Exercises for exercise class 5b in MMS075, Feb 18, 2020

As we did not finish all exercises in exercise class 5a, we start with the three remaining ones.

NOTE: there was a mistake in the specification of exercise 6 from yesterday: all plus signs in the formulas in parts 6a and 6b should have been minus signs. (The formulas with plus signs are also correct and can be proved similarly, but they do not correspond to the desired properties of the odds and logit functions). This mistake has been corrected in the version provided below.

1. For a probability value $0 < p < 1$, the **odds** is defined as $p/(1-p)$, and the **logit** (or **log-odds**) is defined as the logarithm of the odds, i.e. as $\text{logit}(p)=\log(p/(1-p))$. Compute the odds and logit values for the following probabilities:
    a. p=0.1;
    b. p=0.3;
    c. p=0.5;
    d. p=0.7;
    e. p=0.9.

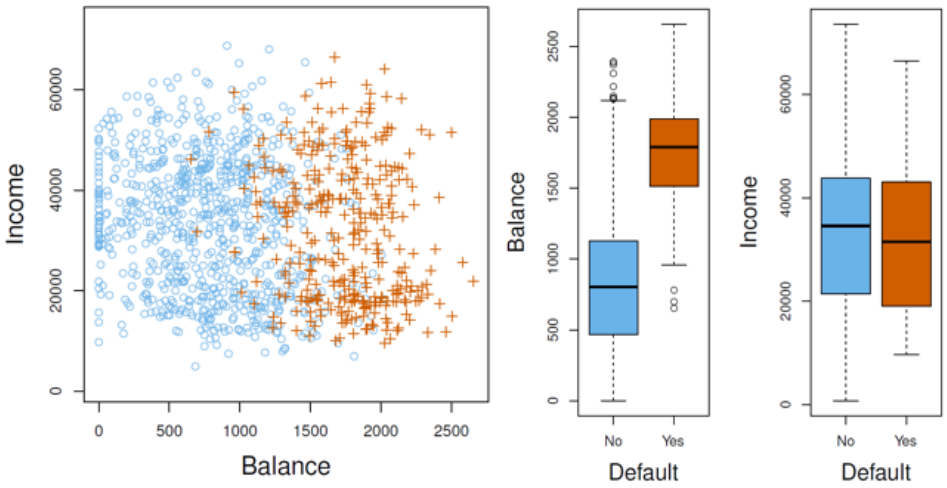
2. Show that the odds and logit functions are increasing in p, i.e., show that:
    a. if $p_2>p_1$, then $p_2/(1-p_2) > p_1/(1-p_1)$;
    b. if $p_2>p_1$, then $\log(p_2/(1-p_2)) > \log(p_1/(1-p_1))$!

3. Express the value of p from the value of the logit function; that is, assuming that we have $\log(p/(1-p)) = y$, express p as a function of y!

4. Consider the logistic regression model predicting the probability of credit card default based on balance, as discussed in the lecture:
$$\log\left(\frac{\hat{p}(X)}{1-\hat{p}(X)}\right) = \hat{\beta}_0 + \hat{\beta}_1 \times \text{balance}$$
Instead of the observed value of 0.0055, describe a correct interpretation of the following coefficient estimates in terms of log-odds, odds and probability:
    a. 1.1;
    b. -1.1;
    c. 0.

5. Consider the same logistic regression model as in part 4, this time with the correct coefficient estimates. We have seen in the lecture that the probability of default can be estimated from this model as follows:
$$\hat{p}(X) = \frac{e^{-10.65+0.0055\times\text{balance}}}{1+e^{-10.65+0.0055\times\text{balance}}}$$
Using this formula, estimate the probability of default for the following balance values:
    a. $1500;
    b. $2000;
    c. $2500.
Which balance value gives the highest probability of default? Could you answer this question before computing the estimates?

6. Relate the results of exercise 5 to Figure 4.1 in ISL, see below. Which plot(s) would help most in correctly anticipating the results before doing the computation?



7. Feedback quiz (optional): Go to www.menti.com and use the code 32 21 69.