JAVIER GALEANO, MARZO DE 2022

Notas escépticas sobre el Machine Learning

Lecture 3: Regression Problems

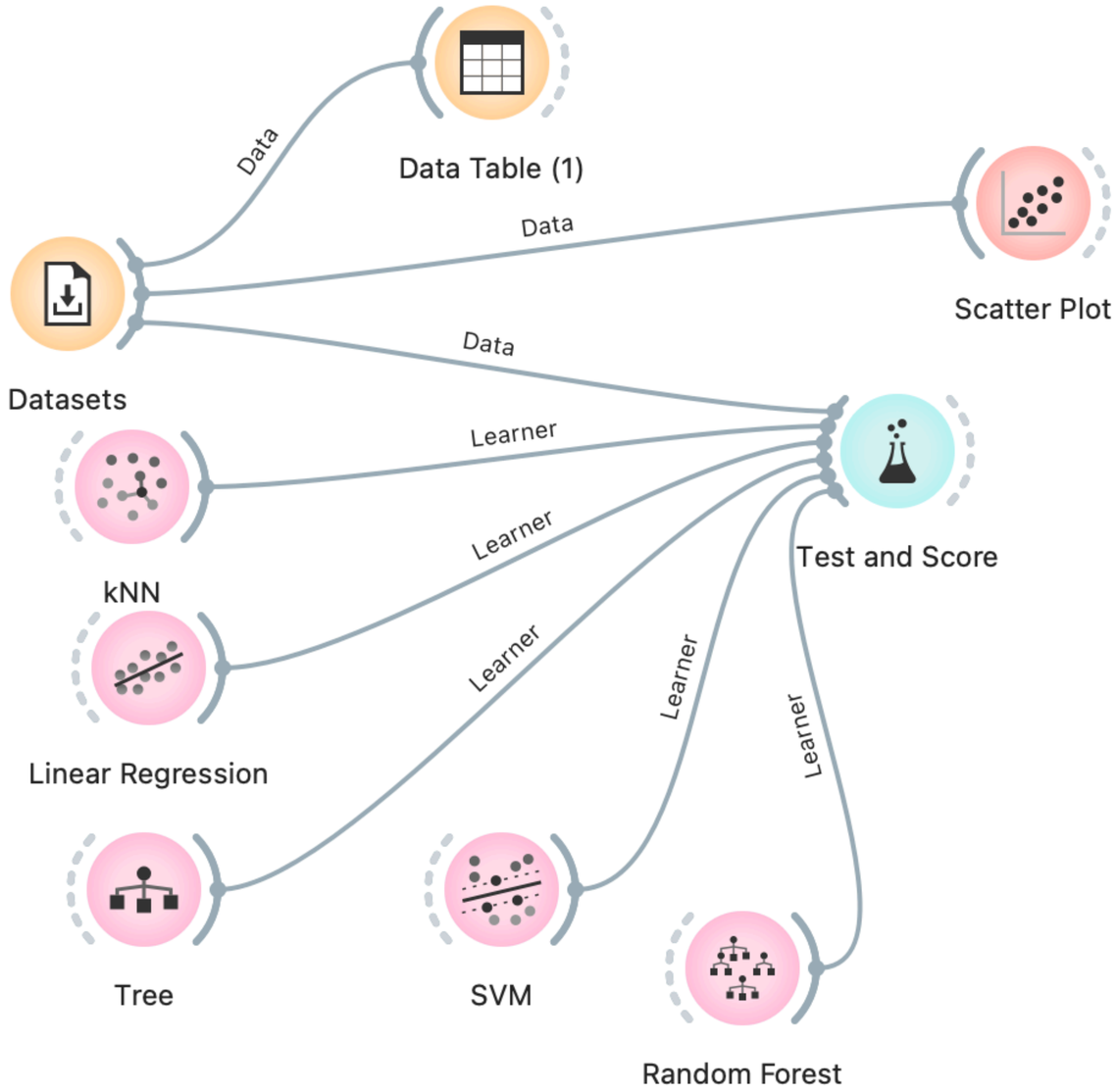
Regression problems in ML

Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Address	DebtIncomeRatio
1	41	2	6	19	0.124	1.073	NBA001	6.3
2	47	1	26	100	4.582	8.218	NBA021	12.8
3	33	2	10	57	6.111	5.802	NBA013	20.9
4	29	2	4	19	0.681	0.516	NBA009	6.3
5	47	1	31	253	9.308	8.908	NBA008	7.2
6	40	1	23	81	0.998	7.831	NBA016	10.9
7	38	2	4	56	0.442	0.454	NBA013	1.6
8	42	3	0	64	0.279	3.945	NBA009	6.6
9	26	1	5	18	0.575	2.215	NBA006	15.5
10	47	3	23	115	0.653	3.947	NBA011	4
11	44	3	8	88	0.285	5.083	NBA010	6.1
12	34	2	9	40	0.374	0.266	NBA003	1.6

