Estimation of Hard-to-Measure Measurements in Anthropometric Survey

Kim Eugene
Dept. of Informational Statistics Korea University.
208, Jochiwon-up, Yongi-gun.
Chungnam, Korea.
sasw@naver.com

Jong Hoo Choi
Dept. of Informational Statistics Korea University.
208, Jochiwon-up, Yongi-gun.
Chungnam, Korea.
jchoi@tiger.korea.ac.kr

1. Introduction

Anthropometric survey is important as a basis for human engineering fields. According to our experiences, there are difficulties in obtaining the measurements of some body parts because respondents are reluctant to expose. In order to overcome these difficulties, we propose a method for estimating such hard-to-measure measurements by using easy-to-measure measurements those are closely related to them. Feedforward Neural Network(FNN) and Projection Pursuit Regression(PPR) model will be used as analytical tools for this purpose. The method we propose will be illustrated with real data from the 1992 Korea national anthropometric survey.

2. Analysis

The raw data contains 4,370 women. Among them, we exclude some individuals who have missing values and outliers those are out of ±3 SD from mean. Therefore, there are 2,850 samples of analytical data set left.

The two target variables are “bimammillary distance” and “lower chest circumference.” We use eight and nine body parts among the total 82 measured body parts as explanatory variables to estimate the target variables respectively, which are derived from stepwise regression analysis.

In order to find the optimal neural network model, we try to compare 1 and 2 hidden layers models with 1 to 10 neurons respectively. Finally, we find the optimal neural network model which has 2 hidden layers with 2 and 6 neurons for “bimammillary distance” and 1 hidden layer with 5 neurons for “lower chest circumference.” We use the MSE criterion for finding the optimal neural network models. And we use fraction of variance unexplained (FVU) criterion for selection of optimal PPR model[1][3]. The smaller value of FVU indicates better one. The optimal model has 11 and 10 smoother terms respectively.

$$FVU = \sum_{i=1}^{q} W_i \sum_{j=1}^{n} w_j (y_{ij} - \hat{y}_{ij})^2 \Big/ \sum_{i=1}^{q} W_i \sum_{j=1}^{n} w_j (y_{ij} - \bar{y}_{ij})^2$$

PPR Models are submitted with SAS Macros for projection pursuit regression based on the algorithm proposed by Friedman[2] [4].

3. Conclusion

In order to evaluate the PPR model and neural network models, we adapt these models to the cross validation data set which sample size is 300. In Table 1, The correlation coefficient is related to Y and $\hat{Y}$. As a result, we can find out the neural network model is preferable to the PPR models in the sense of correlation coefficient and MSE.

Table 1 Model Assesment with Cross Validation Set

	Bimammillary Distance		Lower Chest Circumference	
	Correlation	MSE	Correlation	MSE
Neural Network	0.7664	0.5359	0.9452	4.1863
PPR	0.7469	0.6655	0.9216	4.8335

REFERENCE

[1] Hwang, J.-N., Lay, S.-R., Maechler, M., Martin, D., and Shimert, J. (1994) Regression Modeling in back-propagation and projection pursuit learning. IEEE Transactions on Neural Networks 5, 342-353.

[2] Macros for projection pursuit regression and classification in SAS.

[3] Klinke, S. & Grassmann, J.(1998) Projection Pursuit Regression and Neural Networks.

[4] Friedman, J. H. (1984) SMART users' guide. Laboratory for Computational Statistics Technical Report No. 1, Dept. of Statistics, Stanford University.