности их обучения. Схема Вариационного автоэнкодера.
36. Генеративно-адверсивные нейронные сети и методы их построения. GAN, CGAN, CycleGAN.
37. Методы обучения нейронных сетей на основе контрастивного подхода. Схемы Metric learning, SimCLR. Перенос обучения. Методы дистилляции.
38. Методы аугментации данных для обучения нейронных сетей

### ЗАДАЧИ

39. Вычислить ошибку на первом шаге обучения нейрона.
40. Вычислить коррекцию настраиваемых параметров нейрона и измените их при локальном обучении нейрона.
41. Определите значение метрик (для классификации или регрессии)
42. Определите значение потерь (для классификации или регрессии)
43. Оценить степень мультиколлинеарности данных для признаков с взаимной парной корреляцией
44. Выбрать список 3х наиболее важных признаков для модели признаков, где коэффициент корреляции с целевым признаком
45. Целевой признак принял значения : (1, 0, 1, 1, 0, 1, 0, 0, 1, 0). Для соответствующих примеров выход модели равен: (0.9, 0.5, 0.8, 0.7, 0.6, 0.5, 0.2, 0.2, 0.6, 0.3). Определить порог принятия решений (калибровать модель) по метрике f1
46. Определить опираясь на значения настраиваемых параметров линейной регрессии набор из 5 наиболее существенных признаков:  $w_0 = 1, \langle w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10} \rangle = \dots$
47. Значения признака приняли следующие значения: (1., 2., 2., 1.5, 4., 2., 3., 2.5, 3., 2., 1.1 ), целевой признак при это принимал значения: (1, 1, 1, 1, 1, 0, 0, 0, 0, 0) . Оцените информативность признака через коэффициент Джини
48. Определить значение потерь L2 и оценки полных потерь  $Q = \sum L_2(X_0)$  для алгоритма стохастического градиентного спуска, если на шаге t:  $Q(t) = 0.1$  ,  $y_{true} = 1$  ,  $y_{pred} = 0.5$
49. Оцените кросс-энтропию для следующей ситуации (3 класса):  $y_{true} = ((1,0,0), (0,1,0), (0,0,1))$ ,  $y_{pred} = ((0.2, 0.4, 0.4), (0.1, 0.7, 0.2), (0.3, 0.1, 0.6))$
50. Проведите кодирование категориального признака X методом векторного представления (дамми) при условии, что:  $X = (R, G, B, R, G, B, R, G, B)$ ,  $Y = (1, 1, 2, 2, 0, 0, 0, 0, 0)$ .
51. Проведите кодирование категориального признака X методом Label Encoding при условии, что:  $X = (R, G, B, R, G, B, R, G, B)$ ,  $Y = (1, 1, 0, 2, 2, 0, 1, 0, 0)$ . Определите , что необходимо знать, для проведения такого кодирования

### куда смотреть в доп.материалы:

лекции Воронцов - Voron-ML-1.pdf (есть в сети и в книгах)— основные теоретические вопросы для классификаторов, регрессоров, кластеризаторов

Математические методы распознавания образов Курс лекций МГУ, ВМиК, кафедра «Математические методы прогнозирования» © Местецкий Леонид Моисеевич, 2002–2004 <http://www.ccas.ru/frc/papers/mestetskii04course.pdf> - Основная книга про

признаки и их преобразование, основные теоретические вопросы для классификаторов, регрессоров, кластеризаторов

Загоруйко Н.Г. Прикладные методы анализа данных и знаний. Изд. ИМ СО РАН, Новосибирск, 1999.(есть в сети ) - FOREL, ZET

Саймон Хайкин Нейронные сети. Полный курс ([https://avmim.com/wp-content/uploads/2019/02/Neyronnye\\_seti\\_Polny\\_kurs.pdf](https://avmim.com/wp-content/uploads/2019/02/Neyronnye_seti_Polny_kurs.pdf))2006 — PCA, регрессионная модель обучения, теория статистического обучения, VCdim (скачать на след. семестр)

sklearn <https://scikit-learn.org/stable/modules/classes.html>

OpenCV - <https://habr.com/ru/post/519454/>,

[https://docs.opencv.org/4.x/d6/d00/tutorial\\_py\\_root.html](https://docs.opencv.org/4.x/d6/d00/tutorial_py_root.html)

scipy - <https://pythonru.com/biblioteki/scipy-python>

короткое резюме всех методов <https://habr.com/ru/post/448892/>,

[https://neerc.ifmo.ru/wiki/index.php?title=%D0%9E%D0%B1%D1%89%D0%B8%D0%B5\\_%D0%BF%D0%BE%D0%BD%D1%8F%D1%82%D0%B8%D1%8F](https://neerc.ifmo.ru/wiki/index.php?title=%D0%9E%D0%B1%D1%89%D0%B8%D0%B5_%D0%BF%D0%BE%D0%BD%D1%8F%D1%82%D0%B8%D1%8F)

SimCLR : <https://arxiv.org/pdf/2002.05709.pdf>,

<https://wandb.ai/sayakpaul/simclr/reports/Towards-Self-Supervised-Image-Understanding-with-SimCLR--VmlldzoxMDI5NDM>

Metric Learning <https://habr.com/ru/company/ods/blog/695380/>

Transformer <https://habr.com/ru/post/486358/>