GET00211 - Modelos Lineares 2

Rafael Erbisti

Instituto de Matemática e Estatística

Universidade Federal Fluminense

Aula 07

Análise de Resíduos

- A análise de resíduos facilita a investigação de aspectos específicos do modelo.
- Detecta violações de pressupostos (por exemplo: estrutura de variância no modelo Normal).
- Identifica *outliers* e pontos influentes
- E guia ações corretivas: transformação de variáveis, alteração da função de ligação, alteração da família de distribuição.
- Resíduos utilizados nos MLGs: Resíduo de Pearson e Resíduo da Deviance.

Análise de Resíduos

Resíduo de Pearson

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{V}(Y_i)}}$$

- $\hat{\mu}_i$ é o valor esperado pela resposta para a observação i.
- $\widehat{V}(Y_i)$ é a variância estimada da resposta.
- Aqui, $X^2 = \sum_{i=1}^n r_i^2$, que é a estatística de qui-quadrado de Pearson para bondade de ajuste.

Resíduo da Deviance

$$d_i = \text{sinal}(y_i - \hat{\mu}_i)\sqrt{2\left\{\ell(\hat{\mu}_{\max,i}; y_i) - \ell(\hat{\mu}_i; y_i)\right\}}$$

- $\ell(\mu_{\max,i}; y_i)$ é a log-verossimilhança da do modelo maximal. Neste caso $\hat{\mu}_{\max,i} = y_i$.
- $\ell(\mu_i; y_i)$ é a log-verossimilhança da observação com parâmetro μ_i.
- Origina-se da contribuição de cada ponto para a Deviance total do modelo, ou seja, $D = \sum_{i=1}^n d_i^2$.

Análise de Resíduos

- Os resíduos padronizados são definidos por:

$$s_i^* = \frac{r_i}{\sqrt{1 - h_{ii}}}$$

  onde r_i é o resíduo para a observação i e h_{ii} mede o grau de influência da observação no ajuste do modelo (leverage) e é obtido por meio do i-ésimo elemento da diagonal da matriz hessiana $\mathbf{H}$.

- As análises de resíduos em MLG devem ser conduzidas de forma semelhante à dos modelos lineares normais.
- Vale lembrar que não há qualquer pressuposto sobre a distribuição dos resíduos. Neste caso, em um MLG não se deve avaliar normalidade do resíduos.
- Se os dados são binários, por exemplo, ou se n_i é pequeno para a maioria dos níveis das covariáveis, haverá poucos valores distintos dos resíduos e os gráficos serão pouco informativos.
- Neste caso, será necessário confiar na estatística agregada da bondade de ajuste (X^2 e D) e em outros diagnósticos.

Gráficos de Diagnóstico

- Resíduo (Pearson/deviance) vs valor ajustado $\hat{\mu}_i$: verificar padrões (curvas, heterocedasticidade)
- QQ-plot dos resíduos (quando aplicável): procurar desvios sistemáticos.
- Resíduo vs preditor (covariáveis): detectar tendência não modelada. Se o modelo não é adequado, os pontos podem apresentar padrões de comportamento, sugerindo que termos adicionais devem ser incluídos.
- Distância de Cook (DFBetas, DFFits) e Leverage: localizar pontos influentes.

Simulação de Dados para o Modelo Poisson

Exercício

Considere um modelo de regressão de Poisson com preditores x_1 e x_2. Simule um conjunto de dados com $n = 100$ observações a partir do modelo:

$$\log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i},$$

com $\beta_0 = 1$, $\beta_1 = 0.3$ e $\beta_2 = 0.1$, onde:

- $x_1 \sim \mathcal{N}(2, 1)$;
- $x_2 \sim \mathcal{N}(1, 1)$;
- $y_i \sim \text{Poisson}(\mu_i)$.

Gere os dados, ajuste um modelo de regressão de Poisson e realize a análise dos resíduos.

Métodos de Regularização

- Os métodos de regularização são técnicas usadas em modelos de regressão e aprendizado de máquina para evitar overfitting.

- Na regressão usual de um modelo normal, ajustamos o modelo para minimizar o erro quadrático.

- Problema: Se o número de preditores p for grande, ou se houver multicolinearidade (preditores correlacionados), o estimador $\hat{\boldsymbol{\beta}}$ pode ter alta variância e ser instável.

- Ideia da regularização: introduzir informação extra ou uma restrição/ penalização sobre os coeficientes $\boldsymbol{\beta}$, controlando seu tamanho ou complexidade, com objetivo de reduzir a variância do estimador, mesmo que isso introduza um pequeno viés.

Métodos de Regularização

De formal geral, para lidar com esses problemas, o critério de MQO é modificado:

Forma com restrição

$$\min_{\beta \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 \quad \text{sujeito a} \quad \beta \in C$$

onde C é um conjunto de restrições (tipicamente convexo).

Forma Penalizada

$$\min_{\beta \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + P(\beta)$$

onde $P(\cdot)$ é uma função de penalização (tipicamente convexa).

Princípio Fundamental

A regularização nos permite navegar no trade-off entre viés e variância: introduzimos um pequeno viés para reduzir drasticamente a variância do estimador.

