

Medidas de erros em problemas de regressão

@CursoDS_ProfDanilo

Prof. Dr. Danilo Morales Teixeira

Sites e contatos

- Slack: pythonprofdanilo.slack.com
- GitHub: https://github.com/danmorales/CursoDS_ProfDanilo
 - E-mail: cursods.profdanilo@gmail.com
- Facebook: <https://www.facebook.com/profdanilods/>

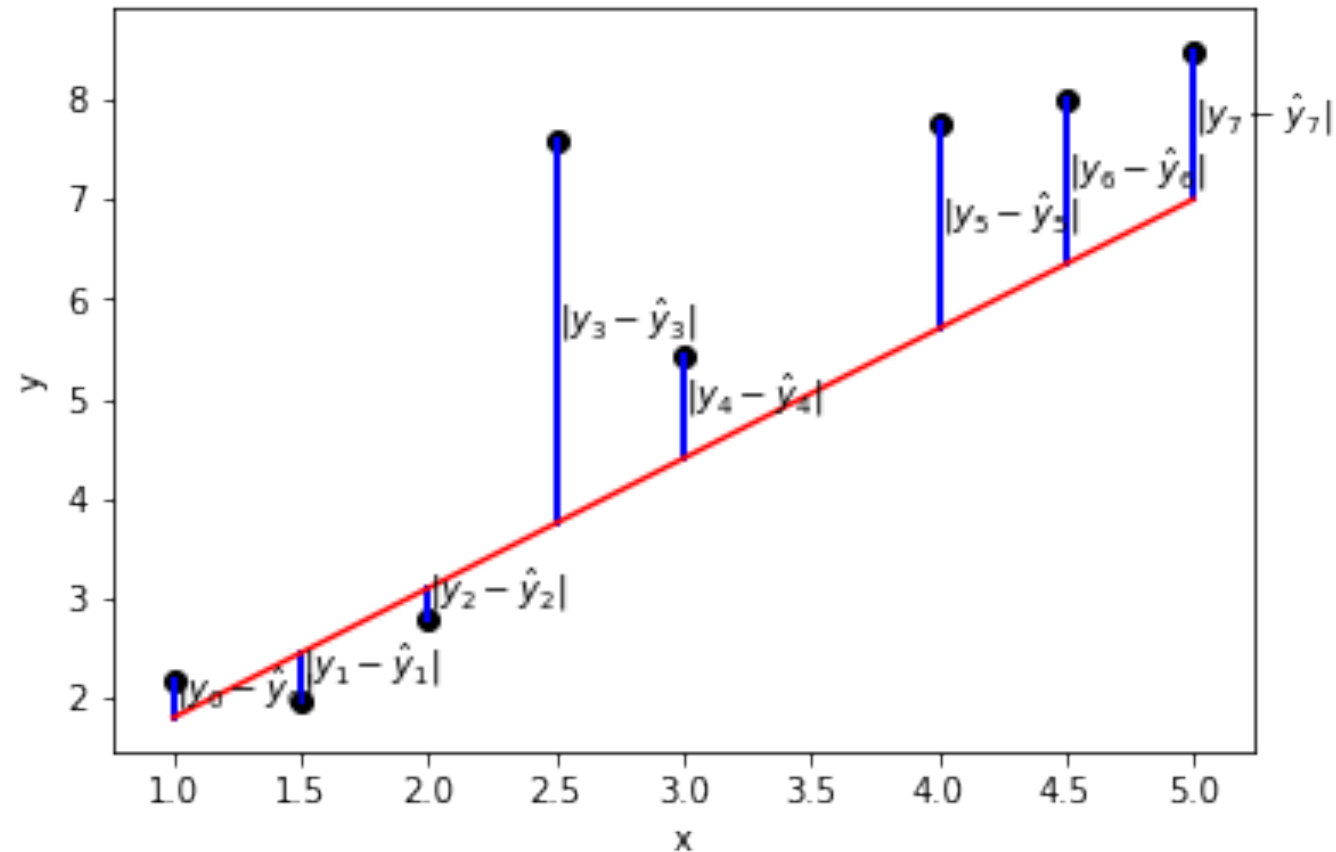
Conteúdo

- Erro absoluto médio
- Erro absoluto percentual médio
 - Erro quadrático médio
 - Raiz do erro quadrático médio
- Erro logarítmico quadrático médio
- Raiz do erro logarítmico quadrático médio

Erro absoluto médio - definição

- Erro absoluto médio (MAE) é a medida da diferença entre duas variáveis contínuas
- Seja Y os valores reais e \hat{Y} os valores previstos de um determinado evento
 - MAE é a distância vertical média entre os pontos
 - MAE utiliza a mesma escala de medida dos dados

