

Билеты к экзамену январь 2025

1. Матричное дифференцирование. Определение матричной производной. Градиентный спуск (GD), стохастический градиентный спуск (SGD), метод Ньютона. Пример нахождения ММП оценок для многомерного нормального распределения.
2. Понятие явления переобучения. Способы борьбы с переобучением. Кросс-валидация, её виды.
3. Метрические методы классификации и регрессии.
4. Решающие деревья. Критерии расщепления, выбор оптимальных порогов. Способы борьбы с переобучением в решающих деревьях. Обработка категориальных признаков, обработка пропусков.
5. Линейные методы классификации и регрессии. Определение отступа. Различные виды верхних оценок на функционал доля ошибок. Логлосс, Hinge-loss. Обобщения линейных моделей для задачи классификации на K классов. Походы One-vs-one и one-vs-all.
6. Логистическая регрессия. Вывод формулы логистической регрессии. Регуляризация функционала в логистической регрессии. Заставляет ли функционал в задаче логистической регрессии предсказывать вероятности?
7. Метрики качества классификации и регрессии. Матрица ошибок. Метрики в случае многоклассовой задачи классификации. Рок-кривая, площадь под ней, явная формула для рок-кривой.
8. SVM. Вывод оптимизационной задачи. Теория двойственности. Сильная/слабая двойственность. Вывод двойственной задачи в SVM.
9. Задача снижения размерности. Вывод постановки оптимизационной задачи PCA, её решение. Критерии отбора оптимального числа главных компонент. Нелинейные методы снижения размерности (MDS, SNE, t-SNE, Autoencoder, ...)
10. Word2Vec. SkipGram NS и CBOW. Fasttext. Применение для задач, связанных с обработкой текстов.

11. Постановка задачи кластеризации. Алгоритмы K-means, DBSCAN, Иерархическая кластеризация. Критерии качества кластеризации.
12. Байесовский подход в машинном обучении. Априорные и апостериорные распределения параметров. Виды точечных оценок параметров (MAP/bayes оценки). EM-алгоритм.
13. Рекомендательные системы: общие подходы (эвристический, коллаборативная фильтрация, контентные модели, гибридные подходы, ...). LFM модели в рекомендациях на основе коллаборативной фильтрации (SVD, ALS). Двухшаговый подход в рекомендательных системах. Типы признаков для моделей второго шага. Примеры моделей, используемых на втором шаге.
14. Bias-variance decomposition. Примеры разложений для некоторых моделей (например для бэггинга).
15. Построение ансамблей. Бэггинг и случайный лес, Градиентный бустинг.
16. Полносвязные нейронные сети. Функции активации, функции потерь. Автоматическое дифференцирование. Прямой проход по сети. Обратный проход по сети. Решение различных задач машинного обучения с помощью нейронных сетей: снижение размерности, классификация. Архитектура Автокодировщика.
17. Задача машинного перевода (seq2seq). Подход с помощью RNN. Механизм внимания (attention). Устройство self-attention. Трансформер.