

ECE ING4 MACHINE LEARNING

Jeremy Cohen



Final Week

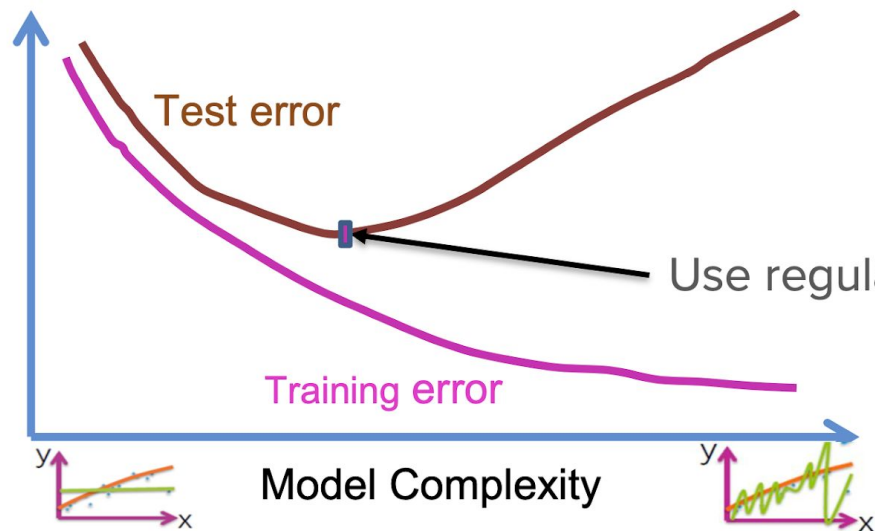
- Full Review
- Regularization
- Mini Project

Full Review

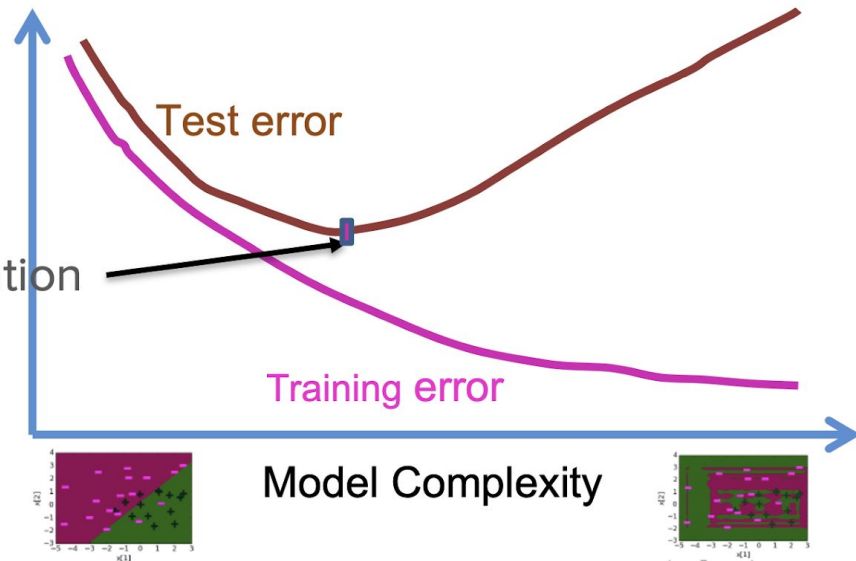
Our Final Course: Regularization

The Problem

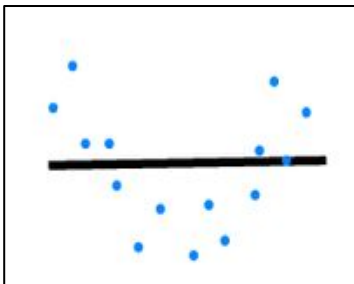
Polynomial Regression



Logistic Regression

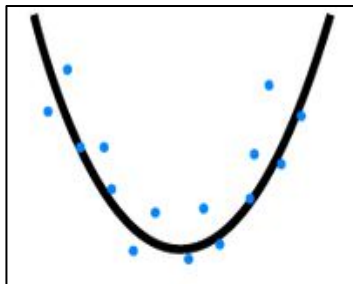


The Problem



UNDERFITTING

$$\theta_0 + \theta_1 x$$



JUST RIGHT

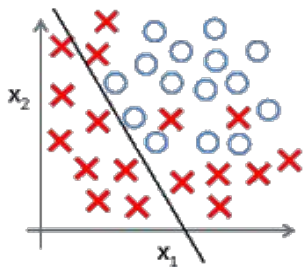
$$\theta_0 + \theta_1 x + \theta_2 x^2$$



OVERFITTING

$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

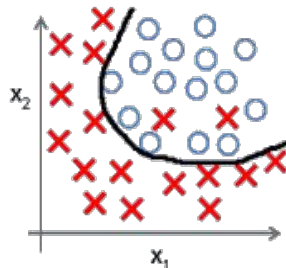
For Logistic Regression



UNDERFITTING

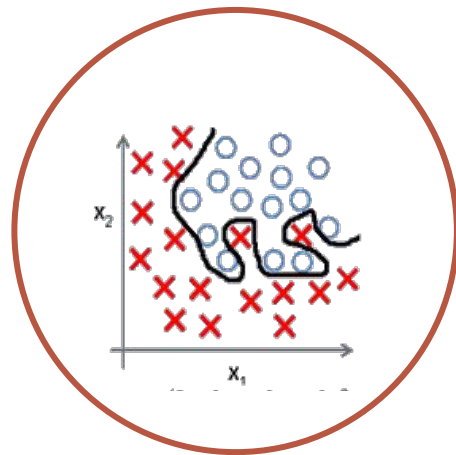
$$h_{\theta}(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

(g = sigmoid function)