Erro absoluto médio - definição



Erro absoluto médio - definição

Erro absoluto médio é definido como:

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

Erro absoluto médio - aplicação

Y	\hat{Y}	$ Y - \hat{Y} $
2.45	2.47	0.02
2.38	2.35	0.03
2.41	2.42	0.01
2.39	2.35	0.04
2.15	2.18	0.04

$$MAE = \frac{\sum_{i=1}^n |Y - \hat{Y}|}{N} = \frac{0.14}{5} = 0.028$$

Erro absoluto médio percentual – definição

Erro absoluto médio percentual médio (MAPE) define a acurácia dos valores previstos por um determinado modelo

$$MAPE = \frac{100\%}{N} \sum_{i=0}^N \frac{|y_i - \hat{y}_i|}{\hat{y}_i}$$

Erro absoluto médio percentual - aplicação

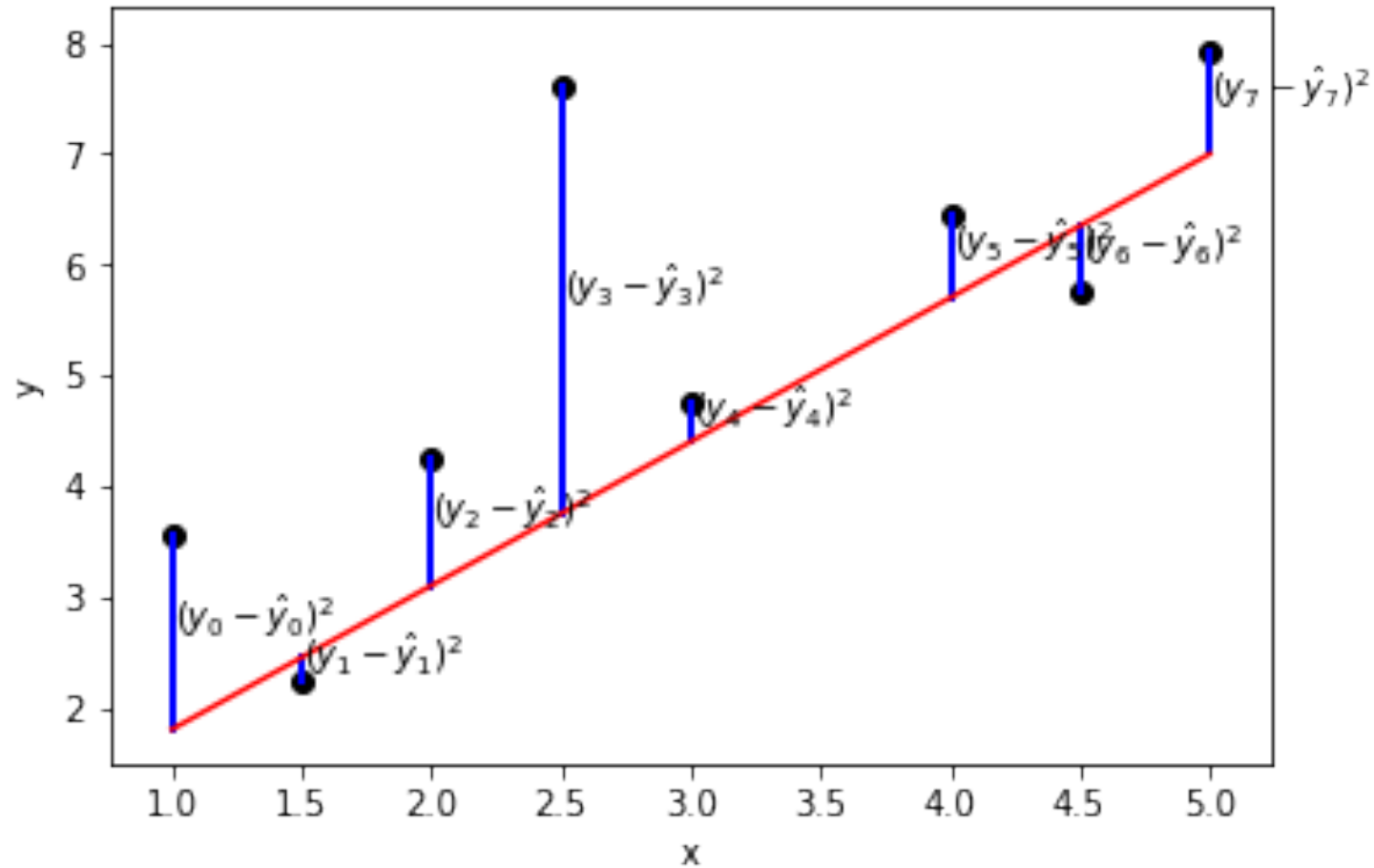
Y	\hat{Y}	$ Y - \hat{Y} $	$ Y - \hat{Y} /Y$
2.45	2.47	0.02	0.0082
2.38	2.35	0.03	0.0126
2.41	2.42	0.01	0.0041
2.39	2.35	0.04	0.0167
2.15	2.18	0.04	0.0186