23/03/2022

Regression

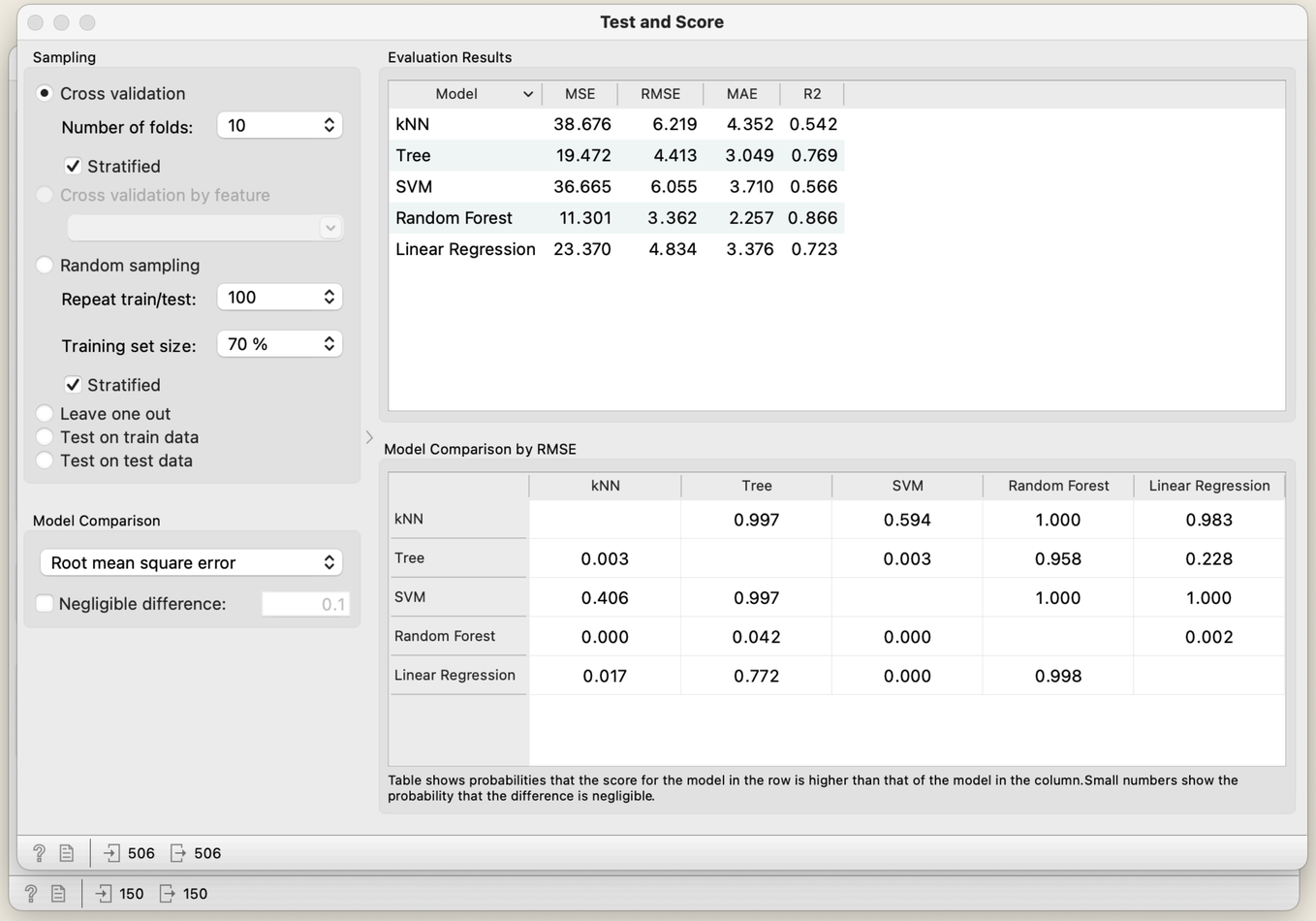
In Regression, the goal is to predict a *continuos value*



23/MARZO/2022

Scoring

- Housing data (numeric Target)
- Test & Score



Evaluation metrics in classification

- We are talking about some metrics:
  - MSE: Mean Squared Error
  - RMSE: Root Mean Squared Error
  - MAE: Mean Absolute Error
  - R^2

