Relationship Between Students’ Study Environments and Academic Results

Gary, Chon Kit Kuok
College of Education, University of Illinois at Urbana-Champaign
Department of Computer and Information Science, Faculty of Science and Technology, University of Macau

INTRODUCTION

This research is intended to investigate the university students’ preferences on different study environments and how it affects their academic accomplishments. Intelligence is not the exclusive element which determines the academic achievement of a student. By analyzing the results, it is suggested that if students want to achieve better academic performance, they should study in places where they wish to, with several other factors’ contribution. The study can provide support for the correct broader teaching management.

The research accumulated data from various academic year’s students. After that the data is interpreted with statistical models and machine learning algorithms. Both quantitative and qualitative methods are involved in the research.

Outcomes recommend that students should study in quiet places such as not limited to libraries, in addition to the help of quietness and isolation, one can also eat some snacks during the study. Students who are not taking too many credits can avail their harvest, too. By doing so, they can focus on the study. Students who are not taking too many credits can

METHODS

In quantitative research progress, statistical models and machine learning algorithms are presented as below:

• Classification by Logistic Regression with Stochastic Gradient Descent

Acquired data have nine parameters, each is regressed to GPA, where each parameter is a variable. We define the best fit parameter that is derived by an iterative search process, is the error derivative.

Gradient descent is to minimize risk function and loss function in order to obtain a more sophisticated result. h(β) is a fitting function, J(β) is a loss function, n is the number of parameters, m is the number of training count, j is a parameter and r is a record. We then obtain the loss function of each sample

\[ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (y_i - h(x_i; \theta))^2 \]

Gradient descent is to minimize risk function and loss function in order to obtain a more sophisticated result. h(β) is a fitting function, J(β) is a loss function, n is the number of parameters, m is the number of training count, j is a parameter and r is a record. We then obtain the loss function of each sample

\[ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (y_i - h(x_i; \theta))^2 \]

The partial derivative with respect to \( \theta \) and update it

Assume the length of descending step is \( \alpha \), we have

METHODS (cont.)

\[ b_j' = b_j + \alpha (y_j - h(x_j; \theta))x_j \]

Prior to get the final result, J(θ) needs to be regularized

\[ J(\theta) = J^1(\theta) + \lambda J^2(\theta) \]

An advantage of using logistic regression rather than linear regression for classifying is all results are simplified into 0 and 1, such output is expressed as ‘possibility’, so that results’ contribution can be as well as eliminating outliers in the data. Since each response is considered as independent, we can measure the accuracy of generalized model by mean squared error and mean absolute error to evaluate goodness of fit

\[ MSE = \frac{1}{m} \sum_{i=1}^{m} (y_i - h(x_i; \theta))^2 \]

Additionally, it’s worth to compute the correlation between these 10 variables to reflect correlation

\[ corr = \frac{\sum_{i=1}^{m} (x_i - \text{mean}(x))(y_i - \text{mean}(y))}{\sqrt{\sum_{i=1}^{m} (x_i - \text{mean}(x))^2 \sum_{i=1}^{m} (y_i - \text{mean}(y))^2}} \]

Next, we used the Linear Regression model to anticipate each candidate’s GPA scale by ordinary least squares formula

\[ m_{\text{linear}}(x) = \mathbf{x}^T \hat{\mathbf{w}} + \epsilon \]

Here the model fits with coefficients \( a = \{a_1, a_2, \ldots, a_9\} \) to minimize the residual sum of squares between observed GPA scale in dataset, the responses predicted by linear approximation.

Apply Frobenius Norm for all values

\[ ||A||_F = \sqrt{\sum_{i,j} |a_{ij}|^2} \]

Last but not least, we used k-means Clustering Algorithm to cluster students into different groups by their sum of index of the nine variables in unsupervised learning setting. The formula is divided into two steps:

Given randomly chosen initial set of m centroids, \( \{c_1, c_2, \ldots, c_m\} \), where centroids are \( c_j \),

1. Assigning step: \( S_j^{(l)} = \{x_j \mid \|x_j - c_j\|^2 \leq \|x_j - c_j^{(l-1)}\|^2 \}, 1 \leq j \leq k \}, \]

2. Updating step: \( m_j^{(l+1)} = \frac{1}{|S_j^{(l)}|} \sum_{x \in S_j^{(l)}} x \)

RESULTS

The survey collected 47 effective responses and 4 successful interviews. The form below shows classified results errors for all responses with 9 originally scaled parameters:

Classified Results Errors with SGD on Different Variables

The correlation matrix below proposed that in fact among all study environments, as long as places offer good isolation, quietness that facilitate them concentrating, their GPA is thus improved. Students who spend much time to study every week can probably obtain excellent performance. They tend to stay in places they are familiar with, so there’s no much difference between library and dorm environments.

However, in these two separate matrices in dorm and library environments, we can clearly see students study in libraries have high correlation with many variables. Such results suggest that for dorm environment students, their improvement is mainly from proximity and time spend on study. In contrast, students study in library expect this place provide them comfortableness, isolation, concentration and thus boosted their GPA if they study alone. They also spend many hours on study (correlation as high as 0.7896) has positive influences on their success, too.

In library matrix, the more green color the block is, the higher correlation it is between two variables. Some red blocks reveal that they even have negative correlation, especially between credit and GPA. This points out that students take many credits in a semester may perhaps perform poorer.

REFERENCES

Influence of Learning Environment on Students’ Academic Achievement in Mathematics: A Case Study of Some Selected Secondary Schools in Yobe State – Nigeria, 2015
The Quality of Learning Environment and Academic Performance from a Student Perspective, 2009
Apache Spark Machine Learning Library, pandas Python Data Analysis Library, scikit-learn Machine Learning Library