JUST RIGHT

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2)$$



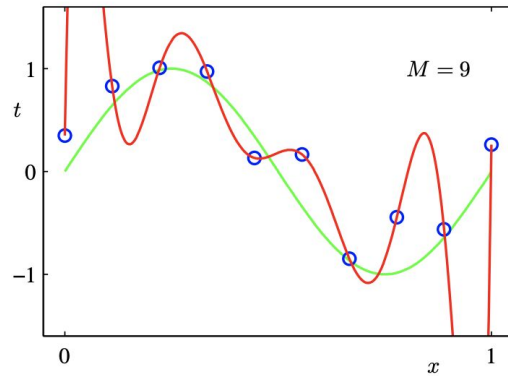
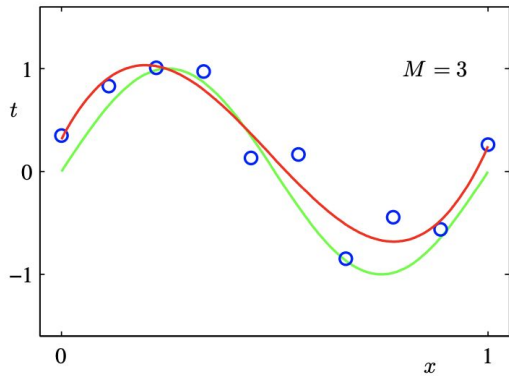
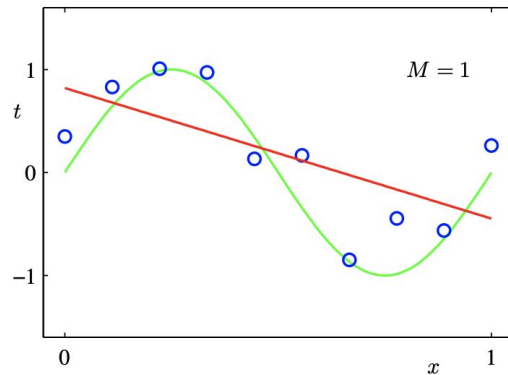
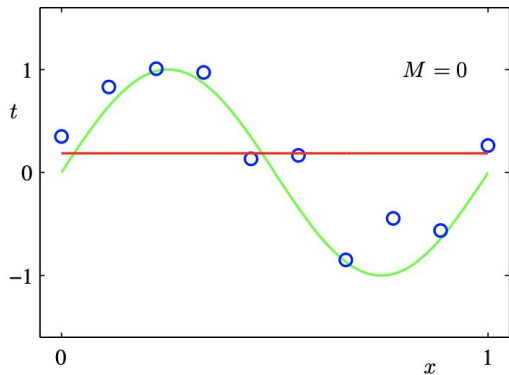
OVERFITTING