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \frac{|Y - \hat{Y}|}{Y} = \frac{100\%}{5} \times 0.0602 = 1.2\%$$

Erro quadrático médio - definição

- Erro quadrático médio (MSE) mede a distância quadrática entre o valor real Y e o valor previsto \hat{Y}
 - Sempre positivo
 - Quanto mais próximo de zero melhor
- Incorpora a variância, ou seja, quão espalhados as estimativas estão dos dados verdadeiros
- Incorpora o bias, ou seja, quão distante os valores médios estimados estão dos valores verdadeiros
 - Penaliza mais os valores mais afastados dos valores reais

Erro quadrático médio - definição



Erro quadrático médio - definição

Erro quadrático médio é definido como:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

Erro quadrático médio - aplicação

Y	\hat{Y}	$ Y - \hat{Y} $	$ Y - \hat{Y} ^2$
2.45	2.47	0.02	0.004
2.38	2.35	0.03	0.009
2.41	2.42	0.01	0.001
2.39	2.35	0.04	0.0016
2.15	2.18	0.04	0.0016

$$MSE = \frac{\sum_{i=1}^n (Y - \hat{Y})^2}{N} = \frac{0.0172}{5} = 0.00344$$

Raiz do erro quadrático médio - definição

- Raiz do erro quadrático médio (RMSE) é a raiz quadrada do MSE
 - Definido como sendo:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

Raiz do erro quadrático médio - aplicação

Y	\hat{Y}	$ Y - \hat{Y} $	$ Y - \hat{Y} ^2$
2.45	2.47	0.02	0.004
2.38	2.35	0.03	0.009
2.41	2.42	0.01	0.001
2.39	2.35	0.04	0.0016
2.15	2.18	0.04	0.0016

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y - \hat{Y})^2}{N}} = \sqrt{\frac{0.0172}{5}} = \sqrt{0.00344} = 0.059$$

Erro logarítmico quadrático médio - definição

- Erro logarítmico quadrático médio (MSLE) é uma medida da razão entre os valores reais e previstos
 - MSLE se importa apenas com a diferença percentual
 - MSLE trata diferenças pequenas e grandes da mesma forma
- Penaliza mais os valores subestimados do que os superestimados
 - Introduz uma assimetria na curva de erro

Erro logarítmico quadrático médio - definição

MSLE é definido como:

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^N (\log(y_i + 1) - \log(\hat{y}_i + 1))^2$$

Erro logarítmico quadrático médio - aplicação

X	Y	Log((X+1)/(Y+1))	[Log((X+1)/(Y+1))]²
1.00	1.02	-0.00432	1.86624e-05
1.30	1.21	+0.01733	0.000300328
1.40	1.45	-0.00895	8.01025e-05
1.45	1.32	+0.02367	0.0005602689
1.34	1.39	-0.00918	8.4272400000000001e-05
1.41	1.51	-0.01765	0.000311522
1.59	1.42	+0.02948	0.000869070

$$MSLE = \frac{0.002224228}{7} = 0.00032$$

Raiz do erro logarítmico quadrático médio - definição

É a raiz quadrada do MSLE

Definido como sendo:

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(y_i + 1) - \widehat{(\log(y + 1))})^2}$$

Raiz do erro logarítmico quadrático médio - aplicação

X	Y	Log((X+1)/(Y+1))	[Log((X+1)/(Y+1))]²
1.00	1.02	-0.00432	1.86624e-05
1.30	1.21	+0.01733	0.000300328
1.40	1.45	-0.00895	8.01025e-05
1.45	1.32	+0.02367	0.0005602689
1.34	1.39	-0.00918	8.4272400000000001e-05
1.41	1.51	-0.01765	0.000311522
1.59	1.42	+0.02948	0.000869070

$$RMSLE = \sqrt{\frac{0.002224228}{7}} = \sqrt{0.00032} = 0.0178$$

Aplicações

Vejam os vídeos da descrição com aplicações implementando as equações do zero e utilizando a biblioteca SciKit-Learn do Python

Obrigado

- Gostou? Deixe o seu like
- Curta o canal para receber notificações de novos vídeos
- Dúvidas? Mande um e-mail ou deixe o seu comentário aqui no vídeo