Report

Huanshuo Dong, Zhiwei Zhuang, Haoran Wang

2024/01/19

1 Introduction

Weather forecasting plays a vital role in all aspects of daily life. In this paper, we will train the training data to obtain the corresponding prediction model by the approach of logistics regression.

2 Preliminaries

Logistic regression is a statistical method used for binary classification problems. Unlike linear regression, logistic regression aims to predict the probability of a binary output variable. Its core is the logistic function (also known as the sigmoid function), which maps the linear combination of inputs to probability values in the range $(0, 1)$.

2.1 Mathematical Model

The basic form of the logistic regression model can be represented as:

$$P(Y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1^T x)}} \tag{1}$$

Here, x is the input vector, and β_0 and β_1 are the model parameters. We can also write the model in a slightly different form to compactly represent the parameters:

$$P(Y = 1|x) = (1 + \exp(-\beta^T \tilde{x}))^{-1} \tag{2}$$

Here, $\tilde{x} = [1, x]$ is an extended input vector, and β is the parameter vector that combines β_0 and β_1.

2.2 Loss Function

To train the logistic regression model, we need to define a loss function to measure the performance of the model. Typically, we use the logarithmic loss function (also called the log loss or cross-entropy loss):

$$L(\beta) = -\sum_{i=1}^{n}[y_i \log(P(Y=1|x_i)) + (1-y_i)\log(1-P(Y=1|x_i))] \quad (3)$$

2.3 Optimization Methods

Our goal is to find a set of parameters β that minimizes the loss function $L(\beta)$. Typically, we use gradient descent to achieve this objective. The basic idea of gradient descent is to update the parameters in the negative gradient direction of the loss function to gradually decrease its value.

The update rule of gradient descent is:

$$\beta := \beta - \alpha \nabla L(\beta) \quad (4)$$

Here, α is the learning rate, and $\nabla L(\beta)$ is the gradient of the loss function $L(\beta)$ with respect to the parameters β. The gradient can be computed using the following formula:

$$\nabla L(\beta) = -\sum_{i=1}^{n} x_i(y_i - P(Y=1|x_i)) \quad (5)$$

2.3.1 Step Size Adjustment

If the loss function increases, the learning rate (lr) should be reduced. If lr $<$ 1e-7, the iteration should be stopped, or if the loss continues to increase, the iteration should be stopped.

2.3.2 Using L1 or L2 Regularization

Mathematical Principles During the training process of logistic regression, we typically minimize the log-likelihood loss function. To control the complexity of the model, we can add a regularization term to the loss function.

The regularization term for L1 regularization is:

$$R_{L1}(\beta) = \lambda \sum_{i=1}^{p} |\beta_i| \quad (6)$$

The regularization term for L2 regularization is:

$$R_{L2}(\beta) = \lambda \sum_{i=1}^{p} \beta_i^2 \quad (7)$$

Here, β is the model parameter, p is the number of parameters, and λ is the hyperparameter that controls the strength of regularization.

L1 Regularization and Parameter Sparsity L1 regularization has an important property of inducing parameter sparsity, where many parameter values are pushed to zero. This is because the L1 norm is not differentiable at zero and the regularization path is nonlinear, resulting in parameter truncation near zero.

Purpose of Regularization The main purpose of regularization is to prevent overfitting and improve the generalization performance of the model.

3 Experimental procedure

3.1 Data cleaning

During the model training process, we have observed that the different datasets and input features have varying formats, and there is a certain amount of missing data. In the face of this phenomenon, it is necessary to perform data cleaning. For input features with a high amount of missing data, we choose to delete those features. For input features with a small amount of missing data, we adopt the method of averaging to fill in the missing information.

3.2 Data processing

We are targeting a task where we predict whether it will rain on a given day based on the weather conditions. During the experiments, we discovered that our data contains information from multiple time periods within a single day. Therefore, we need to consolidate our data. Firstly, for the features that contain text, we transform them into corresponding numerical values using a given method to facilitate classification. Once we have processed all features into real numbers, we proceed to consolidate the data for each day. Our chosen method is to average the data from different time periods within the same day. This allows us to obtain the input features corresponding to each day.

3.3 Model training

We have introduced the theory and details of logistic regression model in the Preliminaries.

4 Precautions and experience

As per the course requirements, we manually implemented gradient descent in this program. During the experiment, we gained the following insights:

- We found that in real-world scenarios, the data we encounter is often imperfect. Therefore, data cleaning and preprocessing before training the model are crucial for achieving better results. We realized that the quality of data and feature selection matters more than having a larger quantity of data. Extracting high-quality features and data can greatly improve the performance of the model.

- During the implementation of different gradient descent methods, we discovered significant differences in training effectiveness and speed. Therefore, when dealing with problems involving different data scales and precision requirements, it is important to choose the gradient descent method that matches the problem. The choice of gradient descent method can have a significant impact on the training performance and speed.

5 Experimental result

```
(TEST_ENVI_20230730) PS D:\whr_laptop\2023Summer\TEST\4thGrade_a\ML\Final_PJ\
test_set accuracy:accuracy:93.2236%
test_set macro_f1:f1_score:0.9185
```

图 1: result

6 Conclusion

In conclusion, our experimental results demonstrate that our logistic regression model performs well in weather prediction tasks. These findings provide valuable insights for further improving weather prediction systems or utilizing logistic regression models in other classification tasks.

7 Related code

Our code will be shown in the big_hw.ipynb.

8 Contribution

- Huanshuo Dong 33.3%
- Zhiwei Zhuang 33.3%
- Haoran Wang 33.3%