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 2 (Logistic Regression)

a) Deriving the gradient of the logistic regression loss function w.r.t. the coefficients β can be done as follows:

$$\begin{aligned}
\frac{\partial}{\partial \beta_j}\ell(\beta) &= \frac{\partial}{\partial \beta_j}\sum_{i=1}^{n}\left[y_i \log p(x_i;\beta) + (1-y_i)\log\left(1-p(x_i;\beta)\right)\right] \\
&= \sum_{i=1}^{n}\left[y_i \frac{\partial}{\partial \beta_j}\log p(x_i;\beta) + (1-y_i)\frac{\partial}{\partial \beta_j}\log\left(1-p(x_i;\beta)\right)\right] \\
&= \sum_{i=1}^{n}\left[y_i \frac{\frac{\partial}{\partial \beta_j}p(x_i;\beta)}{p(x_i;\beta)} + (1-y_i)\frac{\frac{\partial}{\partial \beta_j}(1-p(x_i;\beta))}{(1-p(x_i;\beta))}\right] \\
&= \sum_{i=1}^{n}\left[y_i \frac{\frac{\partial}{\partial \beta_j}p(x_i;\beta)}{p(x_i;\beta)} - (1-y_i)\frac{\frac{\partial}{\partial \beta_j}p(x_i;\beta)}{(1-p(x_i;\beta))}\right] \\
&= \sum_{i=1}^{n}\left[\frac{y_i}{p(x_i;\beta)}\frac{\partial}{\partial \beta_j}p(x_i;\beta) - \frac{1-y_i}{(1-p(x_i;\beta))}\frac{\partial}{\partial \beta_j}p(x_i;\beta)\right]
\end{aligned} \tag{1}$$

b) During the training process, we aim to minimize the log loss function. The log loss function is defined as follows:

$$\ell(\beta) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log p(x_i;\beta) + (1-y_i)\log\left(1-p(x_i;\beta)\right)\right] \tag{2}$$

To better understand how the log loss function beahves, we will examine two distinct cases. In the first case, we will consider the case where the true label is $y_i = 1$. In the second case, we will consider the case where the true label is $y_i = 0$.

Case 1: $y_i = 1$

In this case, the log loss function simplifies to:

$$\ell(\beta) = -\frac{1}{n}\sum_{i=1}^{n}\log p(x_i;\beta) \tag{3}$$

In order for this term to be minimized, we need the values of $p(x_i;\beta)$ to be as close to 1 as possible. This means that the model should be confident that the input x_i belongs to class 1, thus aligning with the true label $y_i = 1$.

Case 2: $y_i = 0$

In this case, the log loss function simplifies to:

$$\ell(\beta) = -\frac{1}{n}\sum_{i=1}^{n} \log\left(1 - p(x_i;\beta)\right) \tag{4}$$

In order for this term to be minimized, we need the values of $p(x_i;\beta)$ to be as close to 0 as possible. This means that the model should be confident that the input x_i belongs to class 0, thus aligning with the true label $y_i = 0$.

c) i) The outputs from the logistic regression model for the given data points are summarized in Table 1.

Table 1: Predictions for the given data points using the logistic regression model. GT: Ground Truth.

x_1	x_2	$p(x_i, \beta)$	**Prediction**	**GT**
1.0	2.0	0.182	0	0
2.0	3.0	0.378	0	0
3.0	4.0	0.622	1	0
4.0	5.0	0.818	1	1
5.0	6.0	0.924	1	1
6.0	7.0	0.971	1	1
7.0	8.0	0.989	1	1
8.0	9.0	0.996	1	1

c) ii) Given the threshold of 0.5, the predictions for the given data points are summarized in Table 1. The model missclassifies only one data point.