

PERGUNTAS E RESPOSTAS – MBA EM DATA SCIENCE E ANALYTICS

Disciplina: Supervised Machine Learning: Modelos para Dados de Contagem II

Data: 21/09/2021

Patricia Barbosa Pacheco

Nós temos direito a ir a biblioteca da Esalq também?

Patrícia, peço a gentileza de que encaminhe e-mail para a madrinha do curso, para que ela possa te encaminhar todas as orientações sobre acesso à Biblioteca.

Gabriel Campos Vieira

Professor, vc comentou sobre discretizar as variáveis no caso de não passar após BC. Como seria essa transformação? Não seria uma manipulação da base de dados?

Gabriel, essa decisão deve ser tomada em cada caso e tomando como base o problema de pesquisa e natureza dos dados. Há vários métodos matemáticos para operacionalizar essa transformação, como por exemplo o Método de Euler-Maruyama. A própria regressão logística pode ser um caso particular desta discretização, em ocasiões em que aproximamos determinada probabilidade dos valores zero ou um.

Alex Jornada Queiroz

Eu não entendi de onde o Favero tirou a equação que ele usou para calcular manualmente o predict da ZINB. Qual slide da apostila?

Alex, o cálculo manual do predict do ZINB é obtido com a substituição dos parâmetros no modelo encontrado. Faz referência à função densidade de probabilidade obtida por meio da equação vista no slide 23 da aula.

Marina Bendit Szacher

essa equação que o professor está escrevendo é para $Y=0$ ou $Y=m$, confirme slide 23?

Marina, se entendi sua pergunta, há duas equações para os componentes do modelo. Uma quando o $y=0$ e outra quando o y é maior do que zero.

Fabício de Oliveira Alves

por que para calcular ($Y=0$) considera-se apenas a variável preditora corruption?

Fabício, foi uma escolha para fins didáticos. Se sabia de antemão que as outras não eram significantes para o fenômeno, mas caso quisesse poderia colocar as outras variáveis sem problemas.

Laila Monte Neto Donni

não entendi como chegamos ao valor do plogit, poderia explicar novamente?

Laila, o valor de plogit é calculado a partir do raciocínio visto na aula de Regressão Logística. Visto que no modelo inflacionado de zeros temos um componente de ocorrência ou não ocorrência de zeros, ou seja, uma condição dicotômica de ocorrência de determinado evento, considerando os zeros estruturais e os zeros amostrais. O plogit nos fornece a probabilidade do evento ser zero, ou a não ocorrência do evento.

Maria Leidiane Kelly Dionísio Barros Sales

Como se faz para discretizar a variável y ?

Maria, essa decisão deve ser tomada em cada caso e tomando como base o problema de pesquisa e natureza dos dados. Há vários métodos matemáticos para operacionalizar essa transformação, como por exemplo o Método de Euler-Maruyama.

Ludmilla Manera Conti

Tem como transformar/cortar uma dispersão de cauda longa sem ser para fins didáticos?

Ludmilla, essa possibilidade existe. Depende do seu problema de pesquisa e se isso tem algum sentido matemático, estatístico, empírico para sua análise. Pode haver algum caso em que a natureza de sua pesquisa imponha uma análise em separado dos dados da cauda e dos demais dados da distribuição.

Patrick Augusto Da Silva Souto

Me parece que após a atualização da lei não temos superdispersão nos dados. Seria válido dividir a base e realizar BNeg antes da lei e Poisson depois da lei para predição?

Patrick, é uma forma de analisar os dados. A depender do seu problema de pesquisa poderia ser válido. Uma aplicação que enxergo seria não para comparar os modelos, mas para verificar as variáveis que podem influenciar o fenômeno nos dois momentos distintos (antes e depois da lei).

Bernardo Silva

No caso da Poisson Gama, qual o nome do teste e qual a hipótese (H_0 e H_1) testada?

Bernardo, se entendi sua pergunta, está falando do teste de Vuong, o qual nos permite escolher entre o modelo Poisson e o modelo ZIP. A estatística do teste de Vuong testa a hipótese nula de que os dois modelos estão igualmente próximos do verdadeiro processo de geração de dados, contra a alternativa de que um modelo está mais próximo da verdadeira distribuição observada.

Lucas Bruno Marques

Professor, se, no caso onde o ϕ não for estatisticamente significativo, tanto o modelo Poisson quanto o Binomial Negativo serão equivalentes, o BN é mais abrangente e superior em todos os casos?

Lucas, essa decisão deve ser tomada em cada caso, considerando os testes estatísticos e o problema de pesquisa. Em estatística é complicado gerar uma lei geral para todos os casos, considerando a multiplicidade de problemas, modelos, contextos do mundo real. Esta preocupação é particularmente importante quando temos o fenômeno da inflação de zeros.

Saulo Galdino Alves Guedes

Semana passada o professor deu um exemplo de usar o modelo para prever dados de contagem em modelos físicos, usando um quadriculado. Como fazemos isso na prática, como especificamos o modelo?

Saulo, cada modelo tem a sua especificação de acordo com as escolhas do pesquisador. Há várias possibilidades de realizar essa especificação a depender do caso concreto.

Gustavo Murad

ainda não vi a acuracidade do modelo como boa; se vc compara valor real da Y com qqr predicao, fica mto longe mesmo com modelo Zero inflation..perdi algo?

Gustavo, o modelo mais adequado para o caso exemplificado pelo professor é o modelo ZINB. Temos um ganho de capacidade preditiva considerável que foi sendo obtida por intermédio das especificações vistas durante a aula.

Daniel Christofolletti

p menor que 0,05

É o nível de significância escolhido para realizar o teste de hipóteses, Daniel.

Daniel Christofolletti

boa noite., Não entendi o teste de Vuong...quando p 0,05 os modelos são estatisticamente diferentes ou não?

Daniel, a estatística do teste de Vuong testa a hipótese nula de que os dois modelos estão igualmente próximos do verdadeiro processo de geração de dados, contra a alternativa de que um modelo está mais próximo da verdadeira distribuição observada. No caso em questão, se entendi bem sua pergunta, rejeitamos a hipótese nula e concluímos que um dos modelos está mais próximo da verdadeira distribuição.

Gustavo Murad

analisando Y reais com os fitted para ambos poisson e binomial, temos que o modelo está muito ruim; entendo que para fins didáticos chegaremos a um modelo melhor ao eliminar o zero inflation?

Exatamente, Gustavo. O caminho percorrido é no sentido de procurar modelos melhores para o exercício em questão.

Flávia Ruiz Leão

Por que distribuições de parâmetros (alfa, beta, teta) regridem a distribuições normais?

Flávia, são premissas estatísticas já observadas.