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \dots)$$

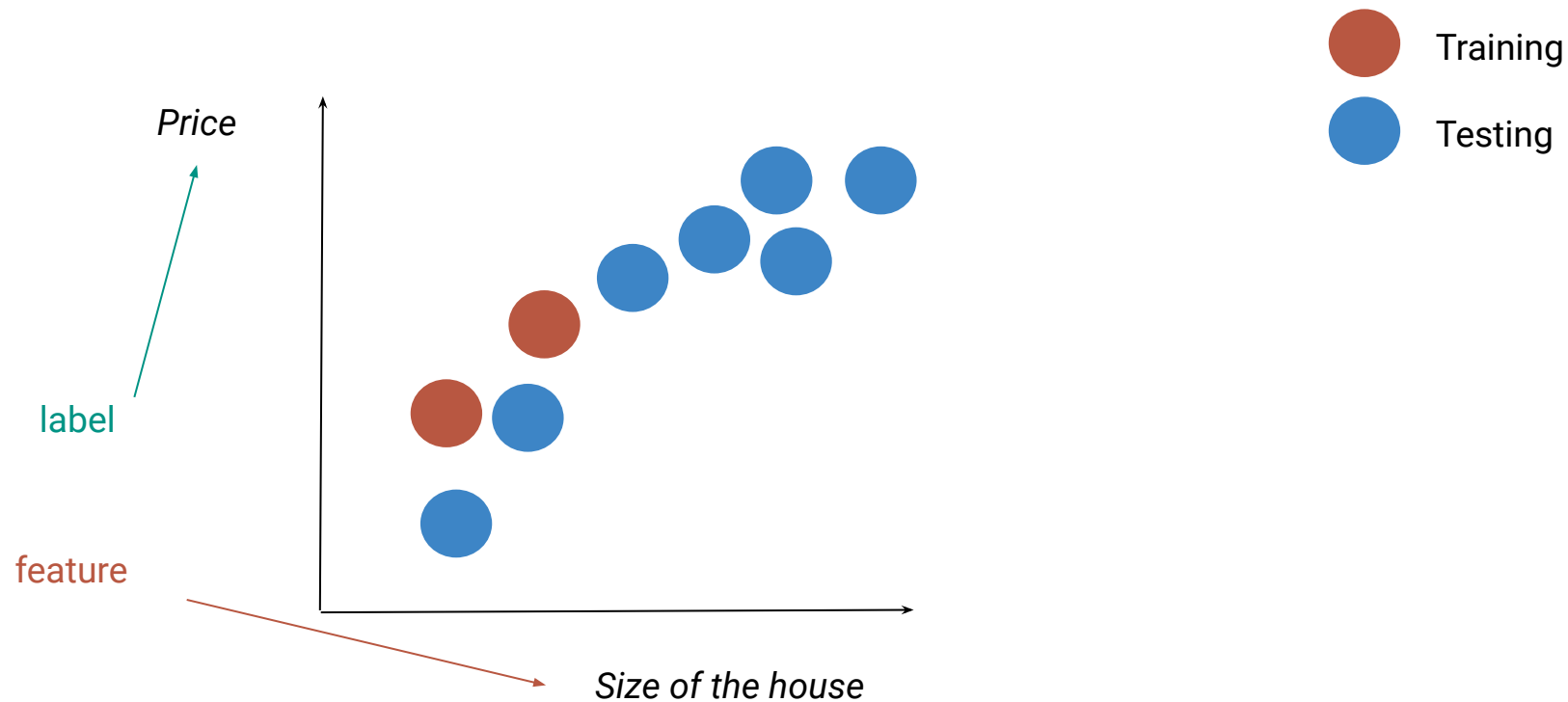
Overfitting

We fit the training data correctly but fail to generalize

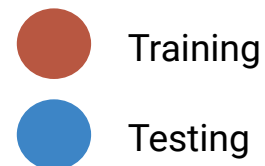
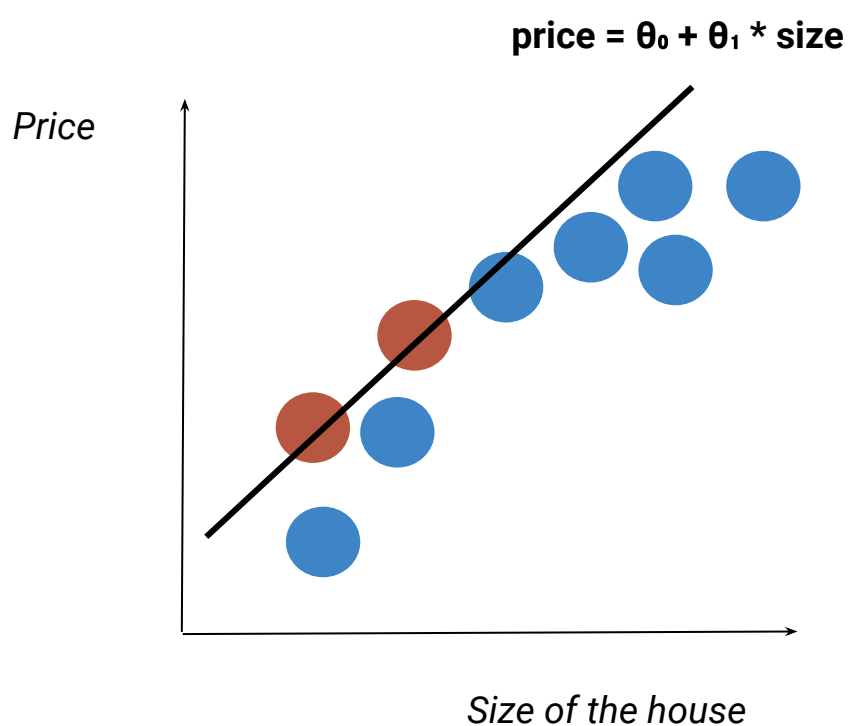
For Logistic Regression



