

CE225 - Modelos Lineares Generalizados

Cesar Augusto Taconeli

13 de setembro, 2018

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 pode ser 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 preditor 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

- 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.
- Diferentemente dos métodos estudados até o momento, a comparação de modelos usando AIC e BIC não requer que eles sejam encaixados (um modelo resultante de alguma restrição nos parâmetros de outro).
- Assim, modelos especificados com diferentes distribuições para o componente aleatório, ou diferentes funções de ligação, podem ter seus ajustes comparados usando medidas de informação.

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;
- 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 (cuja exclusão resulte em maior AIC, por exemplo);
- 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 versão 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 suficiente para entrar* e nenhum termo incluído for *fraco suficiente para sair*.

Métodos de regularização

Métodos de regularização

- Como alternativa aos métodos de seleção apresentados, métodos de regularização têm por objetivo ajustar um modelo em que as estimativas dos p parâmetros sejam *regularizadas* (ou *encolhidas*) em direção a zero.
- O objetivo da regularização é a produção de estimativas com menor variância.
- A contrapartida é a produção de algum viés nas estimativas, mas que deve ser compensada pela menor variância, resultando em maior eficiência.

Métodos de regularização

- O erro quadrático médio de um estimador $\hat{\theta}$, na estimação de um parâmetro θ , é definido por:

$$EQM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2],$$

que pode ser decomposta, facilmente, em:

$$EQM(\hat{\theta}) = E[(\hat{\theta} - E(\hat{\theta}))^2] + (E(\hat{\theta}) - \theta)^2,$$

em que o primeiro componente da soma corresponde à variância de $\hat{\theta}$ enquanto o segundo se refere ao seu viés quadrático.

Métodos de regularização

- Em problemas gerais de estimação, viés e variância caminham em direção oposta, de maneira que quanto menor o viés dos estimadores, maiores suas variâncias, e vice-versa (*tradeoff bias-variance*).
- Os métodos de regularização permitem identificar estimativas com algum grau de viés, mas tais que a redução decorrente na variância remeta a uma estimação mais eficiente.
- As duas técnicas de regularização mais conhecidas são a *regressão ridge* e a técnica *lasso*, conforme descritas na sequência.

Métodos de regularização

- No contexto de MLG, os métodos de regularização baseiam-se na identificação de β_0 (intercepto) e β (demais parâmetros do modelo) que minimizam a seguinte função:

$$-\frac{1}{n} \sum_{i=1}^n l(y_i, \beta_0 + \mathbf{x}'_i \beta) + \lambda [(1 - \alpha) \|\beta\|_2 + \alpha \|\beta\|_1],$$

em que $l(y_i, \beta_0 + \mathbf{x}'_i \beta)$ é a contribuição da observação i para a função de log-verossimilhança.

- Além disso, $\alpha \|\beta\|_1 = \alpha \sum_{j=1}^p |\beta_j|$ e $(1 - \alpha) \|\beta\|_2 = (1 - \alpha) \sum_{j=1}^p \beta_j^2$ são termos de penalização de primeira e segunda ordem, respectivamente.

Métodos de regularização

- Ainda, $\lambda \geq 0$ é um parâmetro que controla a intensidade da penalização, que deve ser determinado separadamente.
- Quando $\lambda = 0$, o termo de penalização não tem efeito, e as estimativas produzidas correspondem às de máxima verossimilhança;
- Quando $\lambda \rightarrow \infty$, então a penalização será tão forte que as estimativas produzidas serão próximas de zero;
- A escolha de λ pode ser feita por validação cruzada, identificando o valor que produz menor erro quadrático médio ou erro absoluto médio.

Métodos de regularização - o método *lasso*

- Para $\alpha = 1$, o método de regularização, baseado em penalização de primeira ordem, é denominado *regressão lasso*;
- Nessa situação, a função a ser minimizada é do tipo:

$$-\frac{1}{n} \sum_{i=1}^n l(y_i, \beta_0 + \mathbf{x}'_i \boldsymbol{\beta}) + \lambda \sum_{j=1}^p |\beta_j|,$$

que é equivalente a minimizar:

$$-\frac{1}{n} \sum_{i=1}^n l(y_i, \beta_0 + \mathbf{x}'_i \boldsymbol{\beta})$$

sob a restrição $\sum_{j=1}^p |\beta_j| \leq \lambda$.

Métodos de regularização - o método *ridge*

- Para $\alpha = 0$, o método de regularização, baseado em penalização de segunda ordem, é denominado *regressão ridge*;

$$-\frac{1}{n} \sum_{i=1}^n l(y_i, \beta_0 + \mathbf{x}'_i \boldsymbol{\beta}) + \lambda \sum_{j=1}^p \beta_j^2,$$

que é equivalente a minimizar:

$$-\frac{1}{n} \sum_{i=1}^n l(y_i, \beta_0 + \mathbf{x}'_i \boldsymbol{\beta})$$

sob a restrição $\sum_{j=1}^p \beta_j^2 \leq \lambda$.

Métodos de regularização

- Valores de α no intervalo $(0, 1)$ também podem ser aplicados, ponderando as duas formas de penalização;
- O método lasso, diferentemente da regressão ridge, pode resultar em estimativas exatamente iguais a zero;
- Assim, além de ser usado como método de regularização, a regressão lasso pode ser usada também como técnica de seleção de covariáveis.