

Processo de pipeline do projeto

Análise descritiva

- separa train test **antes** da descritiva
- na análise dos hist sempre buscar explicação p' os fenômenos
 - ↳ caso ñ encontre, leva p' o homem dos negócios
- ↳ quando cauda muito longa a direita, log pode ser uma ferramenta poderosa, para melhor entendimento pelo modelo

Análise predição

Root Mean Squared Error

$$\begin{aligned}\text{Error} &= \text{valor} - \text{valor} \\ &\quad \text{predito} \quad \text{real} \\ &= \hat{y} - y\end{aligned}$$

$$\begin{aligned}\text{Mean squared error} &= \text{MSE} = \frac{1}{\text{①}} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \\ &\quad \text{n° examples}\end{aligned}$$

$$\begin{aligned}\text{Squared error} &= (\hat{y} - y)^2 \\ \text{error}\end{aligned}$$

$$\text{Root Mean Squared Error} = \text{RMSE} = \sqrt{\text{MSE}}$$

y = median house value

* por ter cauda pesada a direita é interessante utilizar um $(\log y)$

$$\therefore y = \log(\text{preço})$$

↳ \hat{y} não predições de $\log(\text{preço})$

SE \hat{y} está a 1 unidade do RMSE

$$\hat{y} = y + \text{RMSE}$$

$$\text{↳ } \log(\hat{\text{preço}}) = \log(\text{preço}) + \text{RMSE}$$

$$\text{RMSE} = \log(\hat{\text{preço}}) - \log(\text{preço})$$

$$\text{RMSE} = \log_{10}\left(\frac{\hat{\text{preço}}}{\text{preço}}\right)$$

$$10^{\text{RMSE}} = 10^{\log_{10}\left(\frac{\hat{\text{preço}}}{\text{preço}}\right)}$$

$$\frac{\hat{\text{preço}}}{\text{preço}} = 10^{\text{RMSE}}$$

Conclusão:

$$\begin{aligned}\text{Er percentual} &= \left(\frac{\hat{\text{preço}} - \text{preço}}{\text{preço}}\right) \cdot 100 \\ &= 100 \cdot (10^{\text{RMSE}} - 1)\end{aligned}$$

* indiretamente o modelo vai dar + peso para peso para erros maiores

Abordagens sobre y

① $y = f(x) + \epsilon$

② $y \sim P(y|x)$

$\epsilon = \text{média}$

se erro MSE: $f(x) = E_{y|x}[y]$

se erro MAE: mediana $(y|x)$

$$\begin{array}{l} \text{erro} \\ \text{MSE} \end{array} = \begin{array}{l} \text{erro} \\ \text{intrínseco} \end{array} + \begin{array}{l} \text{viés} \\ \text{under} \\ \text{fitting} \end{array} + \begin{array}{l} \text{variância} \\ \text{overfitting} \end{array}$$