Regression



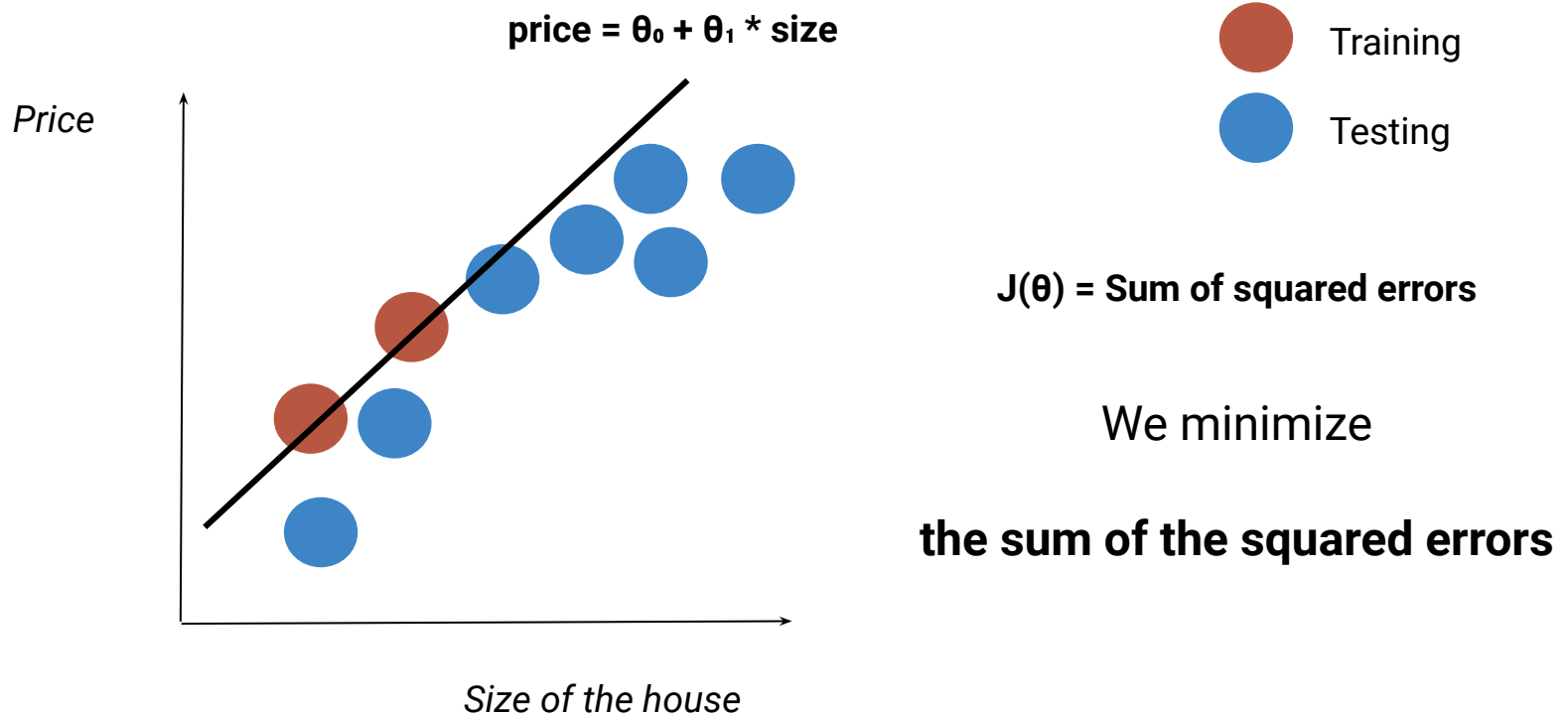
Regression



0 bias

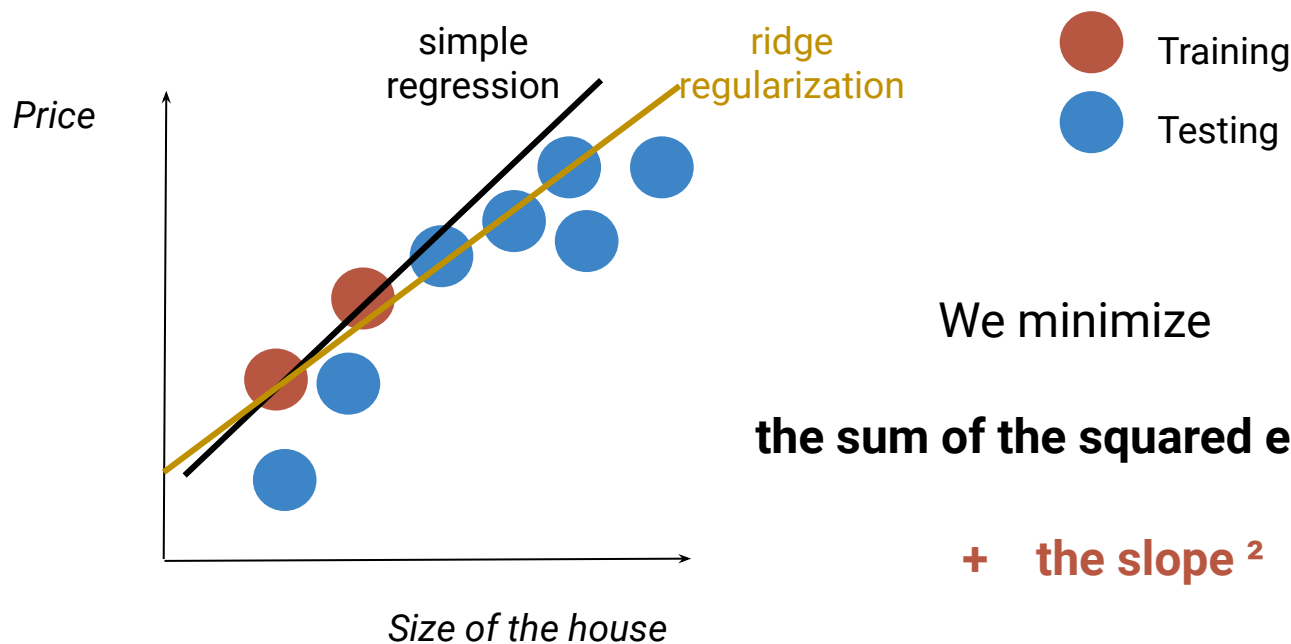
some variance

Regression

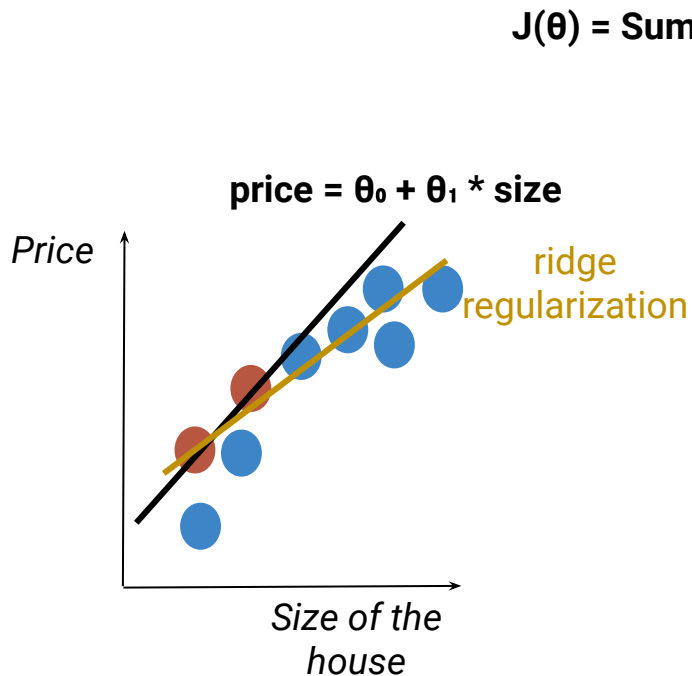


Regularization

$$J(\theta) = \text{Sum of squared errors} + \lambda * \text{slope}^2$$



Regularization



$$J(\theta) = \text{Sum of squared errors} + \lambda * \text{slope}^2$$

determine how strong the penalty is

adds a penalty

We minimize

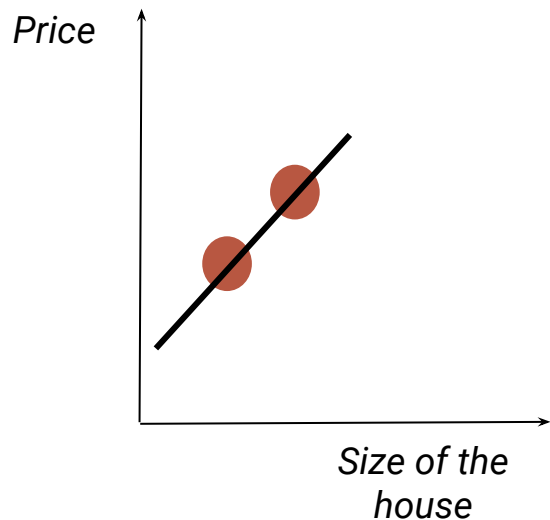
the sum of the squared errors

+ the slope²

Ridge Regularization

$$J(\theta) = \text{Sum of squared errors} + \lambda * \text{slope}^2$$

