How to Zestimate property price?

Presented by:
Dae Han

Problem

How to pick the best features for Zestimate?

Zillow   ♡ Save   ➦ Share   ∘∘∘ More

$6,499,000   5 bd | 7 ba | 4,988 sqft

311 S Lucerne Blvd, Los Angeles, CA 90020

● For sale | Zestimate®: $6,341,209

Est. payment: $31,060/mo $ Get pre-qualified

Contact Agent

How to pick the best features for Zestimate?

Zestimate®: $6,341,209

Zillow

♡ Save   ↪ Share   ∘∘∘ More

$6,499,000   5 bd | 7 ba | 4,988 sqft

311 S Lucerne Blvd, Los Angeles, CA 90020

● For sale | Zestimate®: $6,341,209

Est. payment: $31,060/mo $ Get pre-qualified

Contact Agent

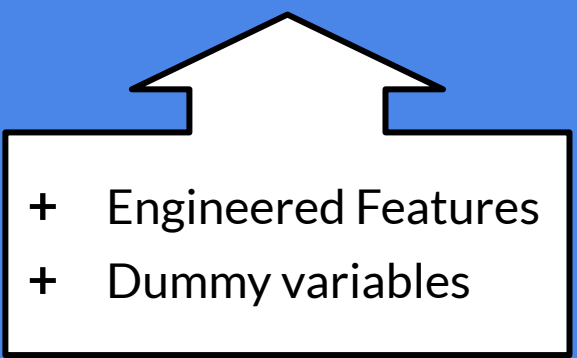[https://www.zillow.com]

