

AVALIAÇÃO TEÓRICA – ESTATÍSTICA AVANÇADA

1. Com relação ao processo de criar variáveis ‘dummy’, no pré-processamento de variáveis categóricas para regressão linear, é correto afirmar:

- a) A quantidade de novas colunas será igual à quantidade de níveis da variável original.
- b) A quantidade de novas colunas será igual ao dobro de níveis da variável original.
- c) Não é necessário transformar as variáveis categóricas em variáveis dummy.
- d) A quantidade de novas colunas será igual à quantidade de níveis da variável original menos 1. RESPOSTA CORRETA.**

e) A quantidade de novas colunas será igual à quantidade de níveis da variável original mais 1.

2. Com relação ao processo de escalonamento de variáveis contínuas, no pré-processamento de variáveis contínuas para regressão linear, é correto afirmar:

- a) Não é necessário fazer o escalonamento das variáveis contínuas.
- b) O processo se dá dividindo cada valor da variável pelo seu desvio padrão.
- c) O processo se dá subtraindo de cada valor da variável a sua média.
- d) O processo se dá multiplicando cada valor da variável por 2.

e) O processo se dá primeiro subtraindo de cada valor da variável e depois dividindo o resultado pelo seu desvio padrão. RESPOSTA CORRETA.

3. Com relação ao pré-processamento da resposta para regressão linear, é correto afirmar que:

- a) Não é necessário avaliar a resposta antes de fazer a regressão.
- b) Deve-se avaliar a normalidade da resposta e se ela não for normal, tentar transformá-la para tentar fazer com que ela seja o mais próximo da normal. RESPOSTA CORRETA.**
- c) A resposta pode ter qual distribuição da família exponencial.
- d) A resposta deve estar entre +1 e -1.
- e) Devemos somente transformar a resposta caso ela possua algum valor maior que 1000.

4. Com relação ao processo de ajuste do modelo, é correto afirmar:

- a) O método dos mínimos quadrados é exclusivo para regressão linear e o método do gradiente descendente por ser utilizado para qualquer modelo, inclusive a regressão linear. RESPOSTA CORRETA.**
- b) Não devemos nunca utilizar o método dos mínimos quadrados.
- c) Não devemos nunca utilizar o método do gradiente descendente.

d) Para regressão linear, os métodos de mínimos quadrados e de gradiente descendentes trarão resultados muito diferentes.

e) Na utilização de ambos os métodos dos mínimos quadrados e gradiente descendente é necessário especificar o 'alpha' (taxa de aprendizado) antes de utilizá-los.

5. Em relação ao cálculo das métricas de qualidade de ajuste (R^2 e raiz do erro quadrado médio) é correto afirmar:

a) Eles devem ser calculados com os dados de ajuste.

b) Eles devem ser calculados com os dados de teste. RESPOSTA CORRETA.

c) Tanto faz qual é o conjunto de dados utilizado para avaliação dessas métricas.

d) Não é possível estimar as métricas com os dados que possuímos, somente com dados que serão gerados após o modelo estar pronto.

e) O mínimo aceitável para o R^2 é 0.98.

6. Em relação a análise de resíduos para regressão linear, é correto afirmar:

a) Após a regressão, os resíduos podem ter qualquer distribuição de probabilidade.

b) Após a regressão, espera-se que os resíduos tenha uma distribuição normal com média zero; sejam independente e homoscedásticos. RESPOSTA CORRETA.

c) Não é necessário realizar análise de resíduos para regressão linear.

d) A heterocedasticidade é a propriedade do resíduo ter variância constante ao longo do espaço de inferência.

e) Não podemos utilizar métodos gráficos para avaliar os resíduos após a regressão linear.

7. Com relação às séries temporais, é correto afirmar que:

a) Modelar séries temporais é exatamente igual a modelar qualquer tipo de dado.

b) Séries temporais nunca exibem nenhum comportamento sistemático.

c) Uma série que possui tendência pode ser chamada de estacionária.

d) Antes de modelar séries temporais, é necessário avaliar quais componentes sistemáticos (como tendência, sazonalidade, homoscedasticidade e autocorrelação) ela possui. RESPOSTA CORRETA.

e) A propriedade de homoscedasticidade não é importante para modelagem de séries temporais.

8. Em relação à regressão regularizada, é correto afirmar:

a) Ela sempre resulta nos mesmos parâmetros da regressão linear ordinária.

b) É necessário decidir o valor do coeficiente de penalização (lâmbida) antes do modelo. Essa definição deve ser feita de modo que o modelo apresente as melhores métricas de qualidade de ajuste, calculadas com os dados de teste. RESPOSTA CORRETA.

c) O valor do coeficiente de penalização é sempre 0.3.

d) As regularizações 'lasso' e 'ridge' sempre resultam no mesmo modelo.

e) Regressões regularizadas não conseguem tratar dados que possuem multicolinearidade.