$$\text{price} = 0.4 + 1.3 * \text{size}$$



for now: $\lambda = 1$

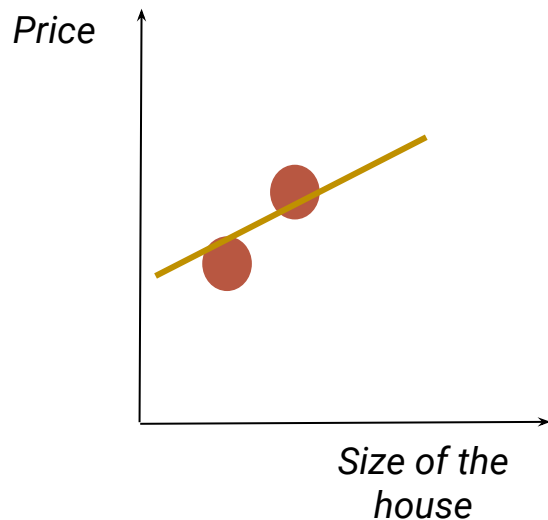
$$J(\theta) = 0 + 1 * 1.3^2$$

$$J(\theta) = 1.69$$

Ridge Regularization

$$J(\theta) = \text{Sum of squared errors} + \lambda * \text{slope}^2$$

$$\text{price} = 0.9 + 0.8 * \text{size}$$



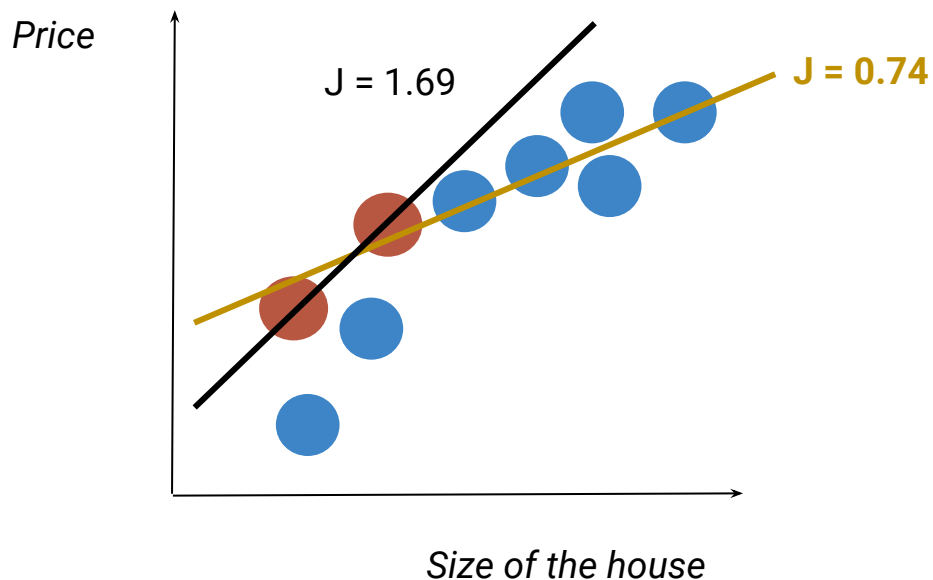
for now: $\lambda = 1$

$$J(\theta) = 0.3^2 + 0.1^2 + 1 * 0.8^2$$

$$J(\theta) = 0.74$$

Ridge Regularization

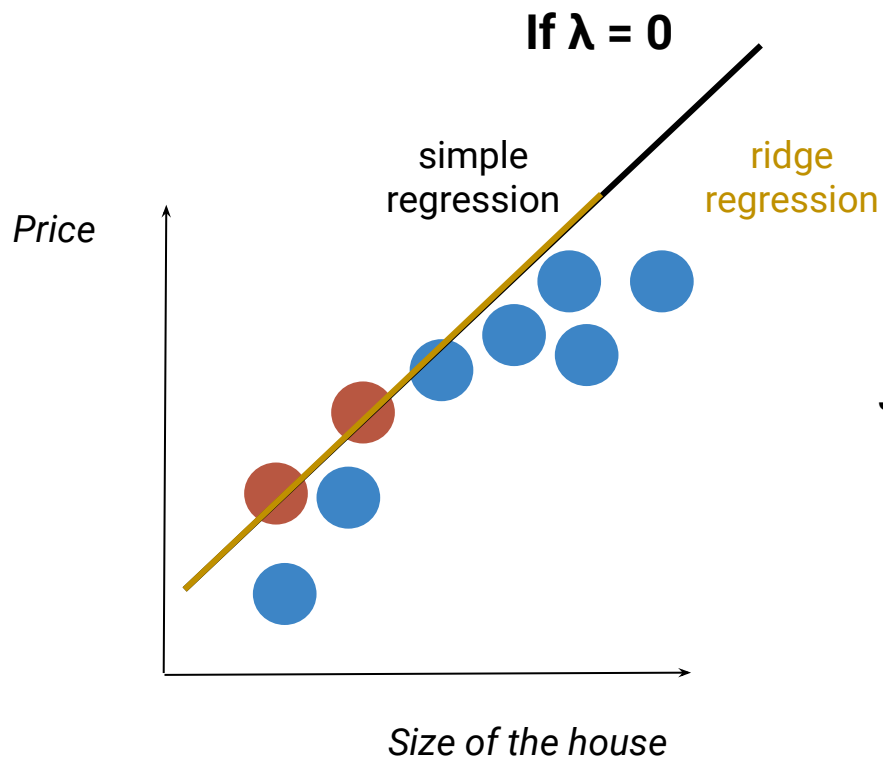
The price is less sensitive to the size of the house