The data contains **a lot of** features

From Ames Assessor's Office
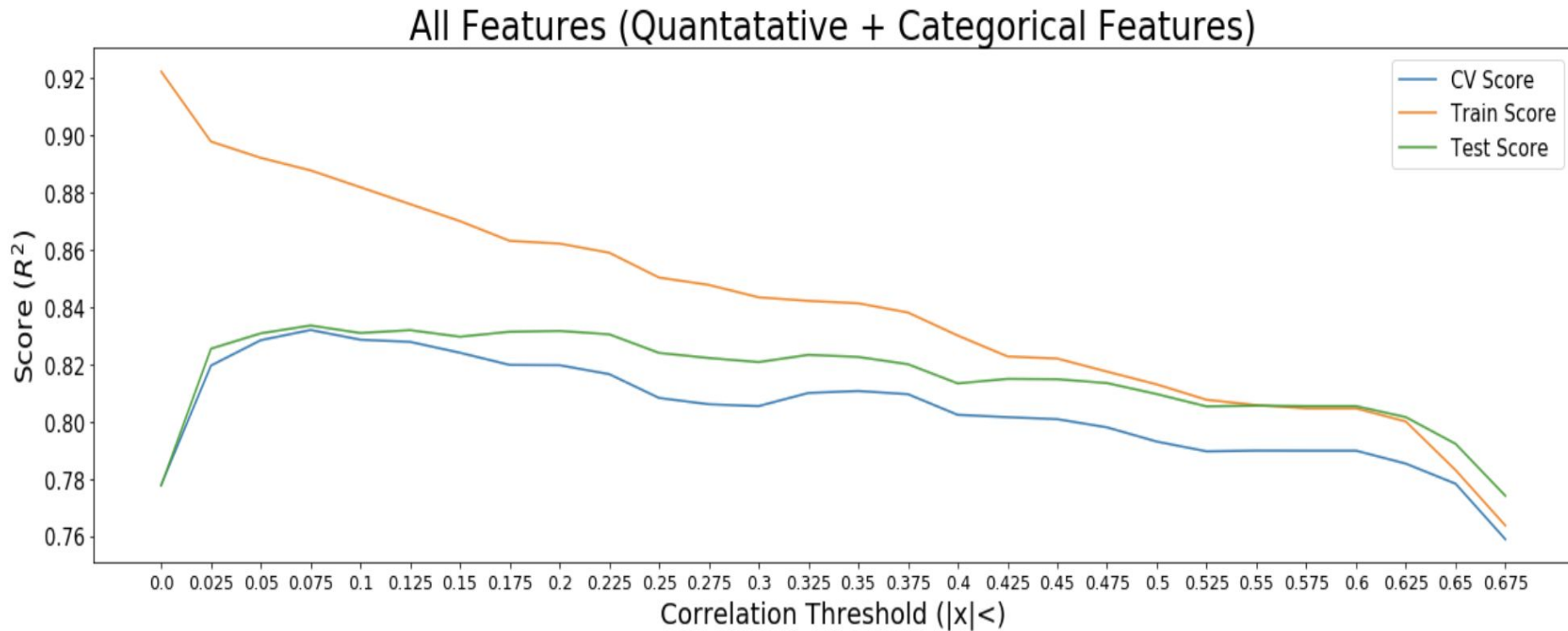
Data about 2051 properties
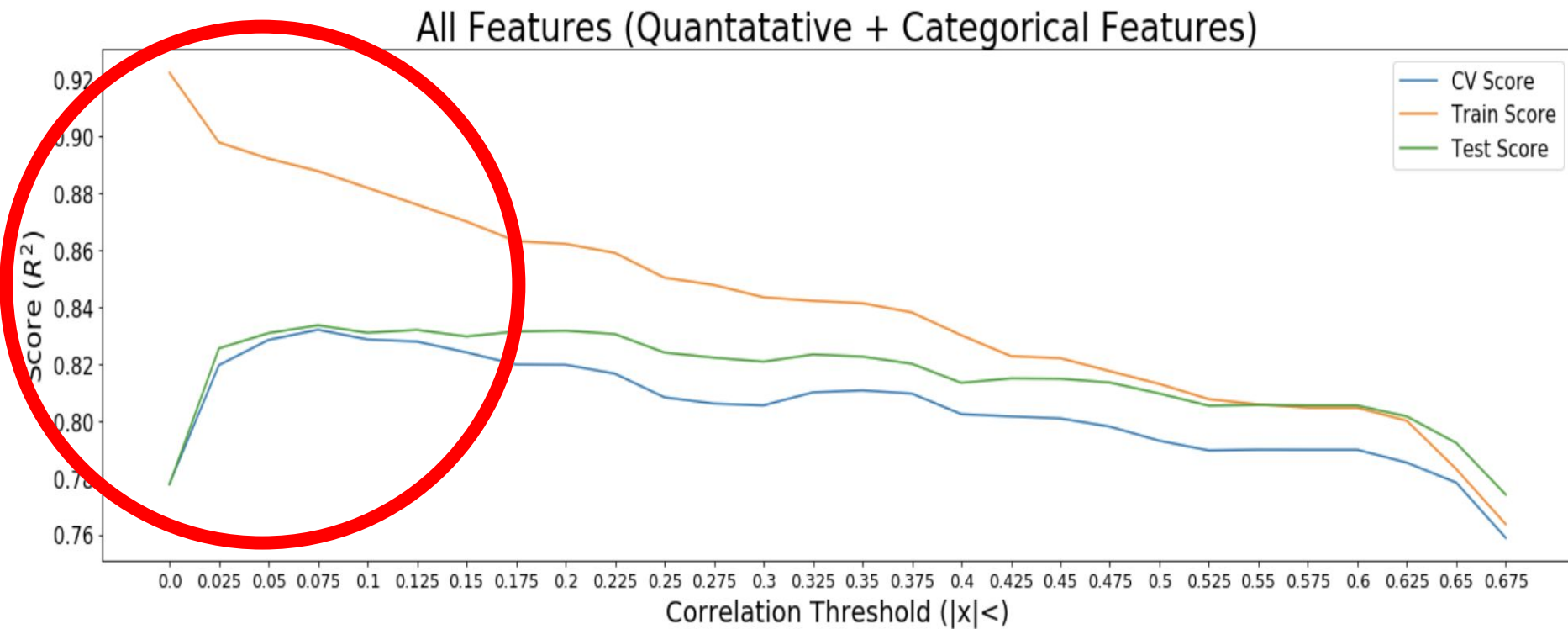
sold in Ames, IA from 2006 to 2010

