

Learning outcomes №2

1. R^2 – коэффициент детерминации: формула расчета, геометрическая интерпретация
2. R^2 скорректированный: формула расчета, отличие от классического коэффициента детерминации
3. Кросс-валидация в контексте регрессионного анализа. Алгоритм k-фолдовой кросс-валидации
4. Проблема переобучения
5. Регуляризация в линейной регрессии. Гребневая регрессия (ridge regression)
6. Связь Variance Inflation Factor (VIF) для предиктора и дисперсии оценки коэффициента при этом же предикторе
7. Сравнение вложенных линейных регрессионных моделей при помощи F-теста
8. Информационные критерии AIC, BIC: понимать суть использования информационных критериев, уметь делать вывод, знать, какой штраф накладывается за добавление новых параметров в AIC и BIC
9. Линейная вероятностная модель применительно к бинарному отклику и ее ограничения
10. Представление модели бинарного выбора:
 - подход, основанный на представлении зависимой переменной как латентной
 - подход, использующий переход от вероятности к шансам
11. Вероятность того, что Y принимает значение 1 как функция распределения от предсказанной части модели
12. Допущения об ошибках: логит- и пробит-модель. Знать функцию стандартного логистического распределения и уметь ее применять
13. Сигмоида: понимать, почему такой характер зависимости вероятности того, что Y принимает значение 1, от значения предиктора
14. Интерпретация оценок коэффициентов:
 - исходные (в терминах склонности)
 - через отношения шансов
 - через предельные эффекты
15. Уметь переводить исходные оценки модели в отношения шансов
16. Тест Hosmer-Lemeshow: какую гипотезу проверяем, логика реализации теста, ограничения теста
17. Понимать, как устроена confusion matrix
18. Уметь считать по confusion matrix (и, разумеется, понимать, что это за величины, что они показывают)
 - ошибку первого рода
 - ошибку второго рода
 - мощность критерия (чувствительность / recall)
 - специфичность

- точность (accuracy) и сравнивать с базовой точностью (baseline accuracy)

19. Понимать дилемму соотношения специфичности и чувствительности
20. ROC-кривая: как устроен график, что показывает AUC (площадь под ROC-кривой)
21. Сравнение вложенных моделей посредством теста отношения правдоподобия (likelihood-ratio test): нулевая гипотеза, как устроена статистика критерия, распределение статистики критерия, итоговый вывод о выборе более / менее экономной модели
22. Представление спецификации логистической модели с порядковым откликом
 - подход, основанный на представлении зависимой переменной как латентной
 - подход, использующий переход от вероятности к шансам
23. Уметь рассчитывать вероятность того, что Y принимает конкретное значение, в порядковой регрессии
24. Уметь рассчитывать функцию распределения от конкретного значения категории в порядковой регрессии
25. Допущение о параллельности регрессий: смысл допущения, последствия нарушения допущения, тестирование допущения
26. Представление спецификации мультиномиальной модели через набор моделей с бинарным откликом