Salary predictor for cities

This is my city: how much should I earn?

Features considered - sources

Estimated Population (2017-07-01)

Land Area (2016-01-01)

} US Census Bureau

Features considered - sources

Estimated Population (2017-07-01)

Land Area (2016-01-01)

} US Census Bureau

Things to do → tripadvisor.com

Features considered - sources

Estimated Population (2017-07-01)

Land Area (2016-01-01)

} US Census Bureau

Things to do → tripadvisor.com

Rent appartment → apartmentlist.com

Features considered - sources

Estimated Population (2017-07-01)

Land Area (2016-01-01)

} US Census Bureau

Things to do → tripadvisor.com

Rent appartment → apartmentlist.com

Murder rate

Assault rate

Median Income

} city-data.com

Method

272 data points

Method

272 data points

Polynomials up to order 4

Method

272 data points

Polynomials up to order 4

Ridge, Lasso and Elastic Net with CV

Method

272 data points

Polynomials up to order 4

Ridge, Lasso and Elastic Net with CV

Mean square error of test as indicator

Method

272 data points

Polynomials up to order 4

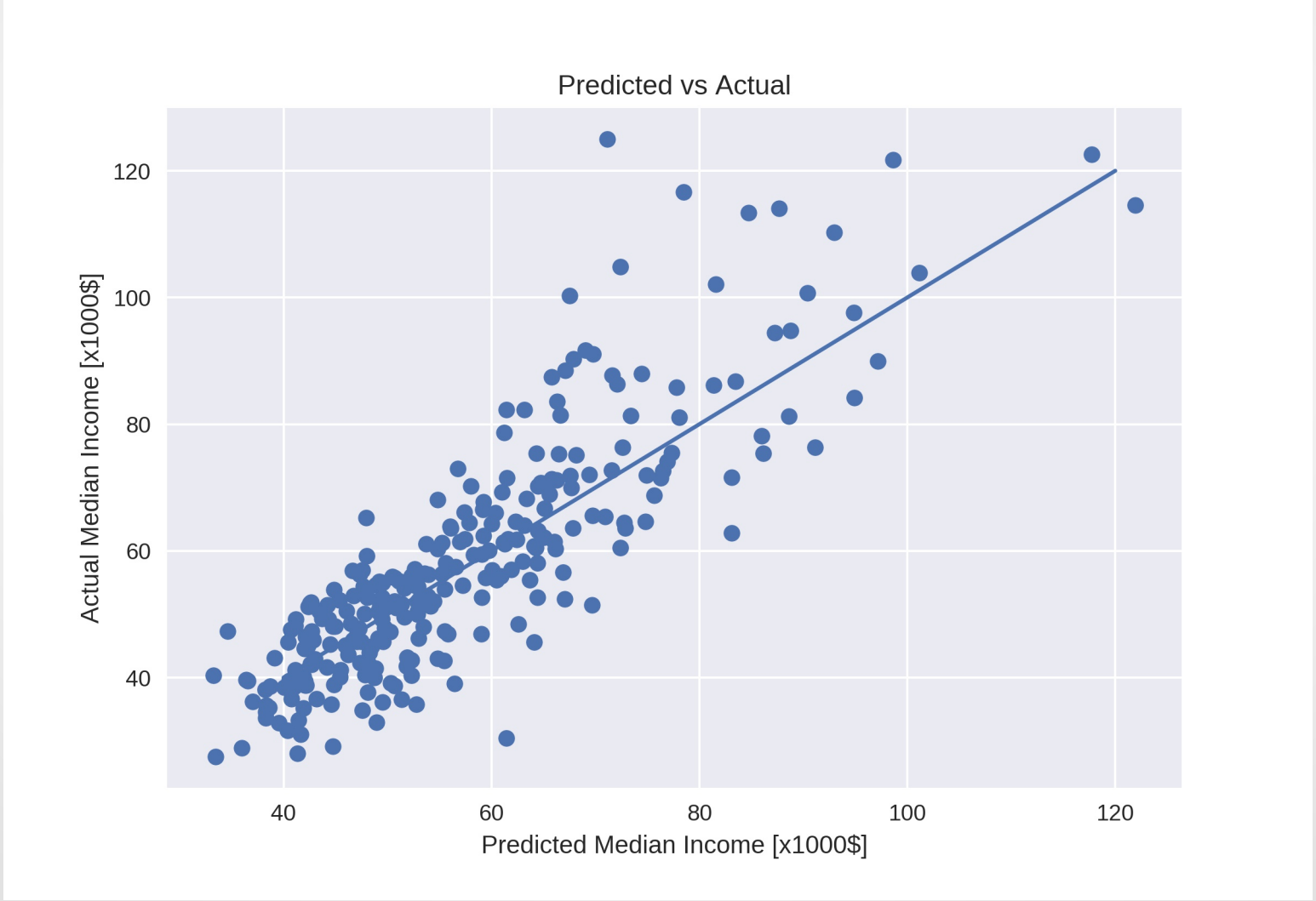
Ridge, Lasso and Elastic Net with CV

Mean square error of test as indicator

Learning curves used to check overfitting

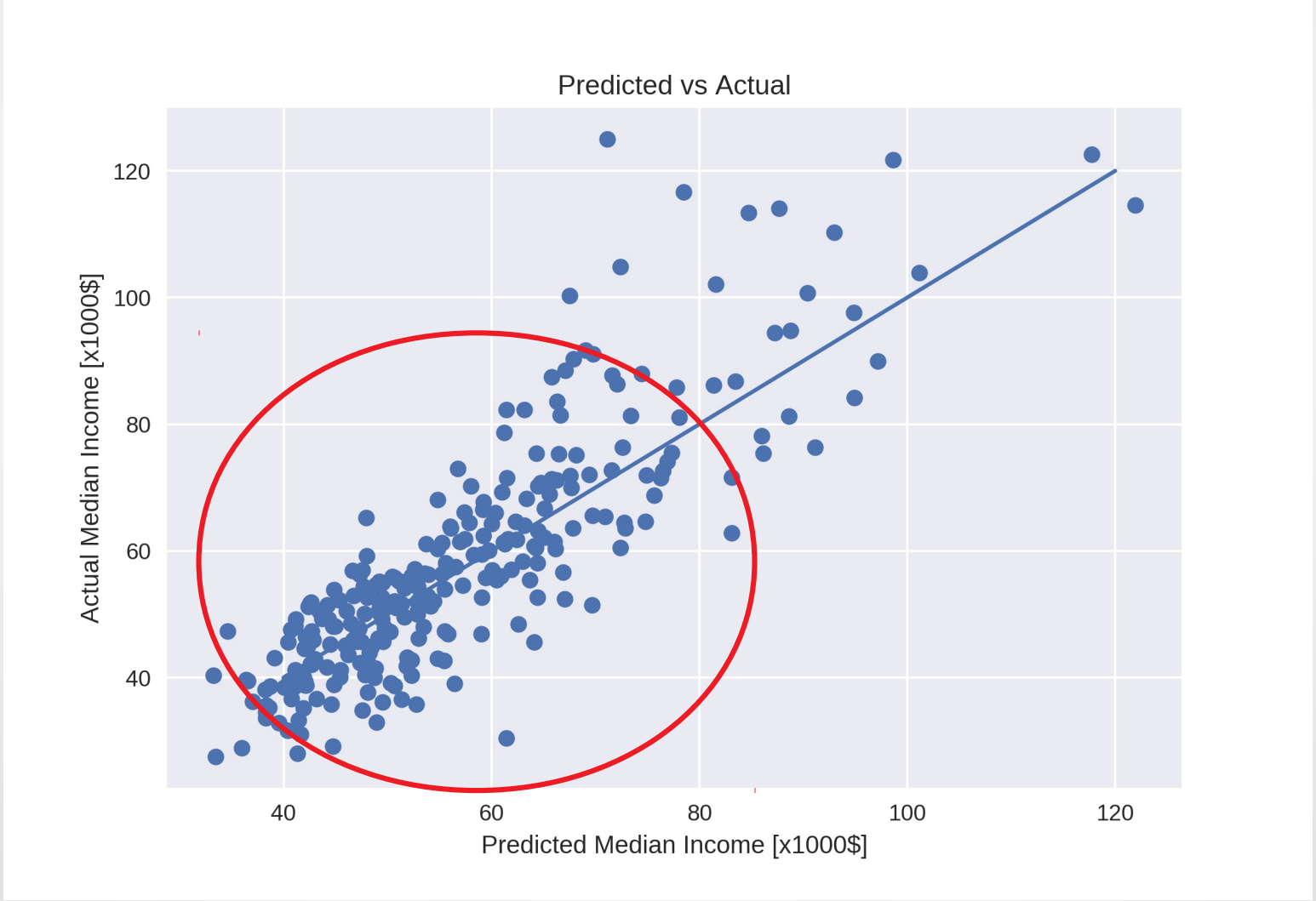
Our chosen model

Lasso regularization polynomial degree 2



Our chosen model

Lasso regularization polynomial degree 2



Influence of features

Population: high population - lower salaries

Influence of features

Population: high population - lower salaries

Area: large cities - higher salaries

Influence of features

Population: high population - lower salaries

Area: large cities - higher salaries

Things to do: interesting cities – higher salaries

