

МАШИННОЕ ОБУЧЕНИЕ

СПИСОК ВОПРОСОВ К ПИСЬМЕННОЙ РАБОТЕ

1. Дайте определение линейного парного коэффициента корреляции. Как проверить гипотезу о его значении? Как построить доверительный интервал? Когда его следует/не следует использовать? Что такое коэффициент корреляции Спирмена и Кенделла? В каких ситуациях их применяют? Как проверить значимость этих коэффициентов?
2. Что такое корреляционная матрица? Как ее интерпретировать? Что такое мультиколлинеарность и наведенная зависимость? Как очистить коэффициент корреляции от влияния других факторов? Что такое коэффициент конкордации?
3. Что такое многократные повторные выборки? Расскажите о критерии знаков и критерии Уилкоксона. Расскажите про метод ANOVA.
4. Опишите метод наименьших квадратов для построения линейной регрессионной модели. Проведите вывод метода из функции правдоподобия, выпишите систему нормальных уравнений.
5. Примените МНК для построения модели $\theta_0 + \theta_1 X_1$. Приведите формулы для коэффициентов модели, постройте доверительные интервалы для них. Сформулируйте теорему Гаусса–Маркова.
6. Как проверить значимость того или иного предиктора в модели? Как проверить гипотезу о равенстве коэффициента модели заданному числу? Как проверить гипотезу о значимости группы факторов модели?
7. Дайте определение состоятельных оценок. При каких условиях оценки коэффициентов модели методом МНК являются состоятельными? Что можете сказать об их асимптотической нормальности?
8. Что такое RSS, коэффициент детерминации, скорректированный коэффициент детерминации, BIC, AIC? Для чего применяют эти показатели?
9. Расскажите о Ридж– и Лассо–регуляризации. Зачем их применяют? Выведите одну из них (любую, на свой выбор) из байесовского подхода к оценке параметров.
10. Что требуют от остатков линейной регрессионной модели? Что такое студентизированные остатки и зачем их вычисляют? Как проверить остатки модели на нормальность?
11. Что такое автокорреляция остатков, когда она может возникнуть? Можно ли это явление заметить на графике остатков? Какие тесты на автокорреляцию вы знаете? Опишите два из них подробно.
12. Что такое гетероскедастичность остатков, когда она может возникнуть? Можно ли это явление заметить на графике остатков? Какие тесты на гетероскедастичность вы знаете? Опишите два из них подробно.

13. Что такое преобразование Бокса–Кокса и Йео–Джонсона, для чего их применяют? Как их применять (опишите не менее трех способов)?
14. Как подбирать набор предикторов модели (опишите метод последовательного включения и метод последовательного исключения)? Что такое ADD–Del–алгоритм?
15. Что такое бинарные переменные и one–hot–кодирование? Как это кодирование применять и как интерпретировать его результаты (приведите несколько примеров)? Покажите на примерах разумность введения в модель предикторов, равных произведению числового и бинарного фактора.
16. Приведите примеры других (кроме квадратичной) функций потерь при построении регрессионной модели. Расскажите о методах LAD, Хубера, Тьюки, LMS. Выведите LAD из метода наибольшего правдоподобия.
17. Сравните на примерах методы LS, LAD и LMS. Выведите регрессию Пуассона из метода наибольшего правдоподобия.
18. Расскажите, как искать выбросы с помощью критерия Граббса, теста Диксона, фильтра Хемпеля. Расскажите про метод isolation forest. Если выбросы уже найдены, то как их исправить (приведите несколько рекомендаций)?
19. Что такое латентные переменные? Опишите ЕМ–алгоритм выделения латентных переменных.
20. Дайте определения временных рядов — White Noise, Random walk, Moving Average, Autoregression, Drifted White Noise. Расскажите об учете сезонных колебаний в аддитивных и мультипликативных моделях. Дайте определение стационарного временного ряда. Опишите тест Дики–Фуллера.
21. Опишите способы моделирования временного ряда: $AR(p)$, $MA(q)$, $ARMA(p,q)$, $ARIMA(n,p,q)$ (приведите уравнение модели и способ вычисления коэффициентов). Как проверяют качество таких моделей?
22. Как построить регрессионную модель, если предикторы и/или таргет являются временными рядами? Что такое детрендрование? Сформулируйте теорему Гаусса–Маркова для случая временных рядов.