

Machine Learning, Winter 2022
Practice Assignment 5

Exercise 5-1 Energy Drink Commercials

A study was designed to compare Red Bull energy drink commercials. Each participant was shown the commercials, A and B, in random order and asked to select the better one. There were 140 women and 130 men who participated in the study. Commercial A was selected by 65 women and by 67 men.

- Find the odds of selecting Commercial A for the men. Do the same for the women.
- Find the log odds for the men and the log odds for the women choosing Commercial A.
- What is the estimated odds ratio for a male to choose Commercial A ($x = 1$) versus a female to choose Commercial A ($x = 0$)?

Exercise 5-2 Stock Options

Different kinds of companies compensate their key employees in different ways. Established companies may pay higher salaries, while new companies may offer stock options that will be valuable if the company succeeds. Do high-tech companies tend to offer stock options more often than other companies? One study looked at a random sample of 200 companies. Of these, 91 were listed in the Directory of Public High Technology Corporations, and 109 were not listed. Seventy-three of the high-tech companies and 75 of the non-high-tech companies offered incentive stock options to key employees.

- What proportion of the high-tech companies offer stock options to their key employees? What are the odds?
- What proportion of the non-high-tech companies offer stock options to their key employees? What are the odds?
- Find the odds ratio using the odds for the high-tech companies in the numerator. Interpret the result in a few sentences.
- Find the log odds for the high-tech firms. Do the same for the non-high-tech firms.
- Define an explanatory variable x to have the value 1 for high-tech firms and 0 for non-high-tech firms. For the logistic model, we set the log odds equal to $\beta_0 + \beta_1 x$. Find the estimates b_0 and b_1 for the parameters β_0 and β_1 .

Exercise 5-3 Regularization
Coding Question

The main aim of this exercise is to compare the effect of different types of regularization on the polynomial (cubic) regression models produced for the Advertising dataset.

For L1 regularization, use `LassoCV(eps=0.1, n_alpha=100, cv=5)`.

For L2 regularization, use `RidgeCV(alphas=(0.1, 1.0, 10.0), scoring='neg_mean_absolute_error')`.

For L1 & L2 combined regularization, use `ElasticNetCV(l1_ratio=[.1, .5, .7, .9, .95, .99, 1], tol=0.01)`.

You will need to output the **RMSE**, **coefficients** and **plot** the output for each.

NB: Do not forget to perform feature scaling on your data before giving it to the models.

EXTRA: how do different values of alpha affect the performance of LassoCV?