Detecting Bovine Lameness Using Three-Dimensional Limb Movement Variable Analysis

UMBC REU Site: Interdisciplinary Program in High Performance Computing

William Dula[1], Jason Glover[2], Ari Goldbloom-Helzner[3], Kayla Makela[4],
RAs: Qing Ji[2] and Sai Kumar Popuri[2], Faculty mentors: Nagaraj Neerchal[2] and Andrew Raim[5], Client: Uri Tasch[6]
[1]Morehouse College, [2]UMBC, [3]Brown University, [4]Michigan State University, [5]United States Census Bureau, [6]StepAnalysis LLC

Introduction

Early detection of lameness is important for dairy farmers to enhance profits and animal welfare. Database consists of lameness score and 76 limb movement variables (LMVs) relating to a cow's gait. Previous work on this project used the TRANSREG procedure in SAS and a binomial (sound/lame) classification. In this project, we create R programs to facilitate exploring a large number of transformations and subsets of LMVs.

Methodology

Logistic regression

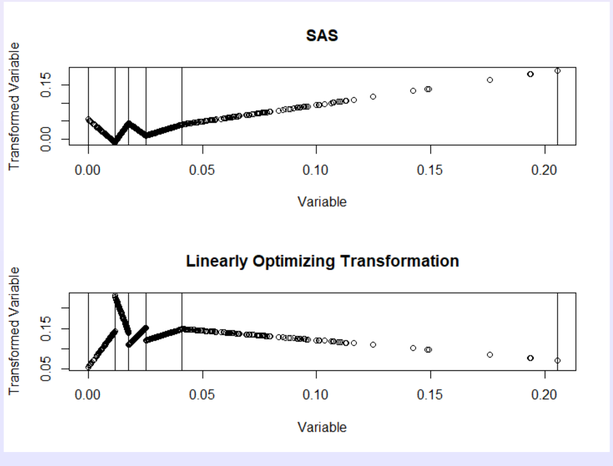
$$P(Lame) = \frac{e^{\beta_0 + \sum(\beta_i * TLMV_i)}}{1 + e^{\beta_0 + \sum(\beta_i * TLMV_i)}}$$

To find piecewise transformations, let **A** be an $n \times (k+2)$ matrix of the form

$$\begin{bmatrix} \vec{x}_1 & \vec{0} & \vec{0} & \dots & \vec{0} & \vec{1} \\ \vec{0} & \vec{x}_2 & \vec{0} & \dots & \vec{0} & \vec{0} \\ \vec{0} & \vec{0} & \vec{x}_3 & \dots & \vec{0} & \vec{0} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \vec{0} & \vec{0} & \vec{0} & \dots & \vec{x}_{(k+1)} & \vec{0} \end{bmatrix}$$

where n is the number of observations and k is the number of knots.

- Linearly optimized:
$$T\vec{x} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{\mathbf{Y}}$$
- Logistically optimized: apply logistic regression on basis for each variable
- Continuous: constrain intercepts
- Test cubic spline vs. linear spline



Lameness Detection Device

Binomial vs. Multinomial

Binomial (2 classes): lame or sound
Multinomial (3 classes): severely lame, mildly lame, or sound
MC = misclassification

Contingency Table for Binary Classifier

	+	-
Predicted +	TP	MC
Predicted -	MC	TN

Multinomial Contingency Table

	1	2	3
1	True	MC	MC
2	MC	True	MC
3	MC	MC	True

AUC = area under the curve of sensitivity (true positive rate) vs. 1 - specificity (true negative rate). Multinomial AUC can be calculated by

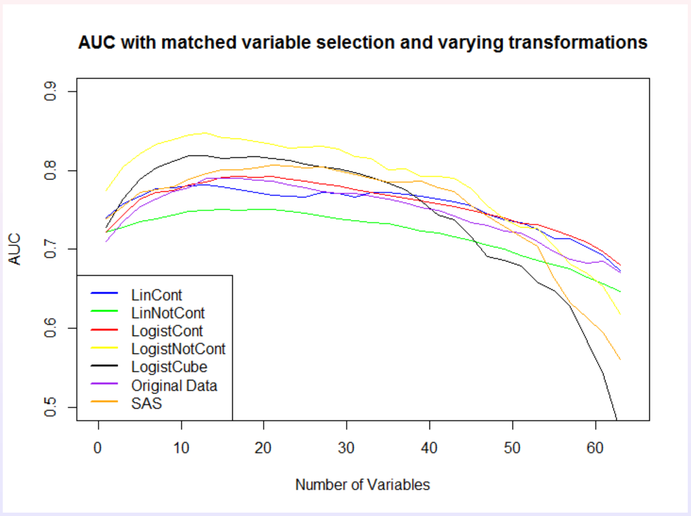
$$\text{AUC}_{\text{total}} = \sum_{c_i \in C} AUC(c_i) * p(c_i)$$

where $AUC(c_i)$ is the AUC of the class and $p(c_i)$ is the prevalence of that class in the dataset.

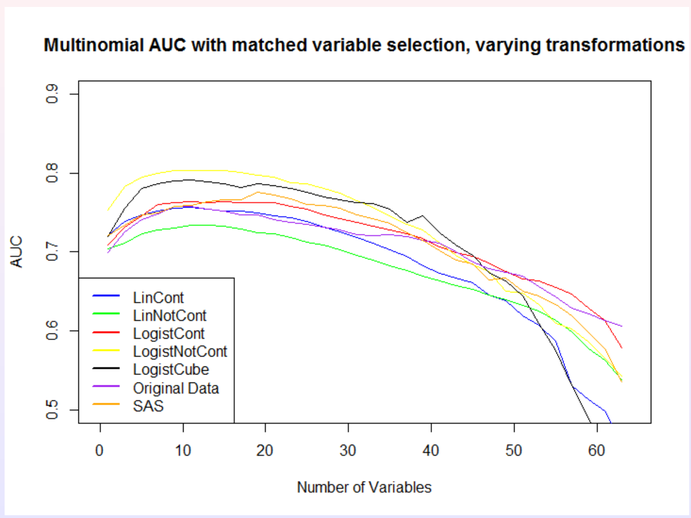
Acknowledgments

- REU Site: `hpcreu.umbc.edu`
- NSF, NSA, DOD, UMBC, HPCF, CIRC
- StepAnalysis LLC

Binomial Classification Results



Multinomial Classification Results



Conclusions

- Logistically optimized transformation achieves higher AUC than TRANSREG procedure [a]
- Multinomial model produces a more specific classification than binomial
- Successfully transitioned code from SAS to open-source R

[a]Pilot Data Set from Dec 2016

References

[1] Dunthorn, J. (2015). Journal of Dairy Research, 82(04), 391-399. doi:10.1017

[2] Fawcett, T. (2005). Science Direct

[3] Full technical report: HPCF–2017–11 `hpcf.umbc.edu` > Publications