1003 features
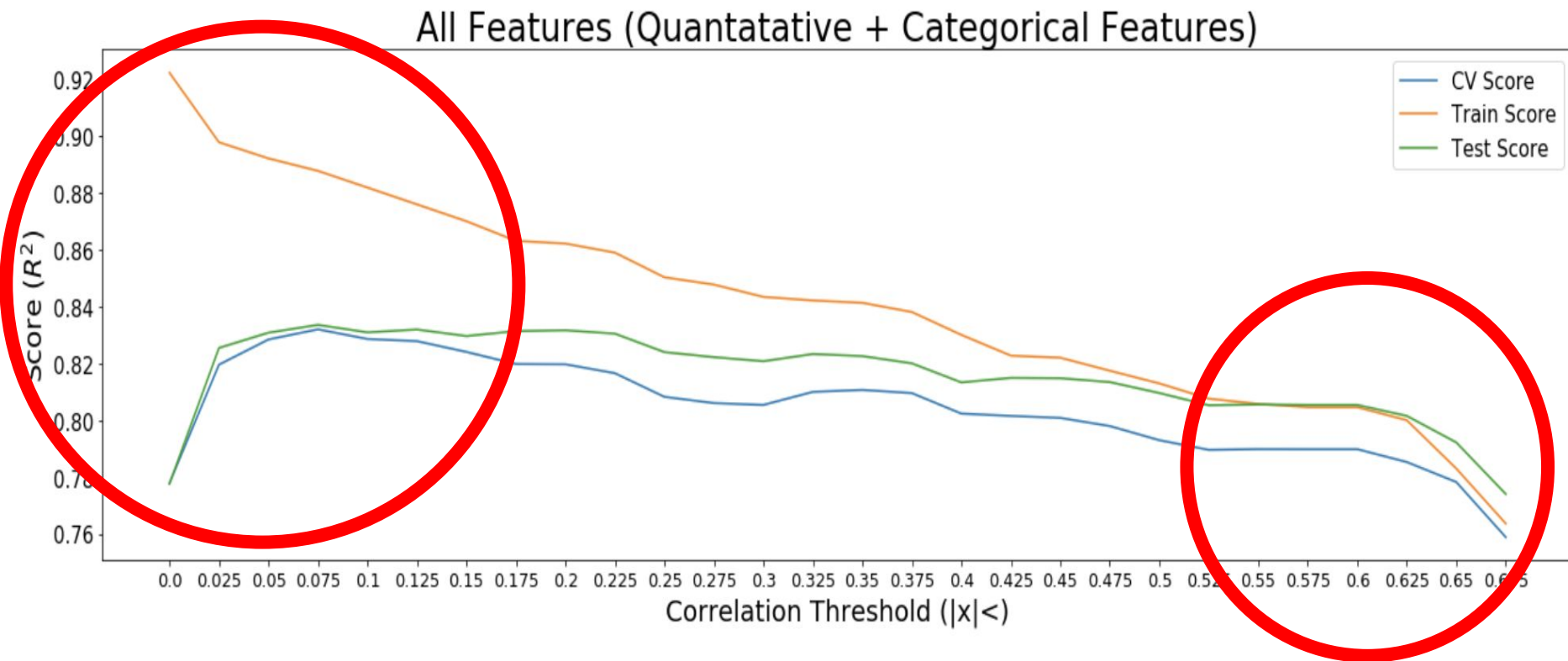

+ Engineered Features
+ Dummy variables
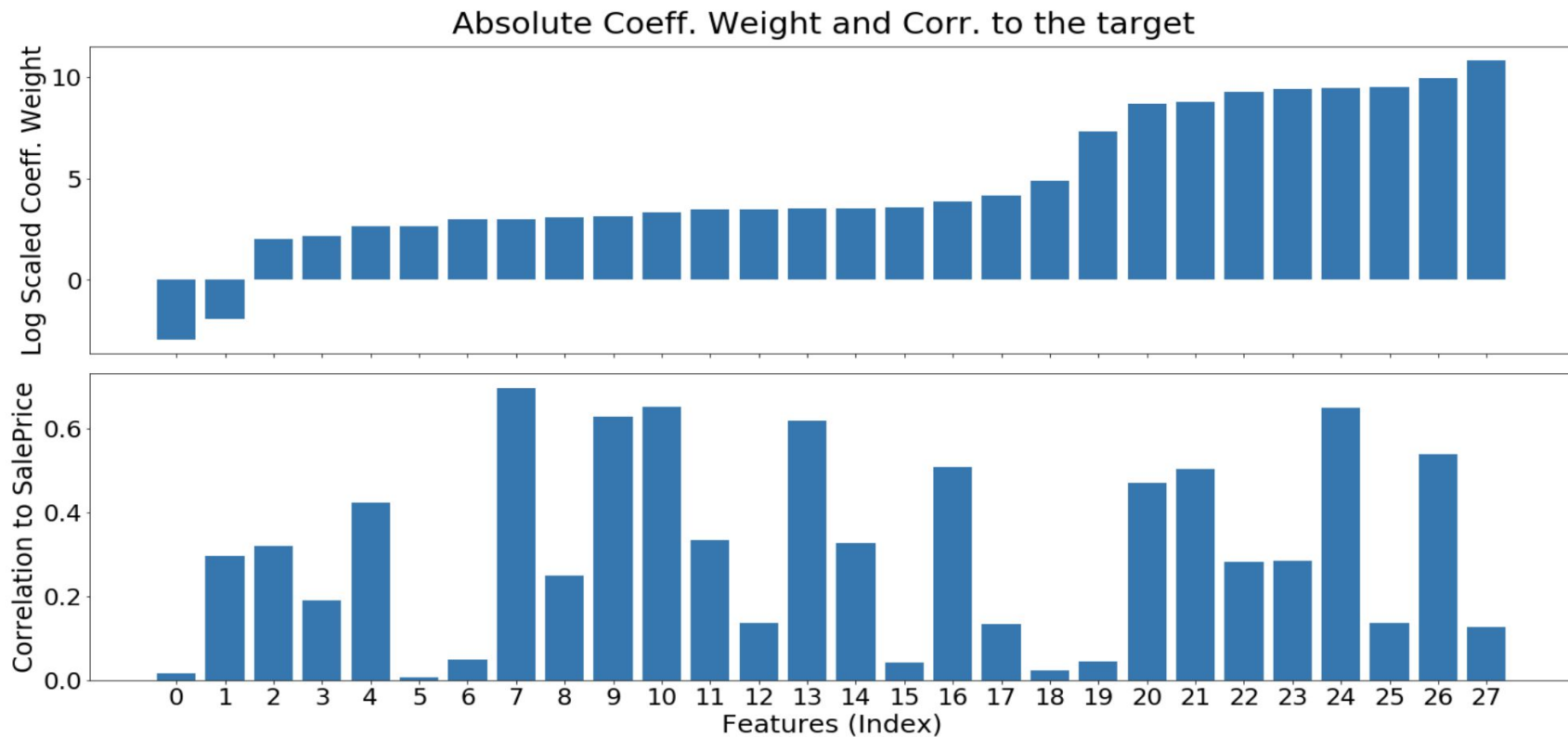
81 features

The more features, the better?   Nop!

All Features (Quantatative + Categorical Features)

The more features, the better?   Nop!

All Features (Quantatative + Categorical Features)

The more features, the better?   Nop!

