Список вопросов к экзамену по дисциплине

«Методы машинного обучения»

Папулин С.Ю.

1. Основные этапы построения моделей в задачах регрессии и классификации.

1 лекция, обьяснить на примере домашней работы классификация текста.

1. Проверка гипотез. t-критерий с одной выборкой.

уровень значимости генеральная средняя нарисовать распределение стьюдента , что такое пи value

1. Выбор модели с использованием кросс-валидации.

В качестве примера степень полинома, как в рк рассказываем, как строится каждая точка

1. Оценка параметров. Метод максимального правдоподобия.

В общем виде про метод, в качестве примера нормальный закон распределения и как в нем найти парамтры.

1. Задача оптимизации. Градиентный спуск.

Что есть функция F(x) ограничения…, градиэнтный спуск это подвид.

1. Регрессия. Метод наименьших квадратов и градиентный спуск.

Методы поиска наименьших параметров есть такие… вводная часть Понимание как работает алгоритм

Сказать как оно расчитывается формула без регуляризации

Градиентный спуск -> функция потерь

1. Регрессия. Градиентный спуск и стохастический градиентный спуск.

По стахостическому -> можем применить к определенной форме нашей функции

1. Регрессия и классификация. Метод ближайших соседей.

Формулы регрессии классификации и ближ соседей

1. Регрессия. Деревья решений. Критерии деления.

Про деревья про бинарное деление на каждом шаге выбираем признак и точку деления что каждый раз подпространство делим на два и что суммы Rss слева и справа минимальна

1. Регрессия и классификация. Ансамблевые моделей. Бэггинг, бустинг и стекинг

Параллельное обучение и последовательное обучения

В качестве бустинга обучение лин регрессии

1. Регрессия и классификация. Деревья решений. Бэггинг, случайный лес и сверхслучайные деревья.

Все в контексте деревьев сделать вводную часть по деревьям

1. Регрессия. L1, L2 регуляризация.

Функцию потерь л1 л2 показать регуляризацию через ограничения

Для чего нужна что он контролирует что если он 0 что если он бесконечность

1. Классификация. Логистическая регрессия (бинарная классификация).

Функция потерь лог регрессии графики ошибок

Форстропия

Написать Yшапка с системой если, перейти от вероятностного к системе уравнений с условием

1. Классификация. Полиномиальная логистическая регрессия (многоклассовая классификация).
2. Классификация. Методы определения качества классификаторов.

acuracy precision recall f1 f2

Вводная часть

1. Классификация. Наивный байесовский классификатор с нормальным законом распределения.

Общая формула базовую часть а потом для разных законов + нормального распределения формула

Параметры полученные это сред значение для каждого класса и дисперсия

1. Классификация. Метод опорных векторов.

До задачи оптимизации прямой что такое гиперплоскость

Что такое зазор

Для линейноразделимого пространства достаточно

1. Классификация. Деревья решений. Критерии деления.

В регрессии ищем признак

Улучшение распределенности информации справа и слева

1. Классификация. L1, L2 регуляризация.
2. Классификация текстовых документов с использованием наивного байесовского классификатора. Модель Бернулли.

В контексте текстовых методов с наивным байесом в файлах старосты(связанном с рк 2) (Minor\_Lecture\_5\_NB вроде)

Общая постановка задачи и классияикация параметров

1. Классификация текстовых документов с использованием наивного байесовского классификатора. Мультиномиальная модель.

Как в 20

1. Классификация. Многослойная нейронная сеть прямого распространения.

В общем виде про классиф неиронную сеть

И алгоритм функция активации и какие они бываают

Для разных задач разный входной слой

Напр бля бинарной классиф сигмоид

Про обучение не надо

1. Кластеризация. Метод k-средних. Выбор начальных значений центров кластеров.

Алгоритм к средних описать алгоритм и формулу

Формулу внутренне кластерной вариации

1. Кластеризация. Иерархическая кластеризация. Агломеративные методы.

Общие принципы,

Меры близости между наблюдениями

Про обьединение кластеров

4 варианта: полные связи, центроидные,…

Что такое полные связи

1. Уменьшение размерности. Метод главных компонент PCA.

Про размерность

Исход набор данных в новом пространстве, в котором наши признаки независимы

Для этого использовать преобразование, это вектора нашей матрицы

Нужна диагональная корреляционная матрица в итоге, такое должно быть преобразование (суть в том что корреляции только на диагонали)

Вывод не надо писать

Про п вектора

1. Рекомендательные системы. Коллаборативная фильтрация на основе сходства пользователей.
2. Рекомендательные системы. Коллаборативная фильтрация на основе сходства элементов.
3. Рекомендательные системы. Факторизация матрицы рейтингов. ALS.
4. Распределенные алгоритмы. Расчет среднего значения и стандартного отклонения.
5. Распределенные алгоритмы. Расчет косинусной меры сходства.
6. Распределенные алгоритмы. Градиентный спуск.
7. Распределенные алгоритмы. Стохастический градиентный спуск.
8. Распределенные алгоритмы. Факторизация матрицы (один из вариантов).