STA 3180 Statistical Modelling: Logistic Regression

Logistic Regression

Logistic regression is a type of statistical modelling used to predict the probability of a binary outcome. It is used to model the relationship between one or more independent variables (predictors) and a binary dependent variable (outcome). Logistic regression is a powerful tool for predicting the probability of an event based on the values of one or more predictor variables.

Key Concepts:

• Binary Outcome: A binary outcome is a result that can only have two possible values, such as yes/no, true/false, or 0/1.

• Odds Ratio: The odds ratio is a measure of association between two binary variables. It is the ratio of the odds of an event occurring in one group to the odds of it occurring in another group.

• Logit Function: The logit function is a mathematical function used to transform a continuous variable into a binary outcome. It is used in logistic regression to transform the linear relationship between the independent variables and the dependent variable into a logistic curve.

• Maximum Likelihood Estimation: Maximum likelihood estimation (MLE) is a method of estimating the parameters of a model from a set of data. In logistic regression, MLE is used to estimate the coefficients of the logistic regression equation.

Definitions:

• Logistic Regression: Logistic regression is a type of statistical modelling used to predict the probability of a binary outcome. It is used to model the relationship between one or more independent variables (predictors) and a binary dependent variable (outcome).

• Coefficient: In logistic regression, a coefficient is a numerical value that represents the strength of the relationship between a predictor variable and the outcome variable.

• Intercept: The intercept is the point at which the logistic regression line crosses the y-axis. It is the estimated probability of the outcome when all predictor variables are equal to zero.

Rules:

• Logistic regression is used to predict the probability of a binary outcome.

• The independent variables must be linearly related to the dependent variable.

• The independent variables must be independent of each other.

• The dependent variable must be binary.

• The coefficients of the logistic regression equation must be estimated using maximum likelihood estimation.

Examples:

• Logistic regression can be used to predict the probability of a student passing an exam based on their performance on previous exams.

• Logistic regression can be used to predict the probability of a customer making a purchase based on their past purchases.

• Logistic regression can be used to predict the probability of a patient developing a disease based on their medical history.