

CE225 - Modelos Lineares Generalizados

Cesar Augusto Taconeli

13 de setembro, 2017

Aula 9 - Seleção de covariáveis

Seleção de covariáveis

- O processo de seleção de covariáveis tem por objetivo a identificação de um modelo parcimonioso: que seja simples, com reduzido número de parâmetros, mas capaz de ajustar satisfatoriamente os dados.
- Para problemas em que se tem um pequeno número de fatores e covariáveis, a busca por um modelo com melhor ajuste em geral é feita a partir da análise de deviances, com base em resultados de testes de hipóteses.
- Em aplicações que envolvem um número mais elevado de covariáveis (e fatores), usar algum algoritmo de seleção, para identificação de um modelo adequado, pode ser útil.
- Em qualquer situação, analisar primeiramente o efeito marginal de cada covariável (usando gráficos de dispersão, por exemplo) é sempre recomendável.

Critério de Informação de Akaike

- O critério de informação de Akaike (*Akaike Information Criterion - AIC*) é uma importante medida usada para avaliar a qualidade do ajuste de modelos.
- Para um modelo qualquer, seja \hat{l} a log-verossimilhança maximizada e p o número de parâmetros estimados. Então:

$$AIC = -2\hat{l} + 2p. \quad (1)$$

- O AIC é uma medida de qualidade de ajuste penalizada pela complexidade do modelo (número de parâmetros).
- Para um problema qualquer, pode-se ajustar diferentes modelos e optar por aquele que produzir menor AIC.

Critério de Informação de Akaike

- Diferentemente dos métodos estudados até o momento, a comparação de modelos via AIC não requer que eles sejam encaixados (um modelo resultante de alguma restrição nos parâmetros de outro).
- Há diversas variações do AIC, dentre as quais o critério BIC:

$$BIC = -2\hat{l} + \ln(n) \times p. \quad (2)$$

- O BIC penaliza mais fortemente modelos mais complexos para maiores tamanhos de amostras.

Algoritmos de seleção de covariáveis - o método Forward

- Um primeiro algoritmo de seleção de covariáveis é o **método forward**;
- Começando pelo modelo nulo (apenas com o intercepto), o método seleciona, dentre todas as covariáveis, aquela que proporciona maior ganho de ajuste (segundo algum critério como, por exemplo, menor AIC);
- Nos passos seguintes, uma a uma as demais covariáveis são inseridas ao modelo, sempre selecionando aquela que proporciona maior ganho de ajuste na presença das covariáveis já inseridas ao modelo;
- O processo encerra quando nenhuma das covariáveis fora do modelo contribui para um melhor ajuste, segundo o critério adotado.

Algoritmos de seleção de covariáveis - o método Backward

- Outro algoritmo de seleção de covariáveis é o **método backward**;
- Começando pelo modelo mais complexo a ser considerado, exclua do modelo aquele termo (covariável, fator, interação. . .) com menor contribuição para o ajuste;
- Nos passos seguintes, um a um os demais termos são excluídos do modelo, sempre selecionando aquele que proporciona menor contribuição para o ajuste na presença dos termos remanescentes;
- O processo encerra quando a exclusão de qualquer termo do modelo resulte em um pior ajuste (segundo o critério adotado).

Algoritmos de seleção de covariáveis - método combinado

- Uma terceira variação de algoritmo de seleção combina os dois métodos anteriores.
- Neste caso, inicia-se com o modelo com todos os termos e seleciona-se para exclusão o de menor contribuição para o ajuste;
- A cada nova exclusão, no entanto, verifica-se também a possibilidade de inclusão de algum termo excluído nos passos anteriores;
- O processo encerra quando nenhum termo excluído tiver *força para entrar* e nenhum termo incluído for *fraco suficiente para sair*.