Lista de Exercícios sobre Ajuste de Curvas

Prof.: Fabrício Murai

Informações importantes:

- Data de entrega: até 23:59 do dia 16/05/2018.
- Questões podem ser discutidas entre até três alunos. Nomes dos colegas precisam ser listados. Contudo, a escrita das soluções e submissão deve ser feita individualmente.
- Submissão deve ser feita em formato PDF através do Moodle, mesmo que tenham sido resolvidas a mão e escaneadas.
- Todas as soluções devem ser justificadas.
- Se puder, peço por favor que marque o tempo gasto para resolver a lista, para que o tamanho da lista de exercícios seja ajustado em semestres futuros.

1. Considere a função $f(x) = 0.5 + 2x$ e os pontos a seguir:

x	0	1	2	3	4
y	0.523	3.275	4.319	5.511	8.052

   Calcule o erro do ajuste, segundo:

   (a) o erro máximo $E_\infty(f)$;

   (b) o erro médio $E_1(f)$;

   (c) a raiz do erro médio quadrático $E_2(f)$.

2. **(Obrigatória para TB1, Extra para TN)** Usando o conjunto de dados `Cereals` apresentado em sala, escreva um programa que encontra a regressão linear simples rating $= \beta_0 + \beta_1 x$ que apresenta o menor desvio D, dentre todos os preditores $x \in$ protein, fat, sodium, fiber, carbo, sugars, potass, vitamins.

3. Utilizando o método dos quadrados mínimos, derive as equações que devem ser satisfeitas para fazer o ajuste das seguintes funções:

   (a) $f(x) = \frac{1}{2\beta}(e^{\beta x} - e^{-\beta x} - 2)$

   Dica: A equação final deve ser

$$\frac{1}{\beta^2}\sum_k (e^{\beta x_k} + e^{-\beta x_k} - 2)^2 + \frac{1}{\beta}\sum_k (e^{\beta x_k} + e^{-\beta x_k} - 2)x_k(e^{\beta x_k} - e^{-\beta x_k}) = \frac{1}{\beta}\sum_k y_k(e^{\beta x_k} + e^{-\beta x_k} - 2) + \sum_k y_k x_k(e^{\beta x_k} - e^{-\beta x_k})$$

   (b) $f(x) = \beta x$

4. Considere os pontos a seguir:

x	2.0	3.5	4.0	5.1	7.0
y	2.2	2.0	3.0	6.0	5.0

- Mostre o diagrama de dispersão destes pontos (pode ser feito à mão ou no computador).
- Usando o método dos quadrados mínimos, encontre os parâmetros da regressão linear simples $f(x) = \beta_0 + \beta_1 x$. **Atenção:** você não pode resolver esta questão usando um método que retorne os coeficientes da regressão.

5. Considere a série de pontos a seguir:

x	1	2	3	4	5	6
y	-4.501	83.453	112.953	123.824	170.335	183.008

Suponha que a relação entre x e y seja dada por $y = \beta_1 x + \beta_2 \log x + \epsilon$. Obtenha os valores de β_1 e β_2 através do método dos quadrados mínimos. (Dica: a função y pode ser vista como uma regressão linear múltipla em x, onde $x_1 = x$ e $x_2 = \log x$.)

6. A tabela a seguir mostra o número de semanas x_i que um candidato gastou estudando para um exame e a probabilidade y_i de passar no exame.

x_i	1	2	5	10
y_i	0.14	0.17	0.27	0.50

A função logística mapeia um número real t para um valor entre 0 e 1 e é definida por

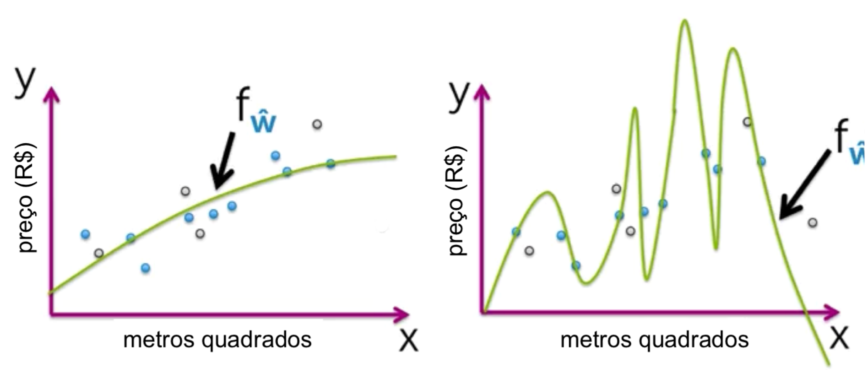
$$y = \frac{1}{1 - e^{-t}}.$$

Suponha que a relação entre o número de semanas x_i que um candidato i gastou estudando para o exame e a probabilidade y_i de i passar seja dada por uma função logística, onde $t_i = \beta_0 + \beta_1 x_i$.

(a) Determine as equações normais a serem resolvidas para obter β_0 e β_1 pelo método dos mínimos quadrados. **Dica: note que a função logística não é linear nos parâmetros. É necessário linearizar essa função.**

(b) Se $\beta_0 = -2$ e $\beta_1 = 0.2$, qual a probabilidade de passar no exame após estudar por 20 semanas?

7. Deseja-se usar a regressão polinomial $f(x_i) = w_0 + w_1 x_i + w_2 x_i^2 + \ldots + w_p x_i^p$ para estimar a relação entre a metragem (em m^2) de um imóvel e o seu preço em um bairro de Belo Horizonte. As figuras abaixo ilustram (não são uma representação exata) os resultados obtidos para $p = 3$ e $p = 7$, respectivamente. Qual das regressões possui o menor desvio? Qual dos valores de p é mais adequado e por quê?



8. Considere as situações a seguir e assinale **I** quando a interpolação polinomial é mais adequada, **R** naquelas em que a regressão é preferível e **A** quando ambas são equivalentes (isto é, geram o mesmo resultado e tem o mesmo custo computacional).

( ) Quando deseja-se descobrir uma fórmula para $\sum_{i=1}^{n} i^3$.

( ) Quando são fornecidos os pontos (x_i, y_i, z_i) e quer-se aproximar uma função $z = f(x, y)$.

( ) Quando são fornecidos $n = 100$ pontos (x_i, y_i), sem erros de medição, e sabe-se que y é um polinômio de quinto grau em x.

( ) Dados 10 pontos, deseja-se obter uma função $f(x) = \beta_0 + \beta_1 x + \ldots + \beta_9 x^9$ que aproxima a função desconhecida.

9. Sobre coeficiente de determinação R^2. **Atenção:** você não pode resolver esta questão usando um método que retorne o R^2.

   (a) Calcule o coeficiente de determinação para a regressão linear simples obtida na Questão 4.

   (b) A regressão é uma boa aproximação para a função desconhecida? Explique.

   (c) Sem fazer nenhuma conta, o que podemos dizer sobre o coeficiente de determinação da regressão $f(x) = \alpha_0 + \alpha_1 x + \alpha_2 \text{sen}(2\pi x)$, obtida pelo método dos mínimos quadrados?

   (d) Isso quer dizer que a função anterior é melhor ou pior do que a regressão linear simples?

   (e) Caso não seja possível afirmar nada, qual seria uma maneira mais adequada de comparar as duas regressões?