

PERGUNTAS E RESPOSTAS – MBA EM DATA SCIENCE E ANALYTICS

Disciplina: Supervised Machine Learning: Modelagem Multinível I

Data: 28/09/2021

Cleverson de Souza

Erros durante o embrulho: cyclic namespace dependency detected when loading 'msm', already loading 'msm' Error: no more error handlers available (recursive errors?); invoking 'abort' restart?

Cleverson, quando pedir 'Do you want to install from sources the package which needs compilation?' Tente ir na opção "No". Caso não seja esse o problema, peço que, por gentileza, volte com a gente no e-mail da monitoria, que iremos prontamente lhe auxiliar.

Yuri José de Santana Furtado

Podemos usar vários modelos GLM nos subníveis de um só modelo multinível?

Yuri, se entendi bem sua pergunta, essa especificação é possível com a utilização de modelos FMM (Finite Mixture Models). Com esta modelagem seria possível capturar comportamentos distintos para as funções de nível contextual

Carlos Henrique de Oliveira

Professor, não abriremos a "caixa preta" de funções como esta criada pelo professor Rafael?

Carlos, as funções são acessíveis no próprio R. Todos os parâmetros e argumentos já se encontram disponíveis.

Marcelo Ferrari

A média da variável "texp", neste banco de dados, não seria uma forma arbitrária para dar valor ao "W" ? Seria recomendável o uso de variáveis não calculadas ou de PCA para esta atribuição?

Marcelo, não consigo ver a arbitrariedade neste caso, é uma variável contextual que, devido a uma subjacência teórica ou empírica, pode influenciar no comportamento das notas dos alunos e não varia dentro das observações de uma mesma escola, ou seja, os requisitos para este tipo de variável estão cumpridos. É uma variável construída para fins didáticos, mas há inúmeras possibilidades neste caso, outras variáveis são passíveis de serem incluídas nesta análise, como a estrutura física da escola, a quantidade de aulas postas à disposição dos alunos, o fornecimento de livros didáticos, a quantidade de investimento por aluno, enfim, há inúmeras possibilidades. Mas concordo que uma PCA é sim passível de utilização, de modo a encontrar os fatores determinantes de um dado resultado dos alunos por escola, parece um bom método.

Marcos Antonio Lopes Freixo Filho

O que é uma variável "W"? (perdi alguma(s) coisa(s) em algum momento)

Marcos, é uma variável explicativa, com a diferença de ser invariável para todos os alunos de uma determinada escola. Neste caso foi somente uma mudança de nomenclatura para fins didáticos, ela é uma variável explicativa normal (um beta comum).

Alex De Lima Bassi

Professor, concordo que a primeira afirmação é incorreta, pois a modelagem de bancos de dados é por natureza multinível(normalização) e os dados em qualquer tabela respeitam um contexto (log, cadastro)

Alex, não entendemos sua pergunta. Caso queira, nos encaminhe sua dúvida no e-mail da monitoria que iremos prontamente lhe auxiliar.

Carlos Rodrigo Costa

Professor nos grupos de estudo, uma das situações que mais nos trava no desenvolvimento dos estudos é a manipulação do dataset, no curso haverá alguma abordagem sobre tratamento de dados?

Carlos, é sempre um dos maiores entraves em relação à manipulação dos dados. Iremos repassar essa sugestão à coordenação.

Yuri Wladimir Pitthan

Professor, há alguma relação entre modelos multiníveis e SEM (Structural Equation Modeling)?

Yuri, são técnicas com objetivos distintos. A modelagem multinível objetiva capturar os distintos contextos que podem haver em uma determinada amostra. Já a modelagem de equações estruturais (MEE), grosso modo, pode ser vista como uma combinação entre as técnicas de análise fatorial e regressão. A MEE trabalha a partir de construções teóricas que podem ser desenvolvidas a partir dos chamados construtos latentes (variáveis não diretamente observáveis pelo pesquisador), sendo que as relações entre as construções teóricas são representadas por coeficientes de regressão.

Eduardo Fajardo

Professor , Boa noite , podemos dizer que há etapas hierárquicas que vão melhorando o modelo a medida que vamos reduzindo a variância do modelo

Eduardo, o que estamos fazendo ao longo da trajetória, construindo diferentes tipos de modelos, é tentar capturar ao máximo esse fenômeno, explicando a maior parcela possível dessa variação. Seria como se tivéssemos tentando aproximar as retas (ou hiperplanos) construídas a partir dos modelos o máximo possível dos dados reais, de modo a diminuir os resíduos ou erros.

