

## PERGUNTAS E RESPOSTAS – MBA EM DATA SCIENCE E ANALYTICS

**Disciplina:** Supervised Machine Learning: Modelos para Dados de Contagem I

**Data:** 14/09/2021

**Francisco Frantz**

Por misturar modelos diferentes e permitir gerar curvas complexas, os FMMs tem potencial de ser superior a qualquer outro modelo? O custo computacional é muito maior que dos modelos "tradicionais"?

**Resposta:** Ótima pergunta. Sim, os modelos FMM são os mais gerais a partir dos quais todos os outros modelos supervisionados derivam. Na realidade, cada modelo supervisionado é uma particularidade de um modelo mais geral FMM. Os modelos multinível são aproximações dos modelos FMM, e iremos estudá-los em duas aulas. Atualmente há um grande desafio, que é a capacidade computacional para estimá-los, principalmente por conta de interações profundas entre as variáveis preditoras, bem como entre estas e componentes de efeitos aleatórios.

**Gabriela Alves De Almeida**

prof em alternativa ao shapiro francian podemos usar o Anderson-Darling test? (ad.test)

**Resposta:** É uma boa pergunta. Sim, pode-se utilizar o teste de Anderson-Darling. Para a comparação entre os diversos testes de normalidade, recomendo a leitura do seguinte artigo: <https://www.tandfonline.com/doi/abs/10.1080/00949655.2014.947986>

**Lays Janaina Prazeres Marques**

Professor, boa noite. Como incluir o intervalo de confiança no output do modelo?

**Resposta:** Por favor, tente o argumento `confint = T`.

**Laila Monte Neto Donni**

Professor, como devemos analisar a deviance nesses modelos?

**Resposta:** Vemos a palavra Deviance duas vezes na saída do modelo. Deviance é uma medida de inadequação de um modelo GLM. Em outras palavras, é uma medida da inadequação do ajuste - valores mais altos indicam um ajuste pior. O R relata duas formas de desvio - null deviance e residual deviance. O null deviance mostra quão bem a variável dependente é prevista por um modelo que inclui apenas o intercepto.

Para um de nossos exemplos, temos um valor de null deviance igual a 570,78 com 297 graus de liberdade. A inclusão das variáveis preditoras diminuiu o desvio para 239,13 (residual deviance) com 294 graus de liberdade, o que representa uma redução significativa no desvio a um nível de significância de 5% ( $p\text{-value } \chi^2 < 0.05$ ).

**Thiago Colette Vegi**

O pseudo  $R^2$  ficou diferente na exibição do modelo `summ` e `export_summs`. Qual é o correto?

**Resposta:** Os pseudo  $R^2$  em ambos são idênticos. Deve-se olhar o pseudo  $R^2$  de Cragg-Uhler.

**Felipe Francisco Nusda**

podemos dizer que a exposição tem relação com a granularidade?

**Resposta:** Sem dúvidas. A granularidade nos dados é decorrente da exposição com que foram coletados.

**Thiago Colette Vegi**

Assistir a série do Netflix falando sobre a challenger. Fiz a estimativa considerando 11 graus celsius e cheguei a ~50%. Eles usaram um modelo similar para estimar a temperatura mínima de 11 graus?

**Resposta:** Boa pergunta. Não tenho conhecimento se fizeram esta análise antes do lançamento, mas deveriam.

**André Sigora**

Professor, a gente viu o teste estatístico da superdispersão para definir se é poisson ou binomial negativa. Mas não deveria acontecer um teste da aderência à própria poisson? E se for uma Weibull?

**Resposta:** Sim, o teste de aderência à própria Poisson pode ser o  $\chi^2$  test, conforme visto mais adiante no mesmo código, quando comparamos os modelos Poisson e binomial negativo com os modelos zero-inflated. Quanto à regressão Weibull (muito utilizada para dados de sobrevivência), a lógica é a mesma, e tratamos desta estimação no Capítulo 17 do Manual de Análise de Dados.

**Leonardo Alves Peixoto**

Em vez de observar os dados por quadrículas diretamente, essas quantidades poderiam ser estimadas por alguma técnica de amostragem (amostragem por quadratura, por exemplo)?

**Resposta:** Boa pergunta. Sim, estas quantidades podem ser definidas a partir de métodos específicos de amostragem.

**Luís Cláudio da Silva e Silva**

Professor, não há uma pequena diferença entre o gráfico com  $\lambda = 1$  gerado no excel e o gráfico gerado no R. Observe o valor máximo. No excel fica em torno de 36% e no R parece passar de 40%.

**Resposta:** Muito bem observado. Os valores das duas funções são idênticos para  $m = 0$  e  $m = 1$ , porém o método de interpolação no R é mais acentuado a fim de garantir uma melhor estética na plotagem.

**Thiago Ricardo**

Boa noite, tudo bem? Vimos em uma aula anterior que não faz nunca sentido calcular  $R^2$  para modelos de classificação. Como isso funciona quando considerando um modelo ensemble do tipo GradientBooster? Isso porque o gradient booster sempre faz uma regressão, mesmo pra problemas de classificação.

**Resposta:**  $R^2$  representa percentual de variância (quadrado do desvio-padrão), e variáveis qualitativas não têm desvio-padrão. Haverá aulas de ensemble models (boosting, por exemplo) no próximo módulo do curso.

**Samya de Lara Lins de Araujo Pinheiro**

Em modelos do GLM para alguns pacotes do R temos opção de usar 'quasipoisson', o que isso quer dizer?

**Resposta:** Boa pergunta. São modelos binomias negativos do tipo 1, ou seja, com especificação linear na expressão da variância.

**Eric Scopinho**

Prof. Em diversos cursos introdutórios de ML, fala-se muito do algoritmo de 'gradient descent'. Ao usar o 'Solver' no Excel, é este o algoritmo utilizado para chegar nos valores dos alfas/betas?

**Resposta:** No Solver é utilizado o algoritmo GRG (Generalized Reduced Gradient) Não Linear. Sugiro a leitura do seguinte link:

<https://www.sciencedirect.com/topics/engineering/reduced-gradient>

Hernandes Matias Junior 124.799.116-40