Elements of Machine Learning
Assignment 2

William LaCroix (7038732)
wila00001@stud.uni-saarland.de

Problem 1 (T, 8 Points). **Logistic Regression**

1. [4pts] In which setting is logistic regression applicable? Explain at least three problems with linear regression when applied in such a setting.

Logistic regression tends to perform well in classification settings where the true decision boundary is linear. Problems with linear regression in these settings include:

- Linear regression describes continuous outcome values, and isn't necessarily satisfactory when our outcomes are discrete variables (fractional riders)
- Linear regression, unless it's slope-0, will predict negative outcomes for some range of predictors (negative riders).
- Similarly, linear regression will predict probabilities p(x) < 0 and p(X) > 1, which don't have a logical interpretation.
- Linear regression also performs poorly with heteroskedastic relationships, as in settings where the variance depends heavily on certain predictors (season / weather, for instance).
- Linear regression with more than one outcome class implies and ordered relationship between outcomes that may not have a logical relationship (stroke vs overdose vs seizure)

2. [1pt] What do we model with logistic regression? How are the independent variables and obtained probabilities related?

Logistic regression models the probability that a response belongs to a particular category, as opposed to measuring the response directly. We do this with a logistic function, mapping predictors onto a response range between 0 and 1, and each unit increase in the predictor X_1 is associated with a log-odds increase in the response by our parameter coefficient β_1.

3. [1pt] In general, what is the meaning of odds? Write down the formula and explain in your own words. How do odds relate to logistic regression?]

Odds is an estimation of likelihood, with low odds meaning something is unlikely (low probability), and high odds meaning something is more likely (high probability).

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

As odds go to zero, p(X) tends toward zero, and as odds approach infinity, p(X) tends toward 1. Thus odds can be used to provide estimations of the probability of an outcome, given predictors, which is exactly the goal of logistic regression.

4. [1pt] Let X be a scalar random variable. Prove that

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

The standard logistic function says that $f\left(x\right) = \frac{e^x}{1+e^x}$. If we begin with our linear regression model of, then $p\left(X\right) = \beta_0 + \beta_1 X = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$, which is just the left side of our implication. Let $p\left(X\right)$ be represented as y, and $e^{\beta_0+\beta_1 X}$ as x. Through simple algebra, we may rewrite:
$[y = \frac{x}{1+x}] \equiv [(1+x)\, y = x] \equiv [y + xy = x] \equiv [y = x - xy] \equiv [y = x\left(1-y\right)] \equiv [x = \frac{y}{1-y}]$

Substituting x and y back in, we get $e^{\beta_0+\beta_1 X} = \frac{p(X)}{1-p(X)}$. Taking the log of both sides gives us $log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$, which is the right side of our implication.

What is the relationship between the logistic and the logit function? Why is this information about the relationship important? Explain.

As seen in the above proof, the logistic function allows us to transform our linear regression, which is linear in X, and map it onto the logit function, which gives us a non-linear probability distribution curve between 0 and 1, that is nonetheless still linear in X. This gives us the ability map the full range of predictors onto a probability distribution which has salient interpretations across its range (as opposed to the below-zero probabilities predicted by simple linear regression)

5. [1pt] Let Y be a binary random variable for which

$$\mathbb{P}\left(Y_\Theta = 1\right) = \frac{e^\Theta}{1+e^\Theta}, \text{ where } \Theta \in \mathbb{R}$$

and define a parameter vector $\beta = [\beta_0, ..., \beta_p]$ and feature vector $\mathbf{x} = [1, x_1, ..., x_p]$. Show that

$$\frac{odds\left(\mathbf{Y}_{\mathbf{x}^\top \beta + \beta_i \delta}\right)}{odds\left(\mathbf{Y}_{\mathbf{x}^\top \beta}\right)} = exp\left(\beta_i \delta\right), \text{ for some } \delta \in \mathbb{R} \text{ and any } i \in \{1, ..., p\}$$

and explain the meaning of this equality in your own words.

The ratio of the odds of Y given some δ increase in parameter β_i is equivalent to e raised to the δ increase in parameter β_i. In other words, the log odds increase in the variable Y is equal to the δ increase in parameter β_i.