Normas Canônicas e Penalização

Duas escolhas fundamentais para a penalização dão origem a métodos clássicos de regularização:

- Norma ℓ_2 (Ridge Regression): Penaliza a magnitude quadrada dos coeficientes:

$$P(\boldsymbol{\beta}) = \lambda \|\boldsymbol{\beta}\|_2^2 = \lambda \sum_{j=1}^{p} \beta_j^2$$

solução fechada, não produz esparsidade e reduz variância.

- Norma ℓ_1 (Lasso): Penaliza a magnitude absoluta dos coeficientes:

$$P(\boldsymbol{\beta}) = \lambda \|\boldsymbol{\beta}\|_1 = \lambda \sum_{j=1}^{p} |\beta_j|$$

convexa e esparsa; trade-off ótimo entre viés e variância.

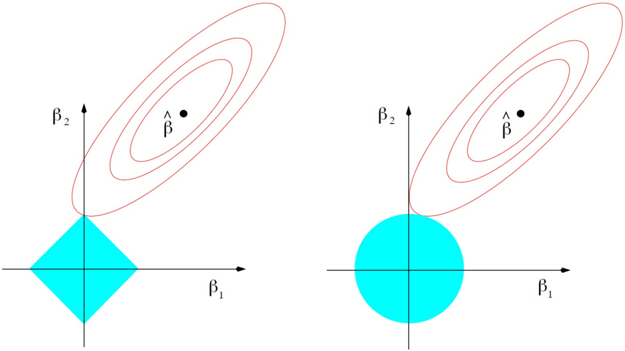
Propriedade 1: Esparsidade

Definição (Esparsidade)

Um estimador $\hat{\boldsymbol{\beta}}$ é esparso se muitos de seus componentes são exatamente zero ($\hat{\beta}_j = 0$). Isso implica uma seleção automática de variáveis.

- Lasso (ℓ_1): produz soluções esparsas. Quanto maior λ, maior a esparsidade. **Pode ser entendido, então, como um método de seleção de variáveis!**
- Ridge Regression (ℓ_2): não produz soluções esparsas. Tipicamente, todos os coeficientes são não nulos.

Por que ℓ_1 induz esparsidade e ℓ_2 não?



- Lasso (ℓ_1): $|\beta_1| + |\beta_2| \leq t$. A bola ℓ_1 tem quinas afiadas alinhadas com os eixos coordenados, levando a soluções esparsas.
- Ridge Regression (ℓ_2): $\beta_1^2 + \beta_2^2 \leq t$. A bola ℓ_2 é suave, sem quinas, tornando improvável que os coeficientes sejam exatamente zero.

Propriedade 2: Convexidade

Definição (Problema de Otimização Convexa)

Um problema é convexo se a função objetivo e as restrições de desigualdade são funções convexas, e as restrições de igualdade são afins.

- Lasso (ℓ_1) e Ridge (ℓ_2): são problemas de otimização convexos.

Por que a Convexidade é Importante?

- Ótimo Global: Qualquer mínimo local é também um mínimo global. Isso garante que o algoritmo converge para a melhor solução.
- Eficiência Computacional: Existem algoritmos muito eficientes e confiáveis para resolver problemas convexos.

Convexidade e Regularização em MLG

- No MLG, a log-verossimilhança $\ell(\boldsymbol{\beta})$ é **concava** em β.
- Portanto, o negativo da log-verossimilhança $-\ell(\boldsymbol{\beta})$ é convexo.
- Maximizar $\ell(\boldsymbol{\beta})$ é equivalente a minimizar $-\ell(\boldsymbol{\beta})$:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \; -\ell(\boldsymbol{\beta})$$

- Ao adicionarmos uma penalização convexa $P(\boldsymbol{\beta})$, obtemos:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \; -\ell(\boldsymbol{\beta}) + \lambda P(\boldsymbol{\beta})$$

- Se $P(\boldsymbol{\beta})$ é convexa (como ℓ_1 ou ℓ_2), o problema continua convexo.

Vantagens da Convexidade

- Ótimo global: não há mínimos locais espúrios.
- Métodos de regularização (Lasso, Ridge, Elastic Net) aplicáveis de forma direta em MLG.

Simulação de Dados para Regressão Logística

Objetivo: Simular um conjunto de dados para ilustrar modelos de regressão logística com múltiplos preditores, incluindo esparsidade (coeficientes nulos).

- Número de observações: $n = 200$
- Número de preditores: $p = 10$
- Matriz de preditores $\mathbf{X}$ gerada de forma aleatória, $X_{ij} \sim N(0, 1)$
- Vetor de coeficientes verdadeiros $\boldsymbol{\beta} = (2, -1.5, 0, 0, 1, 0, 0, 0.5, 0, 0)$
- Probabilidades verdadeiras calculadas pela função logística:

$$\pi_i = \frac{1}{1 + \exp(-\eta_i)}, \quad \eta_i = X_i \beta$$

- Variável resposta binária $y_i \sim \text{Bernoulli}(\pi_i)$
- Alguns coeficientes são zero para permitir avaliação de métodos de regularização (ex.: Lasso)