Introduction to Accurate GLM

CAS Spring Meeting 2023

May 9, 2023

Gary Wang, FCAS, MAAA, CSPA
Senior Consulting Actuary

Agenda

- Introduction
- Piecewise Step Transformation
  - One-hot encoding
  - Ordinal encoding
  - Comparison
- Alternative Formulation of Ordinal Encoding

Introduction

Reference Paper

- Suguru Fujita, Toyoto Tanaka, Kenji Kondo and Hirokazu Iwasawa
- (2020) AGLM: A Hybrid Modeling Method of GLM and Data Science Techniques
- https://www.institutdesactuaires.com/global/gene/link.php?doc_id=16273&fg=1
- Actuarial Colloquium Paris 2020

Main Idea

Encode data to preserve ordering information

Apply penalized regression (e.g., Lasso) to remove insignificant shifts

Piecewise Step Transformation

The Classic GLM Setup

- GLM model structure

$$E[y_i] = g^{-1}(\beta_0 + \beta_1 x_{i1} \ldots + \beta_p x_{ip})$$

  - Relates response to a linear combination of NUMERIC explanatory variables

- Typical application treatment when faced with a categorical variable is to utilize one-hot encoding

One-Hot Encoding (d^U)

- Consider a categorical variable x, with levels L_1 to L_k
- Define, for $x = L_i$,

$$d^U_i(x) = 1$$

$$d^U_j(x) = 0, \text{ for } j \neq i$$

x	$d^U_1(x)$	$d^U_2(x)$	$d^U_3(x)$	...	$d^U_{k-1}(x)$	$d^U_k(x)$
L_1	1	0	0		0	0
L_2	0	1	0		0	0
L_3	0	0	1		0	0
...				...		
L_{k-1}	0	0	0		1	0
L_k	0	0	0		0	1

One Hot Encoding Transformation

- $\beta x \rightarrow \beta_1 d^U_1(x) + \beta_2 d^U_2(x) + \beta_3 d^U_3(x) + \ldots + \beta_{k-1} d^U_{k-1}(x) + \beta_k d^U_k(x)$
- We over-specified by one term
  - Choose one level as base and drop that in setting up the GLM model
- The paper refers to the $d^U_i(x)$'s as the U-dummy variables (U for Usual)

Ordinal Encoding (d^O)*

- Consider that same categorical variable x, with levels L_1 to L_k
- Define, for $x = L_i$,

  $d^O_j(x) = 1$, if $j >= i$,

  and 0 otherwise

x	$d^O_1(x)$	$d^O_2(x)$	$d^O_3(x)$	...	$d^O_{k-1}(x)$	$d^O_k(x)$
L_1	1	1	1		1	1
L_2	0	1	1		1	1
L_3	0	0	1		1	1
...				...		
L_{k-1}	0	0	0		1	1
L_k	0	0	0		0	1

**With apologies to Iwasawa Hirokazu and team, I took the liberty of adding the main diagonal for consistency*

Ordinal Transformation

- $\beta x \rightarrow \beta_1 d^O_1(x) + \beta_2 d^O_2(x) + \beta_3 d^O_3(x) + \ldots + \beta_{k-1} d^O_{k-1}(x) + \beta_k d^O_k(x)$
- Again, we over-specified by one term
  - Note: To line up with the matrix in the paper, we take care of the over-specification by
    - Dropping d^O_k (setting column to 0's), and
    - Moving the column to position 1
- The paper refers to the $d^O_i(x)$'s as the O-dummy variables (O for Ordinal)

