

Data instances: 13
Features: 1
Meta attributes: 1

	name	coef
1	intercept	42.0243
2	Crime rate	-0.112226
3	Residential land	0.0567925
4	Industry	-0.0313188
5	River	2.78605
6	Nitric oxide	-18.0551
7	Rooms	3.51711
8	Age	0.0124839
9	Employment	-1.62943
10	Highways	0.296262
11	Tax	-0.0143417
12	Pupil-teacher ratio	-0.860996
13	Demographic	-0.579762

1. The near 0 values (within 0.1) indicate that it is probably no relation, e.g. how old buildings are don't tend to impact the price of buildings. Nitrous oxide levels have a strong relation to house value (-18.0551) which magnitude is much larger than any other so the correlation much be very strong between the price of houses and lack of Nitrous oxide concentration. The amount of rooms in a house increases with the price, which is what you would expect.

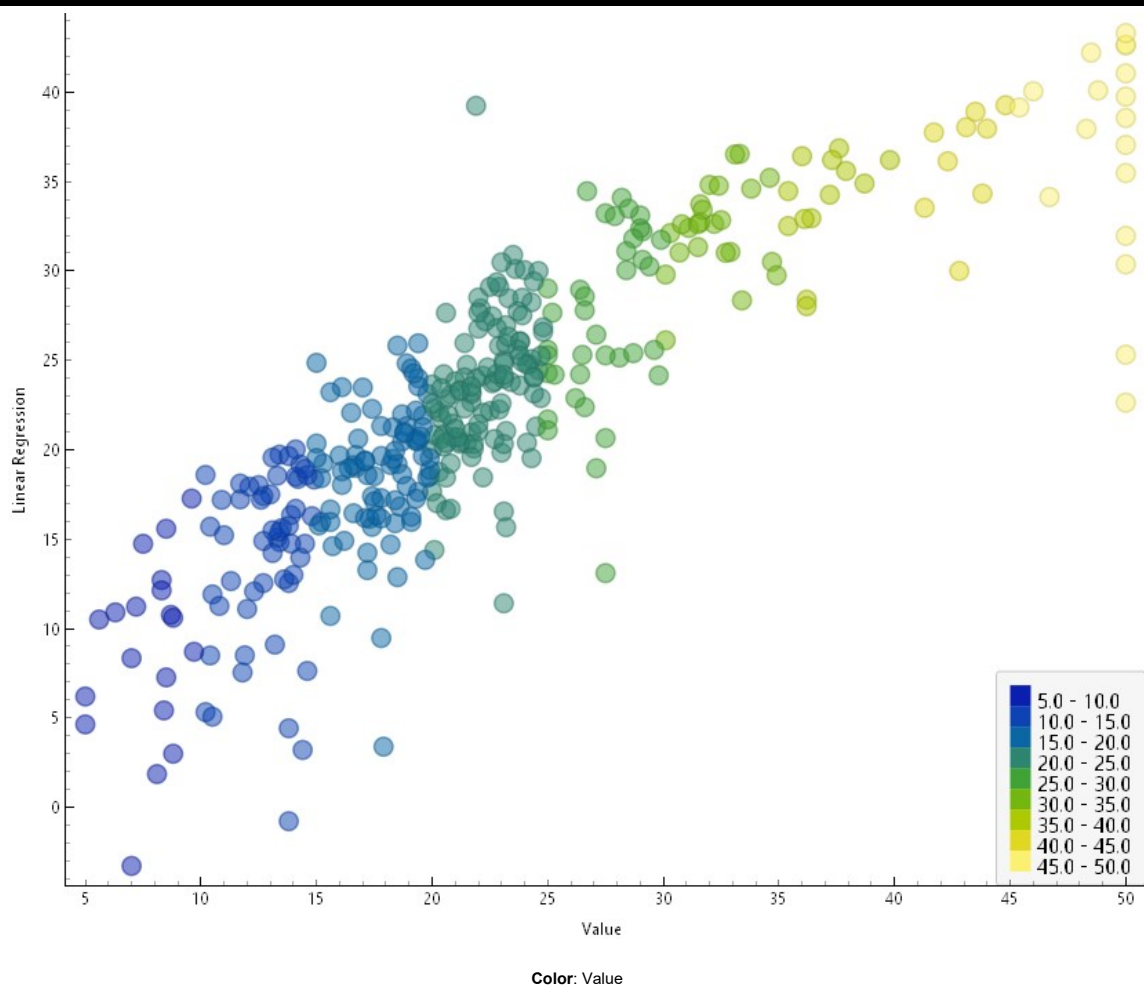
Settings

Sampling type: Stratified 5-fold Cross validation

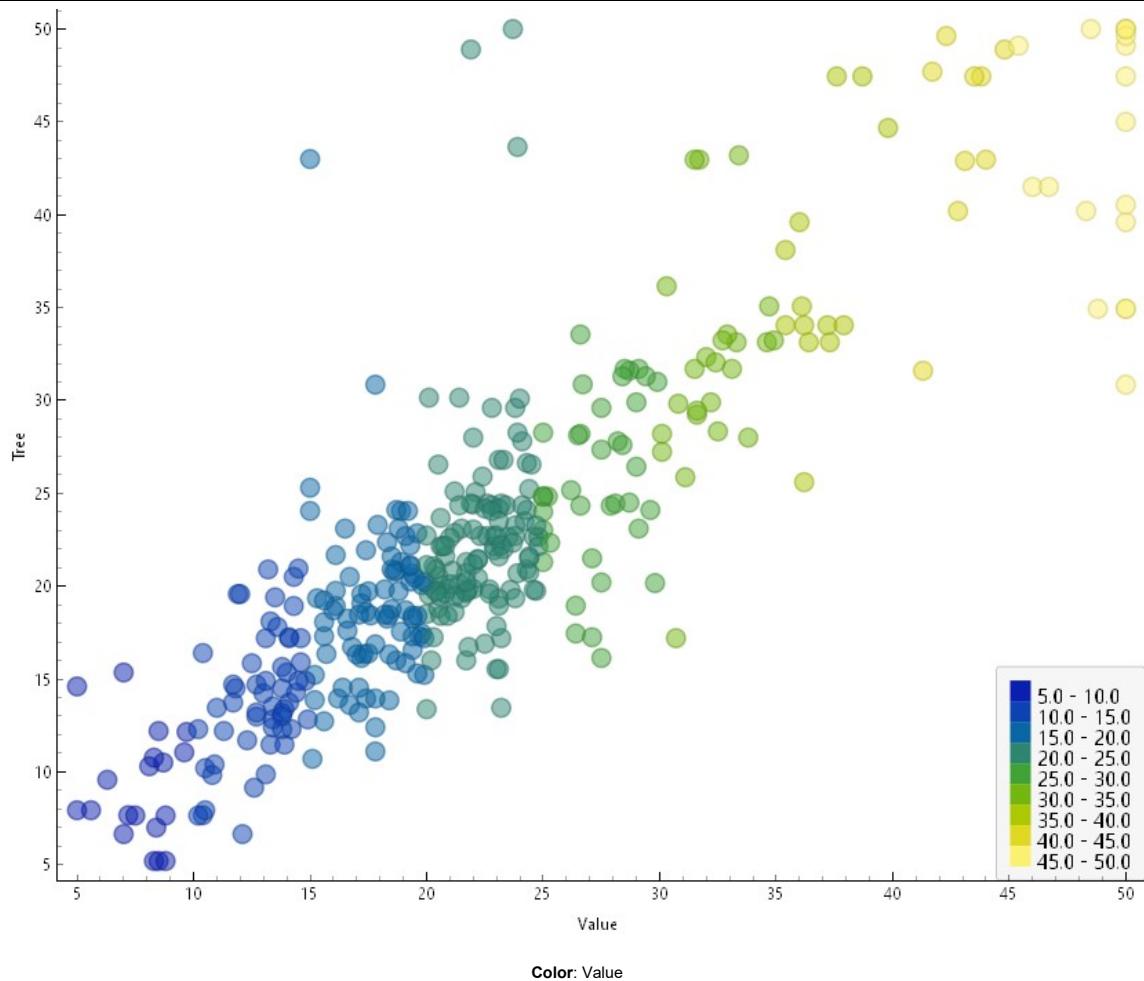
Scores

Model	MSE	RMSE	MAE	MAPE	R2
Linear Regression	25.195	5.019	3.607	18.186	0.721
Tree	24.377	4.937	3.296	16.492	0.730

2. The linear regression model would be a better model. I would expect the model to be right more than tree model most of the time but not all of the time.



The linear regression performs well in the mid range from about 20 - 35 but outside of the range it is more likely to undervalue houses. There are outliers when the value is 50, this is because it appears that there are multiple points in the dataset where the value was set to 50. It struggles when value is less than 15. There are also two negative values in this range which shouldn't be possible.



The linear regression performs very well to about a value of 25 with minimal outliers, but as the value increases from there the accuracy decreases dramatically. There are a few outliers when the predicted value is much higher than the real value. It struggles when value is more than 35.

The linear regression is better at guessing the value over the whole range while the tree model is much more likely to be more accurate with lower values

4. Classification tries to assign to a category while regression assigns to a continuous value. Classification is generally easier to evaluate than regression because it doesn't require as much precision so the accuracy is higher.

5. Ridge regression is a method to reduce over fitting due to the two variables being correlated. It's helpful here because the two variables are related and there are a high number of other variables that are being taken into account.