Influence of features

Population: high population - lower salaries

Area: large cities - higher salaries

Things to do: interesting cities – higher salaries

Rent price: high rent – high salaries

Influence of features

Population: high population - lower salaries

Area: large cities - higher salaries

Things to do: interesting cities – higher salaries

Rent price: high rent – high salaries

Murders and Assaults: high crime – low salaries

Conclusion and outlook

Predictor works best for medium-low salaries

Conclusion and outlook

Predictor works best for medium-low salaries

Most important: things to do, rent prices and crime

Conclusion and outlook

Predictor works best for medium-low salaries

Most important: things to do, rent prices and crime

More features can be included in future models

Conclusion and outlook

Predictor works best for medium-low salaries

Most important: things to do, rent prices and crime

More features can be included in future models

Extension to cities around the world

Coefficients

```
Index(['Population', 'Area', 'Things_to_do', 'Rent', 'Murders', 'Assaults']
```

```
(array([0, 1, 0, 0, 0, 0]), 0.052979594815016046)
(array([0, 0, 1, 0, 0, 0]), 0.04457759078332485)
(array([0, 0, 0, 1, 0, 0]), 0.5456600637371769)
(array([0, 0, 0, 0, 1, 0]), -0.23546102272115763)
(array([0, 0, 0, 0, 0, 1]), -0.33686571267402315)
(array([2, 0, 0, 0, 0, 0]), -2.9348670163295 99e-05)
(array([1, 0, 1, 0, 0, 0]), -0.018717655497216764)
(array([0, 1, 0, 0, 0, 1]), 0.005620153999713671)
(array([0, 0, 2, 0, 0, 0]), 0.014118787982803794)
(array([0, 0, 1, 0, 1, 0]), 0.026423692519553 64)
(array([0, 0, 0, 2, 0, 0]), -0.025840970521524 5)
(array([0, 0, 0, 0, 0, 2]), 0.09101982517795179)
```