Rafael Viegas De Carvalho Carlos Gomes

Está parte que o professor está explicando agora não está claro. Não entendi a relação dos termos de V_{0j} e V_{1j} com o intercepto.

Rafael, os termos V_{0j} e V_{1j} ("V" aqui representativo da letra grega "Ni") representam efeitos aleatórios. São utilizados para modelarmos a situação em que queremos saber se o intercepto e a inclinação variam de escola para escola, situação em que teríamos diferentes retas para cada contexto e seria preferível a utilização da modelagem multinível. O V_{0j} busca verificar a variação do intercepto e o V_{1j} a variação da inclinação, sendo representativos de duas situações, respectivamente: intercepto com efeitos aleatórios e inclinação com efeitos aleatórios.

Guilherme Sampaio Bacelo

Em um modelo de dados em painel, o ano pode ser considerado como um nível?

Guilherme, embora os objetivos sejam diferentes, é um modo de relacionar as duas técnicas. Inclusive algumas técnicas relacionadas aos dados em painel objetivam fazer variar alguns coeficientes ao longo dos painéis.

Vera Lucia Candiani

Prof. há alguma condição para as amostras de cada nível?

Vera, não entendemos a sua pergunta. Caso queira, peço que, por gentileza, encaminhe sua pergunta no e-mail da monitoria, ficaremos felizes em ajudar.

Luciana da Costa Amorim

Esse estudo é um paper?

Luciana, não entendemos a sua pergunta, não sabemos o momento específico da aula ao qual está se referindo. Caso queira, peço que, por gentileza, encaminhe sua dúvida no e-mail da monitoria, ficaremos felizes em ajudar.

Rafael Viegas De Carvalho Carlos Gomes

Essa estrutura multinível não seria um caso de multicolinearidade?

Rafael, se entendi bem sua questão, seria mais um caso de heterocedasticidade, os resíduos acabam por apresentar variâncias diferentes para a amostra. Inclusive é uma das recomendações quando estamos diante de modelos que violam o pressuposto da homocedasticidade dos resíduos.

Ivan Cesar Desuo

sim, mas não é o caso da aula passada. Nesse caso minha Y é logística (0, 1) com prevalência muito baixa, por exemplo 1% de ocorrência (anomalia). Na aula passada, a Y era contagem com zero inflado.

Ivan, se entendemos bem sua pergunta, essa possibilidade existe sim. Modelos logísticos com aplicação de técnicas de modelagem multinível. É uma classe de modelos bastante recente e ainda pouco discutida na literatura. Caso queira indicações de literatura peço que, por gentileza, encaminhe a solicitação no e-mail da monitoria, lá teremos mais espaço e oportunidade para tentar te auxiliar.

Vanessa Hoffmann de Quadros

Professor, nos exemplos dados a forma funcional é sempre a mesma para cada nível. Seria possível um modelo multinível com formas funcionais diferentes para cada grupo?

Vanessa, essa especificação é possível com a utilização de modelos FMM (Finite mixture models). Com esta modelagem seria possível capturar comportamentos distintos para as funções de nível contextual.

Cassiano Pedroso

Prof. quando classificamos as escolas em escola 1,2,3 e etc, não seria isso ponderação arbitrária? Ao invés disso, por que não inserimos o nome da escola? Obrigado

Cassiano, os números aqui são somente categorias semânticas, são indicativos de um contexto qualitativo. Os números são colocados aqui por uma questão didática. Poderia ser colocado também o nome da escola.

EMANUEL RODRIGUES DE VARGAS

Professor, me recordo de ter aprendido que dados em painel segue uma construção temporal, mas entendo agora que não há essa obrigatoriedade.

Emanuel, pertinente a sua observação. São técnicas com objetivos distintos mas com uma relação. De certa forma, nos dados em painel, estamos aplicando uma diferenciação entre determinados grupos considerando determinada base temporal variável.

Arno Paulo Schmitz

professor, e quando a variável W é qualitativa com vários níveis

Arno, o raciocínio é análogo. Seria o caso, por exemplo, de incluirmos o nível "município" ou o nível "estado" na análise exemplificada pelo professor Fávero em aula.

Hernandes Matias Junior 124.799.116-40