A Basic Example

- Consider a six-level variable x, under the two encodings

x	βx	$d^U_1(x)$	$d^U_2(x)$	$d^U_3(x)$	$d^U_4(x)$	$d^U_5(x)$	$d^U_6(x)$
L_1	6	1	0	0	0	0	0
L_2	5	0	1	0	0	0	0
L_3	4	0	0	1	0	0	0
L_4	3	0	0	0	1	0	0
L_5	2	0	0	0	0	1	0
L_6	1	0	0	0	0	0	1
β		6	5	4	3	2	1

x	βx	$d^O_1(x)$	$d^O_2(x)$	$d^O_3(x)$	$d^O_4(x)$	$d^O_5(x)$	$d^O_6(x)$
L_1	6	1	1	1	1	1	1
L_2	5	0	1	1	1	1	1
L_3	4	0	0	1	1	1	1
L_4	3	0	0	0	1	1	1
L_5	2	0	0	0	0	1	1
L_6	1	0	0	0	0	0	1
β		1	1	1	1	1	1

A Basic Example – Dropping d_4 as Insignificant

- Suppose the data has low volume at level L_4

x	βx	$d^U_1(x)$	$d^U_2(x)$	$d^U_3(x)$	$d^U_4(x)$	$d^U_5(x)$	$d^U_6(x)$
L_1	6	1	0	0	0	0	0
L_2	5	0	1	0	0	0	0
L_3	4	0	0	1	0	0	0
L_4	**0**	0	0	0	1	0	0
L_5	2	0	0	0	0	1	0
L_6	1	0	0	0	0	0	1
β		6	5	4	**0**	2	1

x	βx	$d^O_1(x)$	$d^O_2(x)$	$d^O_3(x)$	$d^O_4(x)$	$d^O_5(x)$	$d^O_6(x)$
L_1	6	1	1	1	1	1	1
L_2	5	0	1	1	1	1	1
L_3	4	0	0	1	1	1	1
L_4	**2**	0	0	0	1	1	1
L_5	2	0	0	0	0	1	1
L_6	1	0	0	0	0	0	1
β		1	1	**2**	**0**	1	1

Observations, for Level x = 4

- One-Hot Encoding

  $\beta_4 x = \beta_4 d^U_4(x) = 3(1) = 3$

- Dropping $d^U_4(x)$ as insignificant

  $\beta_4 = 0$

  $\beta_4 x = 0$

- Ordinal Encoding

  $\beta_4 x = \beta_4 d^O_4(x) + \beta_5 d^O_5(x) + \beta_6 d^O_6(x)$

  $= 1(1) + 1(1) + 1(1) = 3$

- Dropping $d^O_4(x)$ as insignificant

  $\beta_4 = 0$

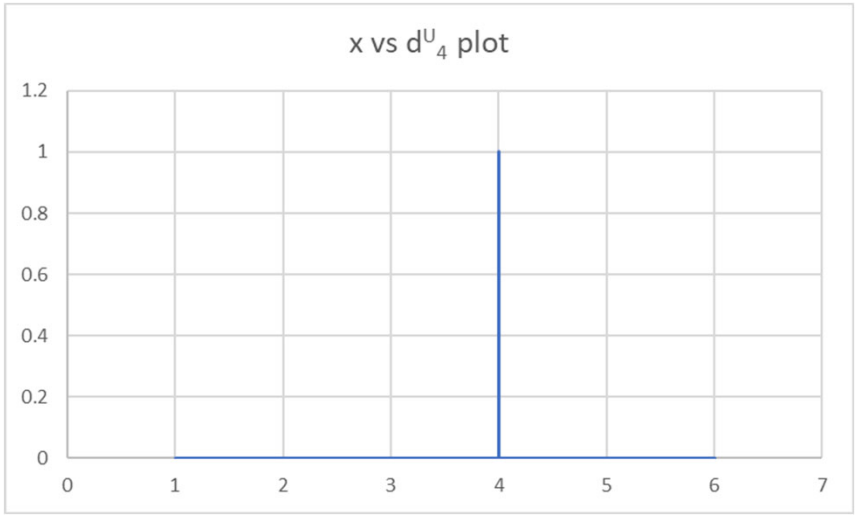
  $\beta_4 x = \beta_5 d^O_5(x) + \beta_6 d^O_6(x)$