some bias

low variance

What about λ

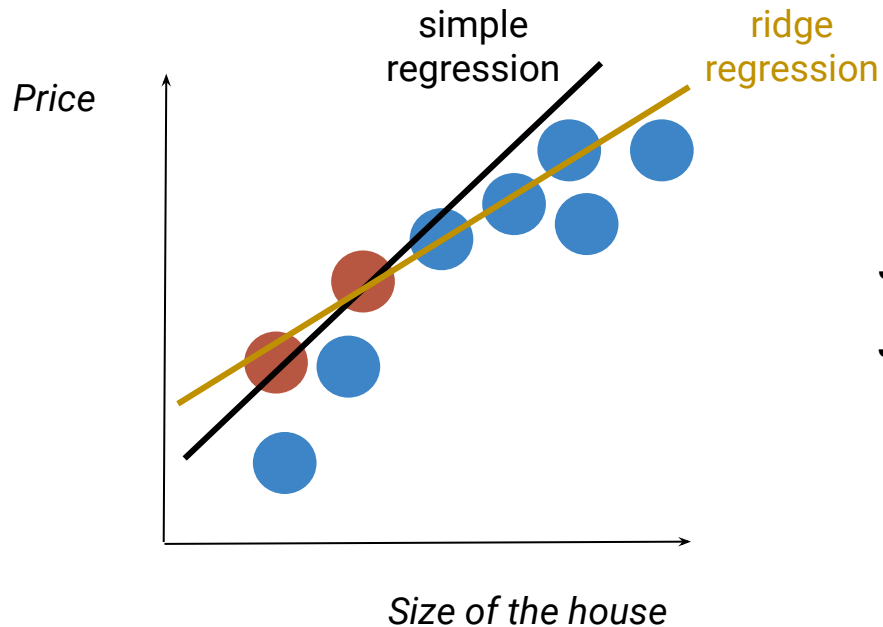


$$J(\theta) = \text{Sum of squared errors} + \lambda * \text{slope}^2$$

$$J(\theta) = \text{Sum of squared errors} + 0$$

What about λ

If $\lambda = 1$



$$J(\theta) = \text{Sum of squared errors} + \lambda * \text{slope}^2$$

$$J(\theta) = \text{Sum of squared errors} + 1 * \text{slope}^2$$

What about λ

If $\lambda = 2$

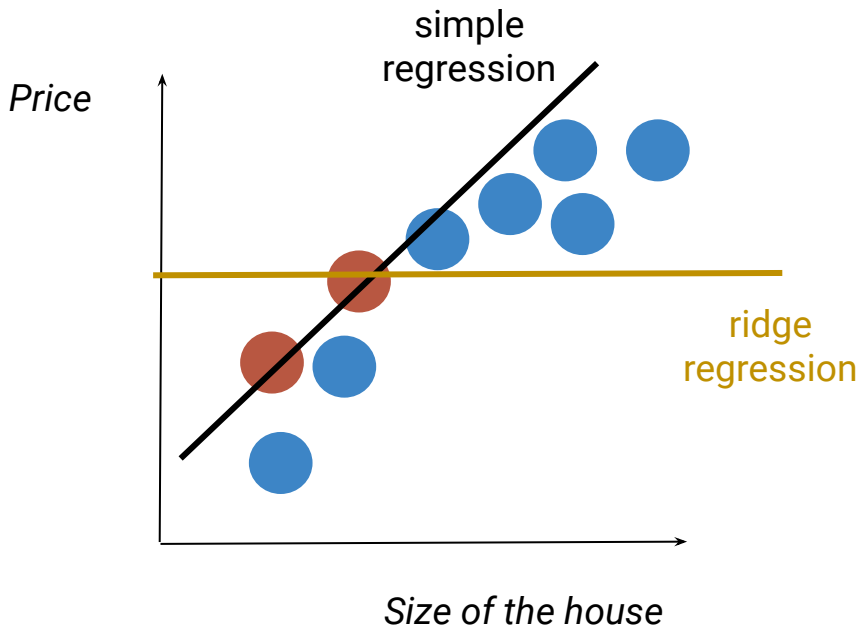


$$J(\theta) = \text{Sum of squared errors} + \lambda * \text{slope}^2$$

$$J(\theta) = \text{Sum of squared errors} + 2 * \text{slope}^2$$

What about λ

If $\lambda = 10000$

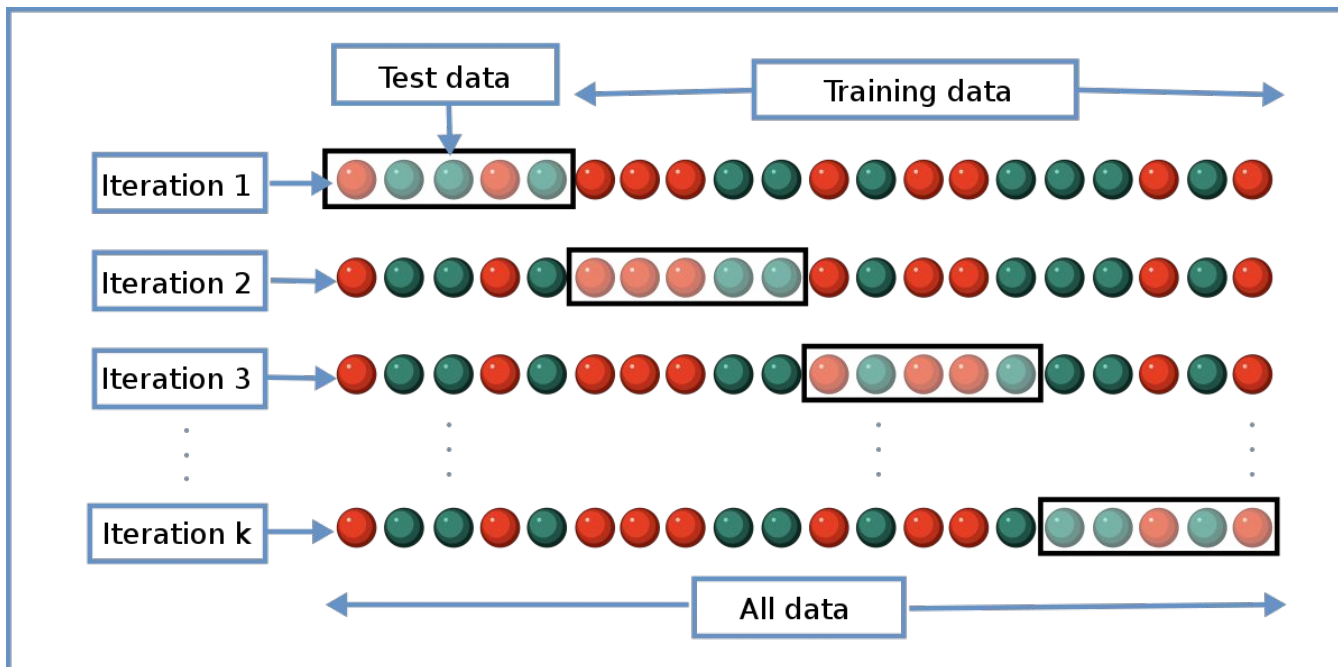


$$J(\theta) = \text{Sum of squared errors} + \lambda * \text{slope}^2$$

$$J(\theta) = \text{Sum of squared errors} + 10000 * \text{slope}^2$$

What about λ

To estimate λ , we use cross-validation



Multiple Features

$$\text{price} = \theta_0 + \text{slope1} * \text{size} + \text{slope2} * \text{rating} + \text{slope3} * \text{number of bedrooms} + \dots$$

