Odds and Logistic Regression

1 Odds

Definition: Odds

A numerical expression, expressed as a pair of numbers

Important: Odds vs Probabilities

Odds are not probabilities, odds is the number of successes vs number of failures, whereas probabilities is number of successes against total number of cases

$$\frac{\text{Probabilitiy(success)}}{\text{Probability(failure)}} = \text{Odds}$$

This is the same as

$$\frac{\text{Probabilitiy(success)}}{1 - \text{Probability(success)}} = \text{Odds}$$

Convention is that p represents the probability of success and q represents the probability of failure

1.1 Log of odds

Where the number of successes > number of failures:

- Odds against success is between 0 and 1
- Odds in favour of success is between 1 and $+\infty$

This makes the odds against success look way smaller, so take the logs to make everything symmetrical

Note here that $\log(\frac{x}{y}) = -\log(\frac{y}{x})$

1.2 Odds ratios

Definition: Odds Ratio

Comparing two different odds by dividing them

This has the same problem that needs solving with logs as with a single odd

Larger odds ratio mean that one thing is a good predictor of another.

2 Logistic Regression

This is similar to linear regression except it predicts true or false

It follows an s shape, fitting to false for low numbers and true for high numbers, although can be adapted to be reversed.

Logistic regression provides probabilities and classifies new samples using continuous and discrete measurements

We use maximum likelihood to draw the regression line in the log graph

$$\log(\frac{p}{1-p}) = \log(odds)$$

$$\frac{p}{1-p} = e^{\log(odds)}$$

$$p = (1-p)e^{\log(odds)} = e^{\log(odds)} - p^{e\log(odds)}$$

$$p + pe^{\log(odds)} = e^{\log(odds)}$$

$$p(1 + e^{\log(odds)}) = e^{\log(odds)}$$

$$p = \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$

To then calculate the probability of success, multiply the probabilities of the points on the logistic regression line

We can use these probabilities to calculate the likelihood, which can then be used to improve the fit of the line.