

CE071 - Análise de Regressão Linear

Cesar Augusto Taconeli

11 de junho, 2018

Aula 8 - Regressão polinomial

- Alguns métodos para a modelagem de relação não lineares entre variáveis:
 - Transformação nas variáveis;
 - Regressão segmentada;
 - Regressão não linear;
 - Regressão não paramétrica (ou semi paramétrica);
 - Regressão polinomial.
- Num modelo de regressão polinomial, uma mesma covariável quantitativa é inserida duas ou mais vezes (de diferentes formas) no modelo;
- Como exemplo, uma covariável x pode ser inserida na forma linear (x), quadrática (x^2), cúbica (x^3),... ou ainda por meio da interação com outras covariáveis (xy , xz),

Regressão polinomial em uma covariável

- O modelo de regressão polinomial de ordem k em uma variável fica especificado por:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k + \epsilon_i. \quad (1)$$

- Particularmente, o modelo polinomial de ordem 2 (quadrático) é definido por:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i. \quad (2)$$

- Assumindo que os erros têm média zero, a média de y , condicional a x , fica dada por uma função quadrática de x :

$$E(y|x) = \beta_0 + \beta_1 x + \beta_2 x^2 \quad (3)$$

Regressão polinomial em uma covariável

- A aplicação de modelo de regressão polinomial pode ser motivada, por exemplo:
 - 1 O analista pode saber, com base na teoria da área, que um modelo polinomial com uma particular ordem é adequado aos dados;
 - 2 A relação entre as variáveis pode apresentar um ponto ótimo (mínimo ou máximo), o que pode motivar a investigação de um modelo quadrático;
 - 3 Em boa parte dos casos, tem-se uma relação desconhecida e não linear entre as variáveis, de maneira que polinômios de diferentes ordens podem ser testados buscando um bom ajuste.

Regressão polinomial em uma covariável - ordem do modelo

- Caso não se disponha, de antemão, de um particular polinômio a ser aplicado, a ordem do polinômio deve ser escolhida de maneira criteriosa;
- Quanto maior a ordem do polinômio especificado, melhor o ajuste aos dados, mas menor a precisão das estimativas (*tradeoff bias-variance*);
- A menos que se tenha uma justificativa adequada para isso, polinômios de ordem superior a 3 devem ser evitados.

Regressão polinomial em uma covariável - ordem do modelo

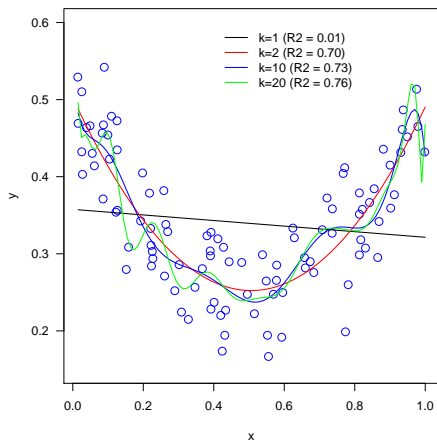


Figura 1: Extrapolação usando um modelo quadrático

Regressão polinomial em uma covariável - ordem do modelo

- A escolha da ordem do polinômio pode ser feita usando estratégias do tipo **forward** ou **backward**;
- Na estratégia **forward** deve-se começar pelo ajuste sem a covariável investigada e incluir, a cada passo, o termo linear (x), depois o quadrático (x^2), o cúbico (x^3),...
- O processo se encerra quando a adição de um termo de ordem $t + 1$ não for significativa, de acordo com o resultado do teste t. O modelo escolhido é o de ordem t .

Regressão polinomial em uma covariável - ordem do modelo

- Na estratégia **backward** deve-se começar pelo ajuste do modelo polinomial de maior ordem a ser investigado (digamos k) e, a cada passo, excluir o termo de maior ordem (primeiro x^k , depois x^{k-1}, \dots), desde que seja não significativo;
- O processo se encerra quando o termo de maior ordem (digamos t) tiver efeito significativo, de acordo com o resultado do teste t . O modelo escolhido é o de ordem t .

Regressão polinomial - extrapolação

- Extrapolar (ou seja, prever y fora do intervalo observado para x) em geral não é apropriado para modelos polinomiais, produzindo resultados incorretos ou mesmo absurdos.
- Uma situação em que a extrapolação pode ser aplicada com maior segurança é quando o modelo de regressão baseia-se em algum modelo teórico da área que descreva adequadamente o fenômeno.

Regressão polinomial - extrapolação

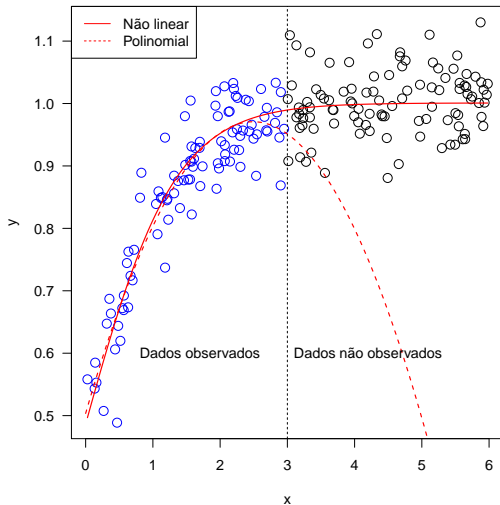


Figura 2: Ajustes de modelos polinomiais de diferentes ordens

Regressão polinomial - multicolinearidade

- Em regressão polinomial, a matriz do modelo é composta por colunas que são potências de uma mesma variável;
- Essa construção implica no mau condicionamento da matriz, ou seja, na quase dependência linear de suas colunas;
- Uma medida corretiva consiste em usar os dados centrados (ou seja, substituir x por $x^* = x - \bar{x}$);
- Uma alternativa mais eficiente é usar **polinômios ortogonais**.