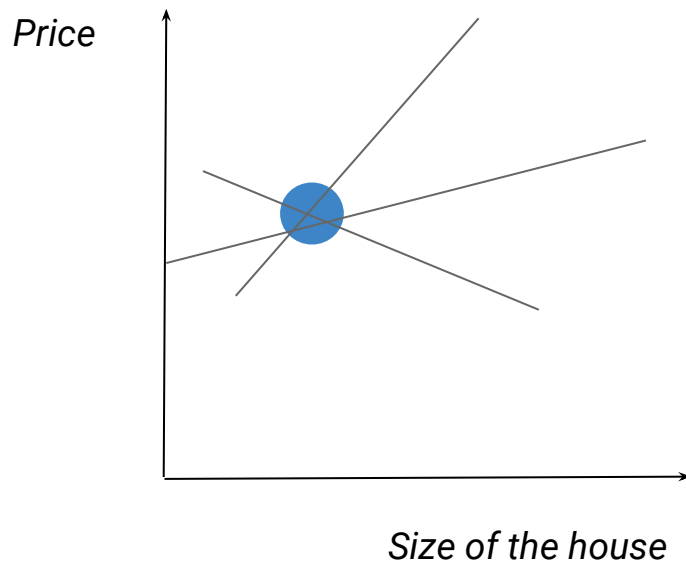
We minimize

the sum of the squared errors

$$+ \lambda (\text{slope1}^2 + \text{slope2}^2 + \text{slope3}^2)$$

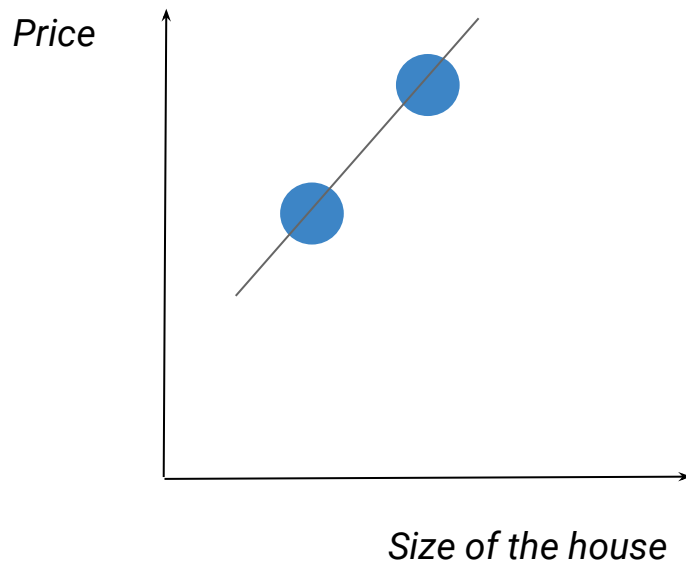
Multiple features

2 features - 1 data point



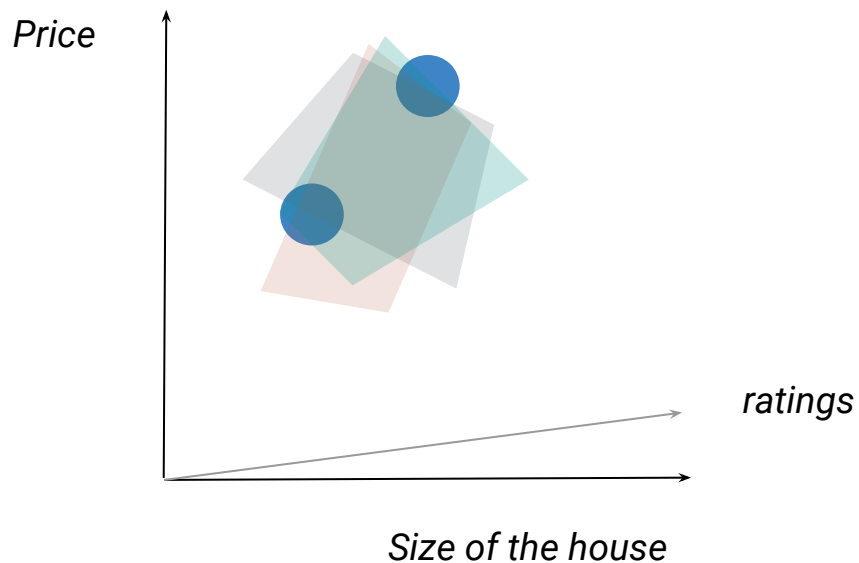
Multiple features

2 features - 2 data points



Thousands of features

3 features - 2 data points



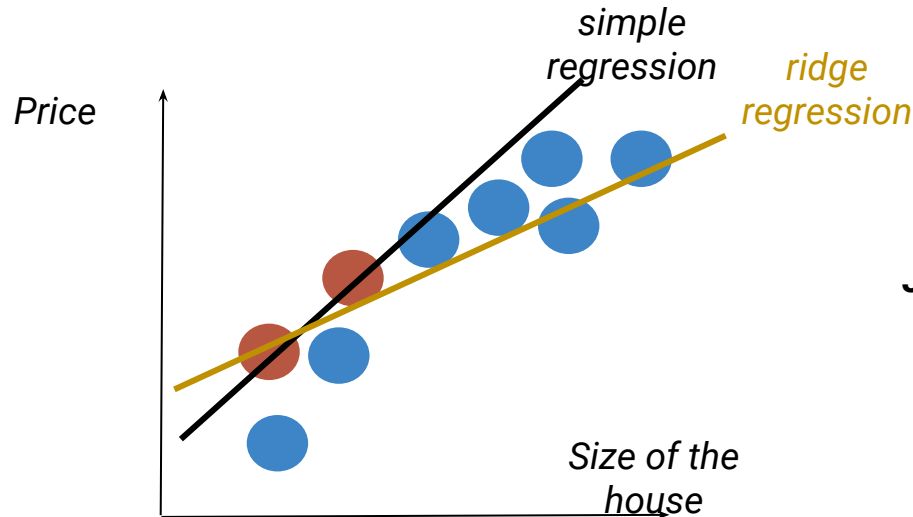
Thousands of features

If we have 4 parameters, we need 4 data points

In case of a dataset with not enough data points compared to the number of features, regularization can help setting some parameters to 0

Summary

Ridge Regularization makes the regression **less sensitive to the training data** (especially when in low number) and helps **reduce overfitting** by **adding a penalty to the cost function**.



$$J(\theta) = \text{Sum of squared errors} + \lambda * \text{slope}^2$$

Ridge Regularization

Ridge Regression (L_2 regularization)

$$J(\theta) = \frac{1}{2m} \left[\sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n (\theta_j)^2 \right]$$

Demo

<http://madrury.github.io/smoothers/>

Lasso Regularization

Ridge Regression (L_2 regularization)