All Features (Quantatative + Categorical Features)

Coefficient weight and Correlation are not related

Absolute Coeff. Weight and Corr. to the target

Use Lasso Regression to select important features

1003 features

Lasso Regression

49 features

Use Lasso Regression to select important features
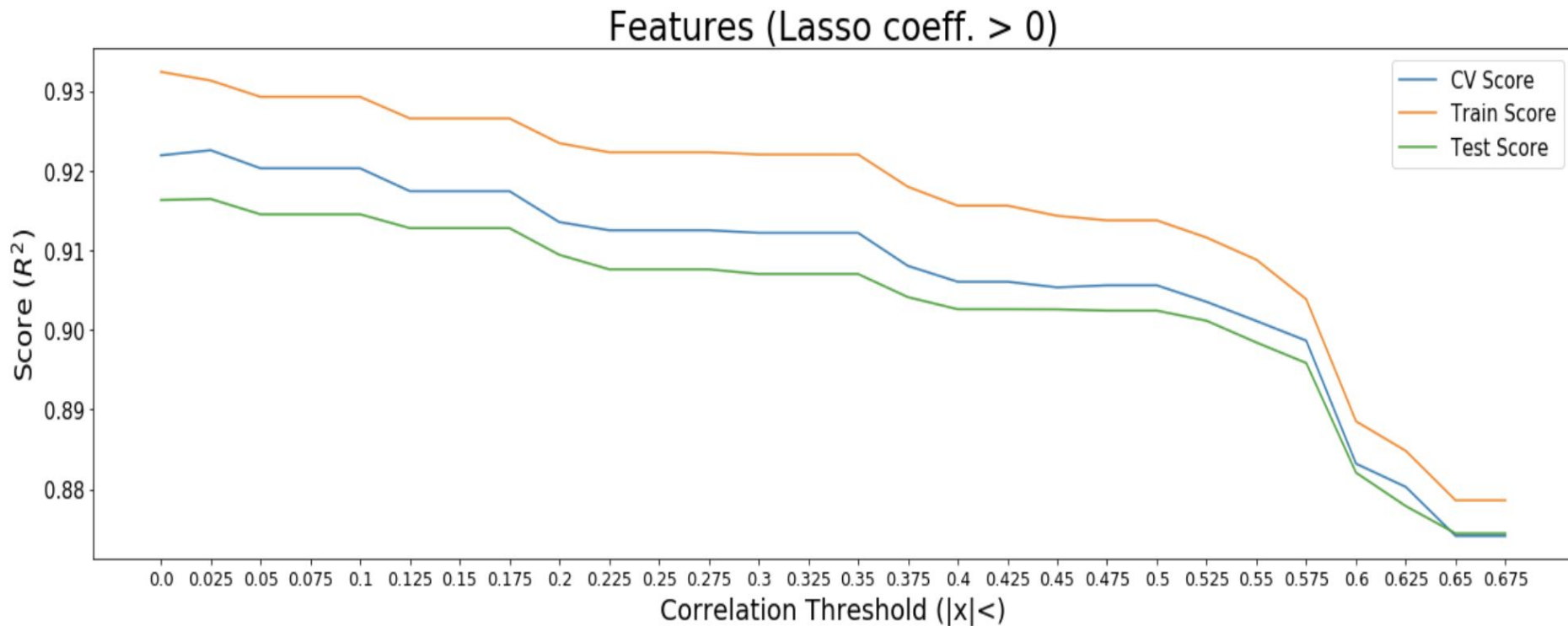
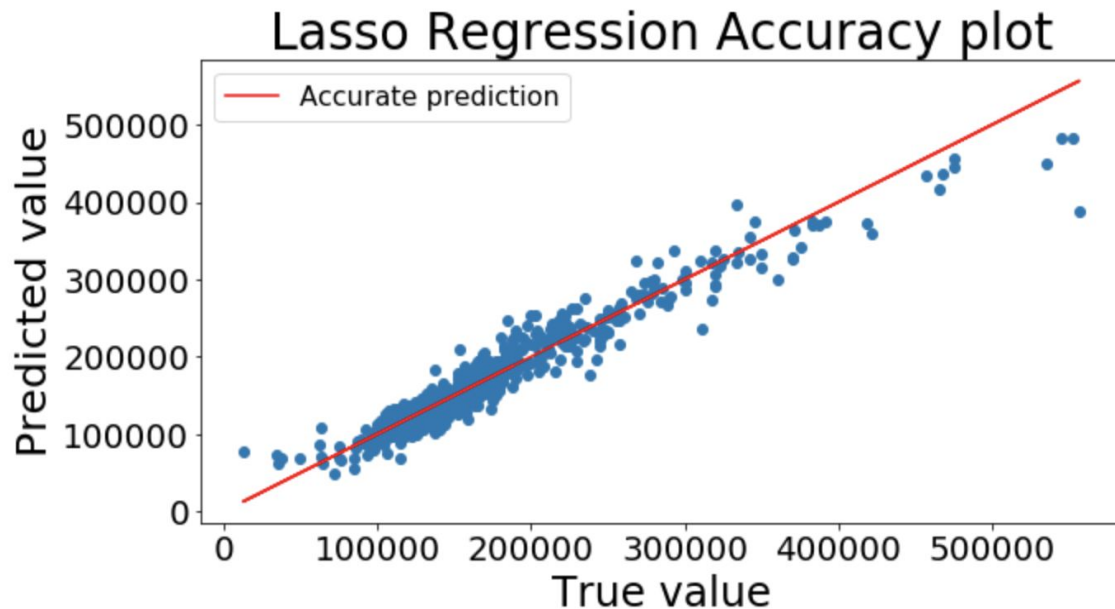# Use Lasso Regression to select important features
