Assignment 2

Muhammed Saeed and Ali Salaheldin Ali Ahmed

November 2022

Problem 1

1.1.

Logistical regression is applicable in the classification setting. In classification, we want to predict a categorical response variable Y for an observation of one or more predictor variables $(X_1, X_2, ..., X_p)$, where each one can be either quantitative or qualitative. We use logistical regression to estimate the probability that the observation belongs to each of the classes.

Linear regression in this setting suffers from a number of problems:

- *Linear regression cannot accommodate a qualitative response with more than two classes.*

  In order to use linear regression, we need to have numeric response values, thus we encode the classes as numbers. If we have more than two classes, such encodings unnecessarily add ordering, which is often not meaningful in practice. This has a significant effect on results of the linear regression, where each numeric encoding would result in a different relationship.

- *Linear regression will not provide meaningful estimates of Pr(Y —X).*

  Even for binary classification where the previous issue has limited effect, linear regression can produce estimate values outside the $[0, 1]$ interval, which are hard to interpret as probabilities.

1.2.

logistic regression models the probability that Y belongs to a particular category. The relationship between the predicted probabilities and the independent variables is given by the equation:

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

The bias term β_0 is responsible for shifting the decision boundaries given that we are satisfied with the classifier slope, depending on whether we are using uni-variant or multi-variant we have parameters from β_1 up to β_n all those parameters help to build the linear classifier or the QDA classifier.

1.3.

Odds are the ratio between probabilities: the probability of the observation belonging to a class, and the probability of that not being the case. Probability is in the interval $[0, 1]$, and odds are in the interval $[0, \infty)$.

$$\begin{aligned}
\text{odds} &= \frac{p(X)}{1 - p(X)} = e^{\beta_0 + \beta_1 * X} \\
p(X) &= \frac{\text{odds}}{1 + \text{odds}}
\end{aligned}$$

where p(X) represents the logistic function. That is, in logistic regression we estimate and transform the odds into a probability value.

1.4.

Starting from the formula for the logistic function.

$$\begin{aligned}
p(X) &= \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \\
p(X)(1 + e^{\beta_0 + \beta_1 X}) &= e^{\beta_0 + \beta_1 X} \\
p(X) &= e^{\beta_0 + \beta_1 X} - p(X) e^{\beta_0 + \beta_1 X} \\
\frac{p(X)}{1 - p(X)} &= e^{\beta_0 + \beta_1 X} \\
\log \frac{p(X)}{1 - p(X)} &= \beta_0 + \beta_1 X
\end{aligned}$$

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

The logistic function can be expressed in terms of the logit function.

$$p(X) = \frac{e^{\text{logit}(X)}}{1 + e^{\text{logit}(X)}}$$

The logistic function maps the values of the logit function into the interval $[0, 1]$, which can be readily interpreted as a probability value. Due to this, it is important to understand this relationship in order to interpret the relationship between the probability and the predictors.

1.5

$$
\begin{aligned}
P(Y_\theta = 1) &= \frac{e^\theta}{1+e^\theta} && \text{where } \theta \in R \\
odds(Y_\theta) &= \frac{P(Y_\theta)}{1 - P(Y_{\theta)}} \\
odds(Y_\theta) &= e^\theta \\
\frac{odds(Y_{x^T\beta+\beta_i\delta})}{odds(Y_{x^T.\beta})} &= \frac{e^{x^T.\beta+\beta_i\delta}}{e^{x^T.\beta}} \\
\frac{odds(Y_{x^T\beta+\beta_i\delta})}{odds(Y_{x^T\beta})} &= e^{x^T\beta+\beta_i\delta-x^T\beta} \\
\frac{odds(Y_{x^T\beta+\beta_i\delta})}{odds(Y_{x^T\beta})} &= e^{\beta_i\delta}
\end{aligned}
$$

This result means that an increment of δ units in the value of the predictor x_i leads to an increase in the odds (of $Y_\theta = 1$) by a magnitude of $e^{\beta_i\delta}$.