

Домашнее задание 3

Deadline: 23.59 10 марта 2021

Общая постановка задачи Задание выполняется на уже знакомом Вам массиве CDAhw2.dta. Описание переменных представлено в этом [файле](#). Вам необходимо смоделировать участие в голосовании на выборах 4 декабря 2011 года. Используйте из предложенного массива в качестве отклика вопрос со следующей формулировкой: q29 — Принимали ли вы участие в выборах в Государственную Думу России 4 декабря этого года?

В рамках домашнего задания 2 Вы уже оценивали логит-модель с самостоятельно выбранным набором предикторов. Используйте эту модель как основную для выполнения последующих заданий.

1. Проверьте гипотезу о согласии наблюдаемых значений (данных) и модели посредством Hosmer-Lemeshow goodness-of-fit test. Кратко поясните, каким образом устроена статистика теста (общий смысл), сделайте вывод на основании полученных результатов.
2. Рассчитайте значение baseline accuracy.
3. Сохраните предсказанные вероятности участия в выборах 4 декабря 2011 г., задайте сами порог отсечения и представьте в качестве результата confusion matrix. По представленной таблице классификации рассчитайте ошибку первого рода, ошибку второго рода и мощность (запишите в явном виде, как рассчитываются эти значения, объясните, что они содержательно показывают в контексте поставленной содержательной задачи в этом домашнем задании). Можно ли говорить о значимых различиях по сравнению с baseline accuracy? Прокомментируйте результаты.
4. Сравните решение (выбранный Вами порог отсечения) с классификацией наблюдений в соответствии с
 - (a) оптимальным порогом, выбранным на основе минимизации ошибки классификации
 - (b) оптимальным порогом, выбранным в соответствии с Youden index (максимизация $Sensitivity + Specificity - 1$)

Примечание: сравните по ряду мер: accuracy, специфичность, чувствительность. Сделайте вывод: какая модель из представленных, на Ваш взгляд, наиболее удачная с точки зрения классификации?

5. Предложите самостоятельно две спецификации вложенных моделей. Протестируйте посредством информационных критериев, какая модель является предпочтительной.
6. Предложите модель, вложенную в Вашу. Запишите ее спецификацию. Проверьте гипотезу об отсутствии различий между более и менее экономными моделями критерием отношения правдоподобия. Сделайте статистический и содержательный выводы.

7. Для порядковой переменной проверьте гипотезу о том, что коэффициенты предикторах-категориях одновременно равны 0 (иными словами, нет статистически значимых различий в их эффекте по сравнению с базовой категорией). Используйте критерий Вальда. Сделайте статистический и содержательный вывод.