5: Logistic Regression

```
$ echo "Data Science Institute"
```

Logistic Regression

Logistic regression models the probability that the response Y belongs to a particular category. Suppose we have a qualitative response Y that has two levels, coded as 0 and 1, and one predictor variable. We want to model

$$p(X) = \Pr(Y = 1 \mid X)$$

The logistic function keeps the probabilities between 0 and 1. For one predictor, the function is

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$ As with linear regression, we are trying to fit β_0, β_1.

Estimating the regression coefficients

β_0 and β_1 are estimated using the training data using a method called **maximum likelihood**. This involves maximizing the likelihood function, but we will not cover the details of this function.

Odds

The **odds** compares the probability of a particular outcome to the probability of all the other outcomes.

$$\frac{p(X)}{1-p(X)} = e^{\beta_0 + \beta_1 X}$$

- takes values between $(0, \infty)$

- odds close to $0 \Rightarrow$ very low probability of the outcome in question

- odds much greater than $0 \Rightarrow$ very high probability of the outcome in question.

Log Odds

The **log odds** (or logit) is obtained by taking the logarithm of the odds

$$\log \left(\frac{p(X)}{1-p(X)} \right) = \beta_0 + \beta_1 X$$

- Increasing X by one unit changes the log odds by β_1.

- If β_1 is positive, increasing X is associated with increasing $p(X)$

- If β_1 is negative, increasing X is associated with decreasing $p(X)$

Making Predictions

Once the coefficients have been estimated predictions can be made for any value of the predictor. Logistic regression will give the probability of the outcome and the classification will be according to some threshold which depends on the problem or how conservative the predictions should be.

Multiple Predictors

Simple logistic regression can be extended to include multiple predictors

$$p(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}}$$

The log odds in this case becomes

$$\log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

As before, the maximum likelihood is used to estimate the coefficients.

Exercise: Logistic Regression

Open the Classification Exercises R Markdown or Jupyter Notebook file.

- Go over "Getting Started" together as a class.

- Go through the "Logistic Regression" as a class.

- 5 minutes for students to complete the questions from "Logistic Regression".

- Questions should be completed at home if time does not allow.

Multinomial logistic regression

We can extend to two-class logistic regression to accommodate K classes. We need to select one class to serve as the **baseline**, so we will choose the Kth class. Then the model becomes

$$\Pr(Y = K \mid X = x) = \frac{1}{1 + \sum_{l=1}^{K-1} e^{\beta_{l0} + \beta_{l1} x_1 + \cdots + \beta_{lp} x_p}}$$

and,

$$\Pr(Y = k \mid X = x) = \frac{e^{\beta_{k0} + \beta_{k1} x_1 + \cdots + \beta_{kp} x_p}}{1 + \sum_{l=1}^{K-1} e^{\beta_{l0} + \beta_{l1} x_1 + \cdots + \beta_{lp} x_p}} \text{ for } k = 1, \ldots, K - 1$$

The interpretation of the coefficients is tied to the choice of the baseline.

References

Chapter 4 and section 2.2.3 of the ISLP book:

James, Gareth, et al. "Classification." An Introduction to Statistical Learning: with Applications in Python, Springer, 2023.