COMP6237 – Tutorial

Logistic Regression Problem Sets

Shoaib Ehsan

s.ehsan@soton.ac.uk

Aims of the Tutorial

1. Reinforce the fundamental concepts of **logistic regression**

2. Enhance students' ability to apply logistic regression to real-world datasets.

Today’s Tutorial

- Quick Recap of Logistic Regression
- Example from lecture
- Problem 1
- Problem 2

Quick Recap of Logistic Regression

- Logistic Regression model is a classification model
- Instead of predicting a continuous value like linear regression, logistic regression predicts the probability of an outcome using the sigmoid function:

$$p(x) = \frac{1}{1 + \exp(-(b + wx))}$$

- This ensures that the output is always between 0 and 1, which can be interpreted as a probability.

Example from Lecture

What is the chance of a film with Box Office takings of $50 million to win an Oscar? The parameters for logistic regression model are -1.75 (intercept) and 0.011.

Solution

- The equation for logistic regression is:

$$p(x) = \frac{1}{1 + \exp(-(b + wx))}$$

Here,

$$b = \text{-}1.75$$

$$w = 0.011$$

$$x = 50$$

Solution

- Putting values in the logistic regression equation:

$$p(x) = \frac{1}{1 + \exp\left(-(-1.75 + 0.011 * 50)\right)}$$

$$p(x) = \frac{1}{1 + \exp\left(-(-1.2)\right)}$$

$$p(x) = \frac{1}{4.32} = 0.231$$

The model predicts that such a film has a 23% chance to win an Oscar.

Problem 1

A data set has been collected to relate the age of a learner to the outcome of driving tests. Carrying out logistic regression, somebody obtains a slope of $w = 0.01$ and an intercept of $b = 0.1$. What are the chances of a 100 years old applicant to pass the test?

Solution

- The equation for logistic regression is:

$$p(x) = \frac{1}{1 + \exp(-(b + wx))}$$

Here,

$$b = 0.1$$

$$w = 0.01$$

$$x = 100$$

Solution

- Putting values in the logistic regression equation:

$$p(x) = \frac{1}{1 + \exp\left(-(0.1 + 0.01 * 100)\right)}$$

$$p(x) = \frac{1}{1 + \exp\left(-(1.1)\right)}$$

$$p(x) = \frac{1}{1.3329} = 0.7502$$

The model predicts that a 100 years old applicant has a 75% chance to pass the driving test.

Problem 2

Somebody collects a data set to analyze examination outcomes (discriminating between fail, pass, and repeat) of students on a three-year BSc degree and carries out multinomial logistic regression to predict the outcome dependent on the year of study. Results give: (i) intercept (fail) = 1, slope (fail) = -1 and (ii) intercept (pass) = 3, slope (pass) = -1/2. What is the chance of a student having to repeat the 3rd year?

Solution

- We are given a multinomial logistic regression model that predicts the probability of three possible examination outcomes (Fail, Pass, Repeat) based on the year of study.
- Let the year of study be denoted by $x = 3$. The regression coefficients are only given for Fail and Pass. Hence, Repeat was the reference category and we have:

$$\log \frac{p_{Fail}}{p_{Repeat}} = 1 - x$$

$$\log \frac{p_{Pass}}{p_{Repeat}} = 3 - \frac{1}{2}x$$

Solution

- This can also be written as:

$$p_{Fail} = p_{Repeat}\, e^{1-x}$$

$$p_{Pass} = p_{Repeat}\, e^{3-\frac{1}{2}x}$$

- As

$$p_{Repeat} + p_{Pass} + p_{Fail} = 1$$

$$p_{Repeat} + p_{Repeat}\, e^{3-\frac{1}{2}x} + p_{Repeat}\, e^{1-x} = 1$$

Solution

$$p_{Repeat} + p_{Repeat}\, e^{3-\frac{1}{2}x} + p_{Repeat}\, e^{1-x} = 1$$

$$p_{Repeat}(1 + \ e^{3-\frac{1}{2}x} + \ e^{1-x}) = 1$$

$$p_{Repeat} = \frac{1}{1 + \ e^{3-\frac{1}{2}x} + \ e^{1-x}}$$

$$p_{Repeat} = \frac{1}{1 + \ e^{3-\frac{1}{2}(3)} + \ e^{1-3}} = \frac{1}{1 + 4.4817 + 0.1353} = 0.178$$

Thus, chances for an examination outcome of repeat for a student in the 3rd year are around 0.178. This means 17.8%.