

Список вопросов к экзамену:

1. Метод ближайшего соседа для задачи классификации KNN, для задачи регрессии.
2. Проблема переобучения. Разделение данных на train-test как способ подбора гиперпараметров модели.
3. Линейная регрессия. Постановка задачи линейной регрессии. Аналитическое решение МНК. Градиентное решение задачи линейной регрессии.
4. Метрики качества регрессии: MSE, MAE, R2, RMSLE
5. Регуляризация линейных моделей. L1 и L2 регуляризация, их вероятностная интерпретация.
6. Наивный байесовский классификатор
7. Алгоритм Решающего дерева. Понятие информации, информационной энтропии. Критерии информативности: энтропийный, Джини.
8. Регуляризация как ограничения на модель (глубина и кол-во деревьев, ограничение числа листьев в дереве и др.)
9. Матрица ошибок (Confusion matrix): Accuracy, precision, recall, sensitivity, specificity, ROC AUC.
10. Работа с категориальными признаками и пропущенными значениями. One-Hot encoding. Mean encoding. Target encoding. Weight of evidence.
11. Подбор гиперпараметров моделей. Кросс-валидация.
12. Процедура bootstrap. Бэггинг. Метод случайных подпространств (сабсэмплинг признаков). Смесь моделей и смесь экспертов. Random Forest.
13. Градиентный бустинг
14. Задача кластеризации. DBSCAN, k-means, Иерархическая кластеризация.
15. Задача снижения размерности. Алгоритм PCA. Другие методы снижения размерности.

План занятий 14 занятий, последнее занятие экзамен:

70%-80% (10-12/1)

- Проверка
- Составление

30% экзамен