Elements of Machine Learning

Assigment 2 - Problem 1

Sangeet Sagar(7009050), Philipp Schuhmacher(7010127)
{sasa00001,phsc00003}@stud.uni-saarland.de

December 1, 2021

Problem 1 (T, 10 Points) Logistic regression

1

(4P) In which setting is logistic regression applicable? Explain at least three problems of linear regression is applied in such a setting. In which setting is linear regression applicable?
Logistic regression is applicable when the dependent variable is binary, i.e. 1 or 0, spam or not spam, fail or pass.

1. Linear regression has continuous variables with continuous values, i.e. can range outside 1 or beyond 0. This can be a problem because, in logistic regression, the output probability must lie between 0 and 1, and a value outside this range can not be interpreted to predict the probability of a class.
2. Linear regression fails miserably when trying to accommodate qualitative responses with more than 2 responses. E.g. predicting the response of a news headline as sports, business or politics using linear regression would depend on the order we are encoding this response. That is to say; each order will be modelled by linear regression differently.
3. It is assumed in linear regression that ϵ (noise in the data) is normal distributed. This assumption is violated because in a logistic regression setting the observations $y \in 0, 1$ follow Bernoulli distribution.

Linear regression could be applicable for a binary qualitative response when the output of the estimate is $[0, 1]$. E.g. $Y = 1$ when a student fails and $Y = 0$ when the student passes. We could predict that for $\hat{Y} < 0.5$, the student passes otherwise fails.

2

(1P) What do we model with logistic regression? How are the independent variables and obtained probabilities related?
We model the probability of a binary event or a class with logistic regression.
The independent variable and the probabilities are non-linearly related using the so-called S-shaped *logistic function* which constraints the probabilities on $(0, 1)$.

3

(1P) In general, what is the meaning of odds? Write down the formula and explain it in your own words. How do odds relate to logistic regression?
Odds are a ratio of the probability of success to the probability of failure. E.g. in horse race betting, if we say the odds of wins are 3 to 2. We can interpret this as for every 5 bets, on average 3 bets are in favour, and 2 bets are against.
Odds have a range between $(0, \infty)$.
Formula

$$Odds = \frac{p(X)}{1 - p(X)}$$

So $p(x)$ is the probability of success, and $1 - p(x)$ is the probability of failure. Considering the same example above, let's consider 3 cases:

1. Odds=3/2. 3 bets in favor and 2 bets against. So, $\frac{p(X)}{1-p(X)} = 3/2$ which in turn gives $p(x) = 3/5$. Hence the odds of success are 1.5. But whats the odds of failures? Simply reciprocal of odds of success i.e 2/3 or 0.67 are the odds of failures (we basically mean to say that on an average 2 out 5 bets are against.).

2. Odds=0. 0 bets in favor and 5 bets against. So, $\frac{p(X)}{1-p(X)} = 0/5$ or $p(x) = 0$

3. Odds=∞. All bets in favor. So, $\frac{p(X)}{1-p(X)} = 5/0$ or $p(x) = 1$

Probability is defined on $(0, 1)$ and odds range between to $(0, \infty)$. While log odds or logit has an advantage that it ranges between $(-\infty, \infty)$ i.e they can take all possible values. Thus, log odds hold a linear relationship with X. This means that in a logistic regression model, one unit increase in X is represented as change in log odds by the regression coefficient β_1. This is how odds are related to logistic regression.

4

(2P) Prove that Equation 4.2 (ISLR p. 134)

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

is equivalent to Equation 4.4 (ISLR p. 135)

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

Starting with the given equation,

$$\begin{aligned}
p(X) &= \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}} \\
p(X) + p(X) \times e^{\beta_0+\beta_1 X} &= e^{\beta_0+\beta_1 X} \\
p(X) &= e^{\beta_0+\beta_1 X}\,(1 - p(X)) \\
\frac{p(X)}{1-p(X)} &= e^{\beta_0+\beta_1 X} \\
\text{Now take log on both sides} & \\
\log\left(\frac{p(X)}{1-p(X)}\right) &= \log\left(e^{\beta_0+\beta_1 X}\right) \\
\log\left(\frac{p(X)}{1-p(X)}\right) &= \beta_0 + \beta_1 X
\end{aligned}$$

5

(1P) Show the equality:

$$\frac{odd(X_i + \Delta)}{odd(X_i)} = exp(\beta_i \Delta)$$

where $\Delta \in \mathbb{R}$. Explain the meaning of this equality in your own words.
From the last question we have the following, which we will be using in this part.

$$\begin{aligned}
\log\left(\frac{p(X)}{1-p(X)}\right) &= \beta_0 + \beta_1 X \\
\log\left(odd(X_i)\right) &= \beta_0 + \beta_i X_i
\end{aligned}$$

Now starting with whats given

$$\frac{odd(X_i + \Delta)}{odd(X_i)} = exp(\beta_i \Delta)$$

$$\text{Take log on both sides}$$

$$\log\left(\frac{odd(X_i + \Delta)}{odd(X_i)}\right) = \log\left(exp(\beta_i \Delta)\right)$$

$$\log\left(odd(X_i + \Delta)\right) - \log\left(odd(X_i)\right) = \beta_i \Delta$$

$$\text{From the above equation}$$

$$\beta_0 + \beta_i(X_i + \Delta) - \beta_0 - \beta_i(X_i) = \beta_i \Delta$$

$$\beta_i \Delta = \beta_i \Delta$$

Therefore LHS = RHS and the equality holds for $\Delta \in \mathbb{R}$.
This means that increasing by X by Δ units lead to change in the log odds by $exp(\beta_1 \Delta)$ units.

6

(1P) Assume the number of features to be $p = 1$, i.e.,

$$Pr(Y = 1|X) = p(X) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}$$

How do we have to choose X if we want to have $p(X) = 0.5$. What does this probability tell us?

$$\begin{aligned}
&\implies \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)} = 0.5 \\
&\implies 2\exp(\beta_0 + \beta_1 X) = 1 + \exp(\beta_0 + \beta_1 X) \\
&\implies \exp(\beta_0 + \beta_1 X) = 1 \\
&\implies \beta_0 + \beta_1 X = 0 \\
&\implies X = \frac{-\beta_0}{\beta_1}
\end{aligned}$$

$p(X) = 0.5$ acts as a threshold value in a logistic regression. If for any variable $p(x) > 0.5$, the logistic function pushes the resulting prediction towards 1 and towards 0 for $p(x) < 0.5$.