STATG006: Solutions to Exercise Sheet #5

The exercises in this sheet focus on linear regression. As before, we provide solutions, sometimes detailed, sometimes a sketch that should point you to the complete solution. Skeetches should not be taken at face value as the level of detail required for an exam answer.

1. I assume this will not be a challenge at all. This exercise provides a further understanding of the role of the parameters.

2. (a) The linear response is $\eta = -6 + 0.05 \times 40 + 1 \times 3.5 = -0.5$. The probability is then given by $1/(1 + e^{-\eta}) \approx 0.38$.

   (b) In this case, we need to solve the system

$$0.5 = \frac{1}{1 + e^{-(-6+0.05x_1+3.5)}},$$

   or the easier and equivalent

$$\log \frac{0.5}{0.5} = -6 + 0.05x_1 + 3.5.$$

3. Let us define Y as a Bernoulli random variable, where $Y = 1$ encodes the event "a stock will issue a dividend", with $Y = 0$ encoding "a stock will *not* issue a dividend". Notation here will get a bit clumsy, as we combine density functions and probability mass functions. I will use $p_Y(\cdot)$ for the (conditional) probability Y taking some value, and $p_X(\cdot)$ for the (conditional) density of X. Bayes' theorem allows us to combine densities and mass functions.

   By the problem description, the estimated model is

$$\begin{aligned} p_Y(1) &= 0.8 \\ p_Y(0) &= 0.2 \\ p_X(x \mid Y = 1) &= \tfrac{1}{\sqrt{72\pi}} e^{-(x-10)^2/72} \\ p_X(x \mid Y = 0) &= \tfrac{1}{\sqrt{72\pi}} e^{-x^2/72} \end{aligned}$$

   By Bayes' theorem,

$$P(Y = 1 \,|X = x) = \frac{p_X(x \mid Y = 1)p_Y(1)}{p_X(x)} = \frac{p_X(x \mid Y = 1)p_Y(1)}{p_X(x \mid Y = 1)p_Y(1) + p_X(x \mid Y = 0)p_Y(0)}.$$

Plugging in $x = 4$ in the above will get us the answer. The lesson here is that we derive a different model for the conditional probability of a Bernoulli given a covariate that is not a generalised linear model (even though we can show that it boils down also to a class of functions close to logistic regression). The issue here is it also needs a model for the distribution of the covariates. In machine learning, this is known as a **generative approach**. See Section 4.4 of ISLR for a in-depth discussion of this approach in the context of binary classification.

4. This exercise further exploits your understanding of odds. In (a), assuming $Y = 1$ encodes the event of defaulting, it boils down to assessing

$$\frac{P(Y = 1)}{1 - P(Y = 1)} = 0.37,$$

and solving for "$P(Y = 1)$" as the unknown.

For (b), it is the opposite: just calculate $0.16/(1 - 0.16)$.

5. See EX5.R.

6. See EX5.R.

7. See EX5.R.

8. See EX5.R.

9. See EX5.R.