Error coefficients

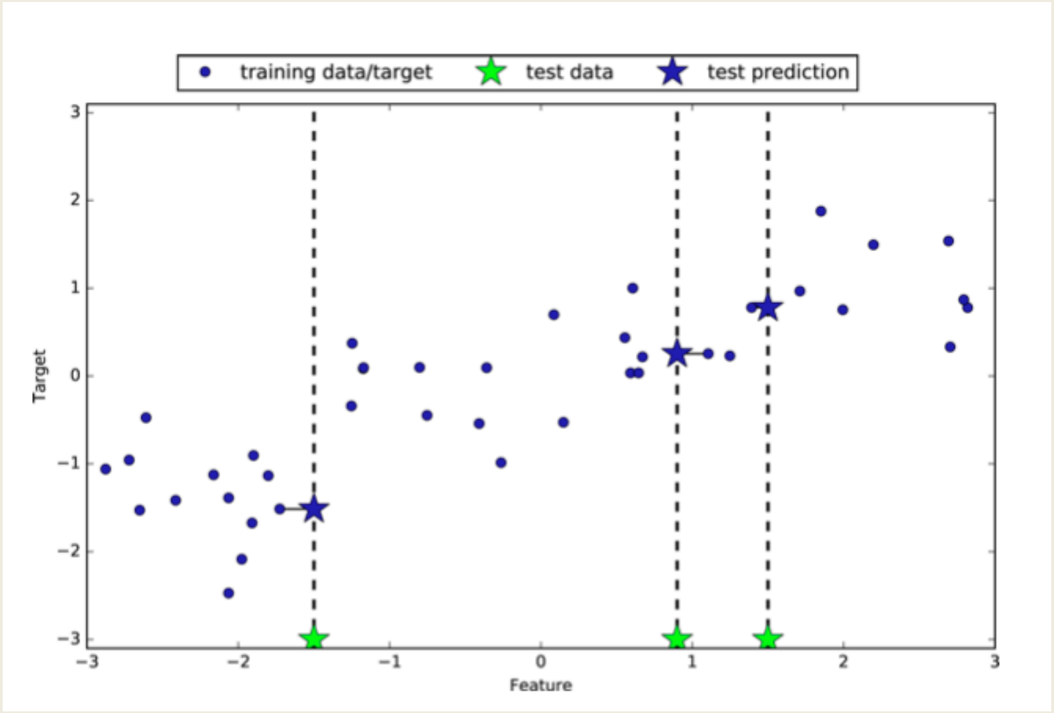
$$\mathrm{MSE} = \frac{1}{n}\sum_{i-1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

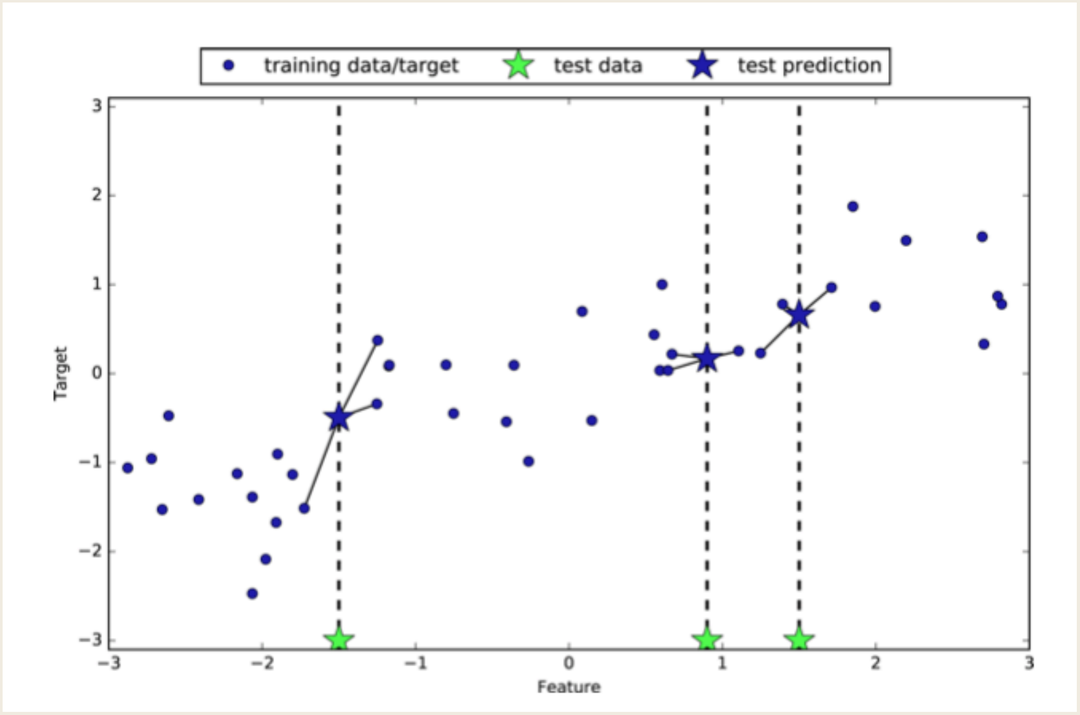
$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i-1}^{n}\left(Y_i - \hat{Y}_i\right)^2}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}\left|y_i - x_i\right|}{n} = \frac{\sum_{i=1}^{n}\left|e_i\right|}{n}$$

