Assignment #2

Elements of Machine Learning

Saarland University – Winter Semester 2024/25

Rabin Adhikari
7072310
raad00002@stud.uni-saarland.de

Dhimitrios Duka
7059153
dhdu00001@stud.uni-saarland.de

Problem 1 (Introduction to Logistic Regression)

a) From the exercise description, we know that we only have a single feature $X \in \mathbb{R}$ as the input to the model $f(X)$. This means the $f(X)$ would have the form:

$$f(X) = X\beta_1 + \beta_0 \tag{1}$$

where β_0 is the intercept and β_1 is the slope of the line defined by $f(X)$.

From the lecture, we know that the logistic function is defined as:

$$p(Y = 1|X) = \frac{e^{f(X)}}{1 + e^{f(X)}} \tag{2}$$

Substituting Equation 1 in Equation 2, we get the following:

$$p(Y = 1|X) = \frac{e^{X\beta_1 + \beta_0}}{1 + e^{X\beta_1 + \beta_0}} \tag{3}$$

Equation 3 represents the logistic regression function given a single input $X \in \mathbb{R}$. The output of this function lies within [0, 1]. More formally, $p(Y = 1|X) \in [0, 1]$ for all X. As a consequence, the output of the logistic regression function can be interpreted as the probability of an input X belonging to class 1.

b) In the context of logistic regression, the likelihood function is defined as the probability of observing the output $y_1, \ldots, y_n$ given the input $x_1, \ldots, x_n$. The likelihood function is given by:

$$p(y_1, \ldots, y_n|x_1, \ldots, x_n) = \prod_{i:y_i=1} p(y_i = 1|x_i) \prod_{i:y_i=0} (1 - p(y_i = 1|x_i)) \tag{4}$$

Log-likelihood is essentialy the logarithm of the likelihood function. It is used to simplify the optimization problem. The log-likelihood function is given by:

$$\log(p(y_1, \ldots, y_n|x_1, \ldots, x_n)) = \log \left[\prod_{i:y_i=1} p(y_i = 1|x_i) \prod_{i:y_i=0} (1 - p(y_i = 1|x_i)) \right] \tag{5}$$

Given that $\log(xy) = \log(x) + \log(y)$, we can rewrite Equation 5 as:

$$\begin{aligned}\log(p(y_1, \ldots, y_n|x_1, \ldots, x_n)) &= \log\left[\prod_{i:y_i=1} p(y_i = 1|x_i)\right] + \log\left[\prod_{i:y_i=0} (1 - p(y_i = 1|x_i))\right] \\ &= \sum_{i:y_i=1} \log(p(y_i = 1|x_i)) + \sum_{i:y_i=0} \log(1 - p(y_i = 1|x_i))\end{aligned} \tag{6}$$

Equation 6 expresses the log-likelihood function. To tailor it for logistic regression, we just need to substitute $p(y_i = 1|x_i) = \frac{e^{x_i\beta_1+\beta_0}}{1+e^{x_i\beta_1+\beta_0}}$. Doing so, we get the following result:

$$\log(p(y_1, \ldots, y_n|x_1, \ldots, x_n)) = \sum_{i:y_i=1} \log(\frac{e^{x_i\beta_1+\beta_0}}{1 + e^{x_i\beta_1+\beta_0}}) + \sum_{i:y_i=0} \log(1 - \frac{e^{x_i\beta_1+\beta_0}}{1 + e^{x_i\beta_1+\beta_0}}) \tag{7}$$

The goal is to maximize the log-likelihood function. However, this problem doesn't have a closed-form solution. Therefore, we have to rely on optimization methods such as the Newton-Raphson method to estimate the parameters β_0 and β_1.

c) Frist we will compare the two types of classifiers in terms of the output. Given an input x, the discriminative classifier estimates $\hat{g(x)}$ of class $g(x)$, while the generative classifier outputs a probability distribution $p_g(x)|g \in G$, where $p_g(x)$ is the probability that x belongs to class g.

Next, we will compare the two types of classifiers in terms of the loss function. The discriminative classifier measures the deviation between the estimates and the output, while the generative classifier measures the (log-)likelihood of the estimator generating the output $\sum_{i=1}^{N} \log p_{g_i}(x)$.

Finally, we will compare the two types of classifiers in terms of optimization. The discriminative classifier aims to minimize the loss function, while the generative classifier aims to maximize the likelihood of the data.