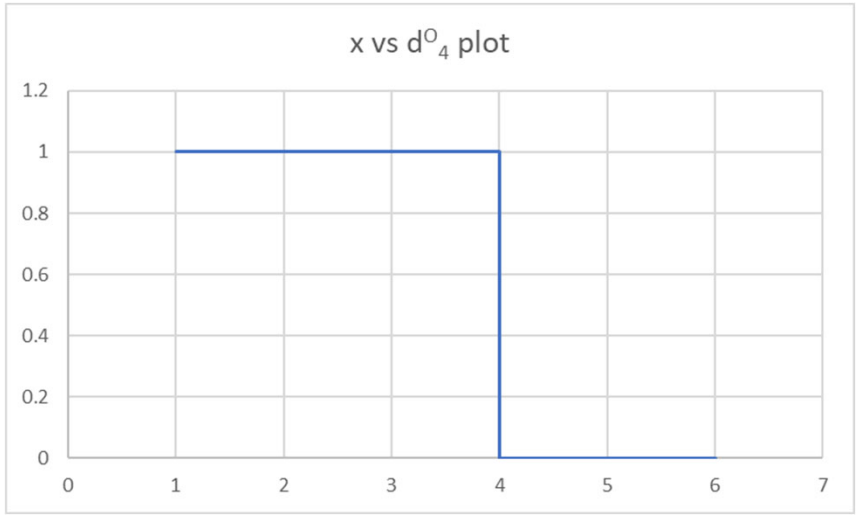
  $= 1(1) + 1(1) = 2 = \beta_5 x$

What the Ordinal Encoding Offers, Conceptually

- Under One-Hot Encoding
  - Dropping a level implies grouping with base level
  - To do otherwise requires modeler intervention
- Under Ordinal Encoding
  - Dropping a level implies grouping with successive level
  - Again, to do otherwise requires modeler intervention

Encoding as Building Blocks





Alternative Formulation of Ordinal Encoding

Ordinal Encoding – Subsequent Grouping

- Consider Ordinal Encoding
  - 6 levels
  - d^O_6 dropped to set L_6 as base

- Recall dropping d^O_4 would group L_4 into L_5, matching the **subsequent** level

X	$d^O_1(x)$	$d^O_2(x)$	$d^O_3(x)$	$d^O_4(x)$	$d^O_5(x)$	$d^O_6(x)$
L_1	1	1	1	1	1	0
L_2	0	1	1	1	1	0
L_3	0	0	1	1	1	0
L_4	0	0	0	1	1	0
L_5	0	0	0	0	1	0
L_6	0	0	0	0	0	**Base**

Ordinal Encoding – Precedent Grouping

- Consider a Transposed Ordinal Encoding
  - Transpose the Triangular Matrix
  - d^O_1 dropped to set L_1 as base

- In this setup, dropping d^O_4 would group L_4 into L_3, matching the **precedent**

X	$d^O_1(x)$	$d^O_2(x)$	$d^O_3(x)$	$d^O_4(x)$	$d^O_5(x)$	$d^O_6(x)$
L_1	Base	0	0	0	0	0
L_2	0	1	0	0	0	0
L_3	0	1	1	0	0	0
L_4	0	1	1	1	0	0
L_5	0	1	1	1	1	0
L_6	0	1	1	1	1	1

Combining the Two Effects – Adjusting the Base Level

- Suppose we want to make level L_3 the base?
  - Let's group toward L_3!

- We can build a subsequent grouping encoding to the left, and a precedent grouping encoding to the right

X	$d^O_1(x)$	$d^O_2(x)$	$d^O_3(x)$	$d^O_4(x)$	$d^O_5(x)$	$d^O_6(x)$
L_1	**1**	**1**	0	0	0	0
L_2	0	**1**	0	0	0	0
L_3	0	0	**Base**	0	0	0
L_4	0	0	0	**1**	0	0
L_5	0	0	0	**1**	**1**	0
L_6	0	0	0	**1**	**1**	**1**

On to the Penalized Regression Application…

[Please see Liam McGrath's presentation material from this AGLM session]

Thank You

Gary Wang

309.807.2348

gwang@pinnacleactuaries.com

PINNACLE
ACTUARIAL RESOURCES