$$R^2 = 1 - \frac{SS_{\mathrm{res}}}{SS_{\mathrm{tot}}} = 1 - \frac{\sum_i\left(y_i - \hat{y}_i\right)^2}{\sum_i\left(y_i - \bar{y}\right)^2}$$

Regression Algorithms

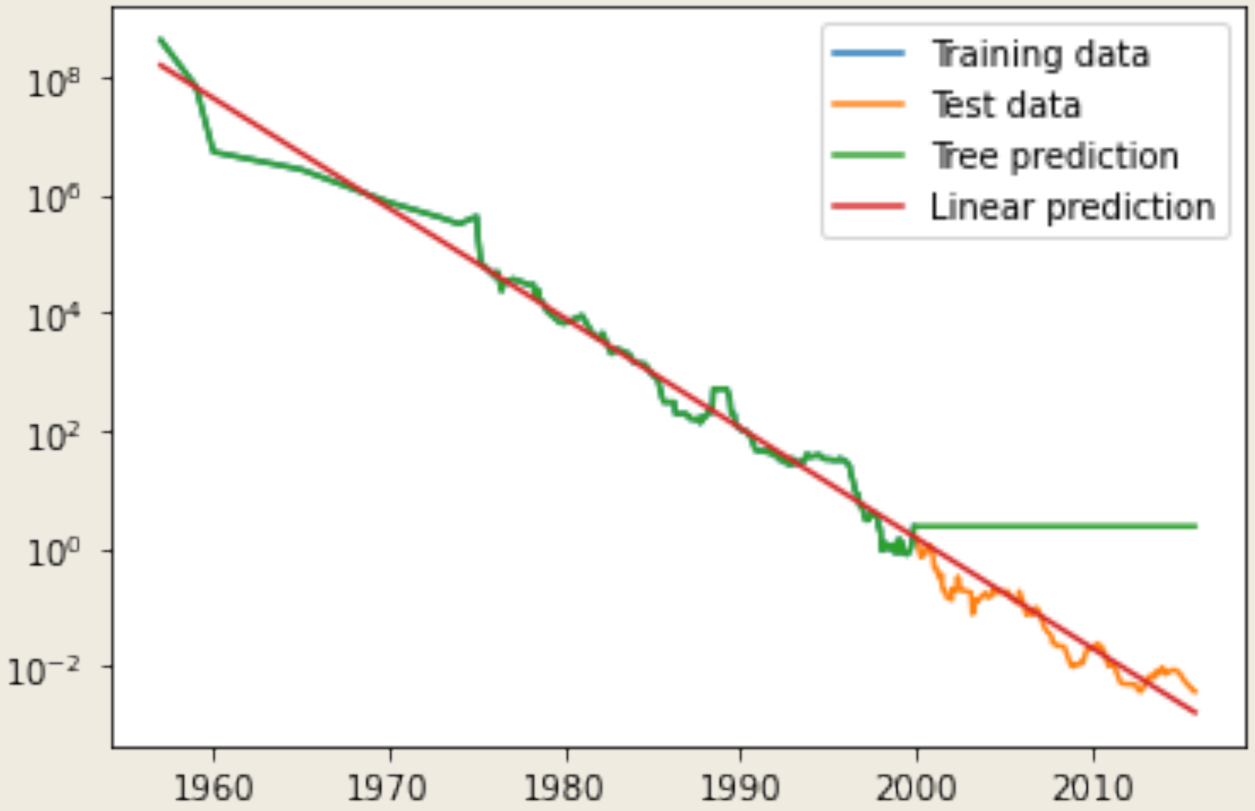




16/03/2022

K-Nearest Neighbors

- Building the k-NN algorithm consists only of storing the training dataset.
- To make a prediction for a new data point, the algorithm finds the closest data points in the training dataset—its “nearest neighbors.”



23/MARZO/2022

Tree vs Linear prediction

- Using ram_prices