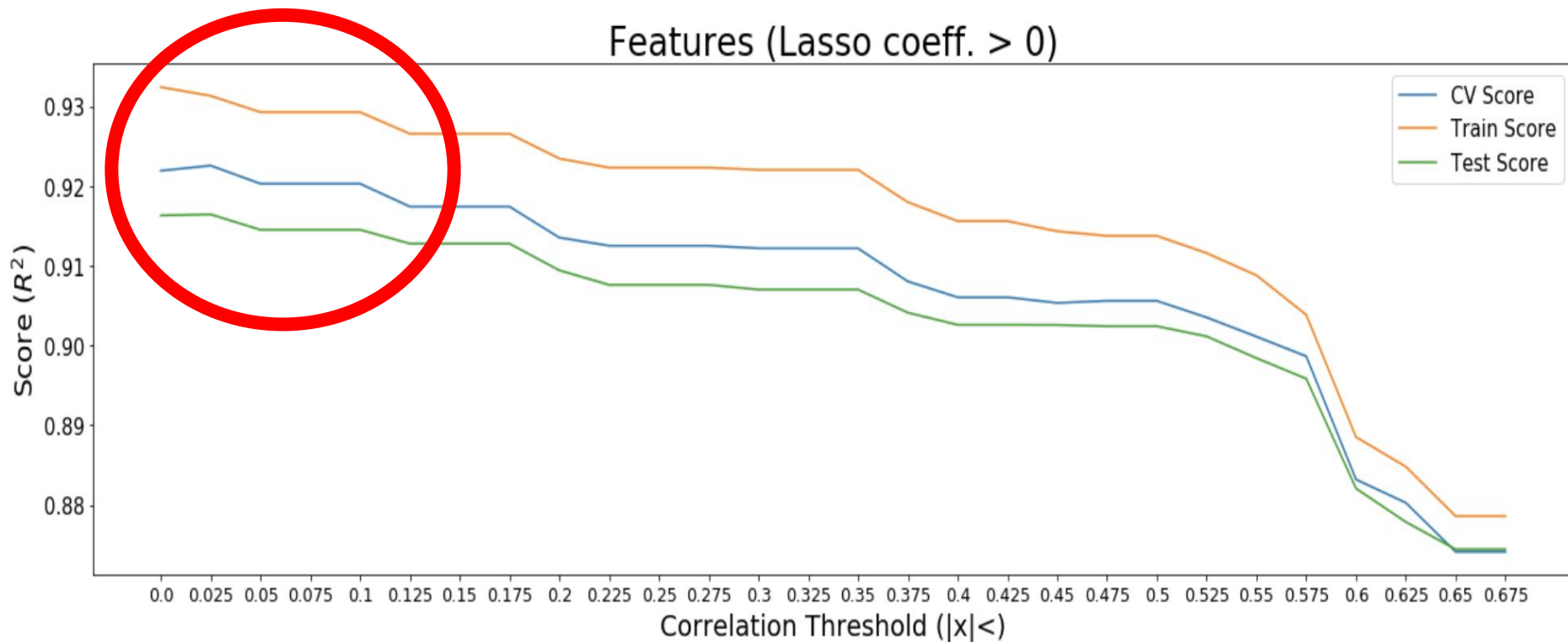
Use Lasso Regression to select important features
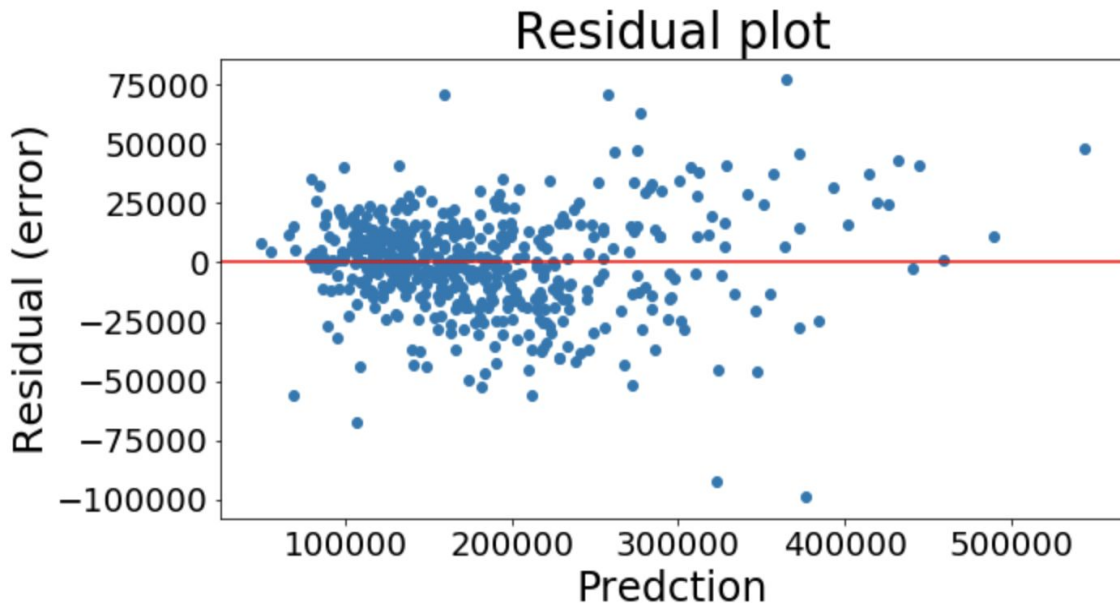
Prediction Score:

R-square: **0.93**

RMSE: **20267 $**

Next Step

Tuning the model to improve overfitting and metrics

- Try Ridge Regression
- Try Log Scaling target and or features

- Aim for
- Normal distribution
- Random scatter of residuals (error)

Residual plot

How to pick the best features for Zestimate?

- **Not Important factors**

  1. Coefficient Weight from Linear Regression

  2. Correlation to the target

- **Use Lasso regression to select important features**

- **Tune your model using Ridge regression, log scaling**

Source:

1. https://www.zillow.com

2. http://jse.amstat.org/v19n3/decock/DataDocumentation.txt

3. https://pixabay.com/