$$J(\theta) = \frac{1}{2m} \left[\sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n (\theta_j)^2 \right]$$

Lasso Regression (L_1 regularization)

$$J(\theta) = \frac{1}{2m} \left[\sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n |\theta_j| \right]$$

λ is the regularization parameter:

- Ridge: Encourages small weights θ but not exactly 0
- Lasso: "Shrink" some weights θ exactly to 0

Gradient Descent

$$\begin{aligned}\theta_0 &:= \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_0^{(i)} \\ \theta_j &:= \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}\end{aligned}$$

With regularization: $\theta_j := \theta_j \left(1 - \alpha \frac{\lambda}{m}\right) - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$

α, λ are learning parameters to choose manually

In practice: $(1 - \alpha\lambda/m)$ is between 0.99 and 0.95

Logistic Regression

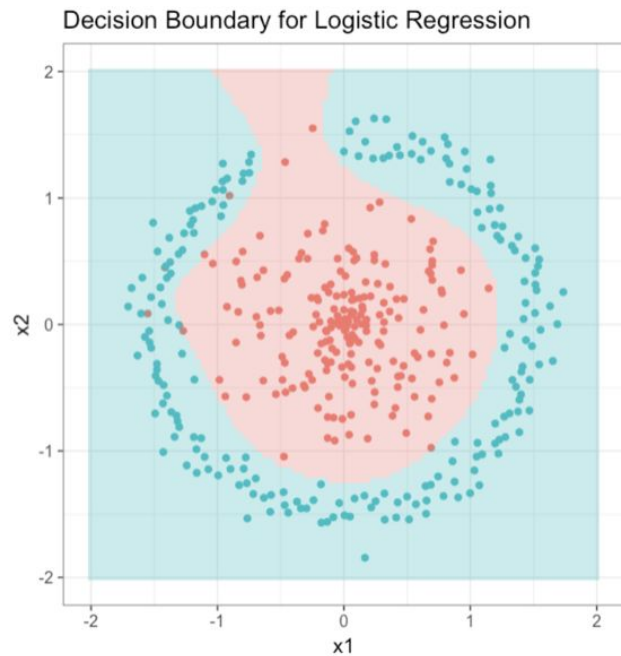
Original Formula

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))]$$

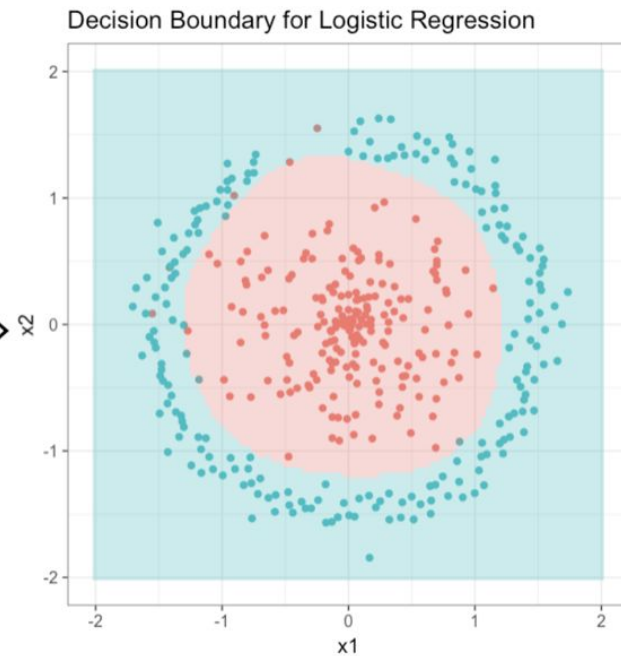
Updated Formula

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))] + \underbrace{\frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2}_{\text{regularization term}}$$

Logistic Regression



Regularization

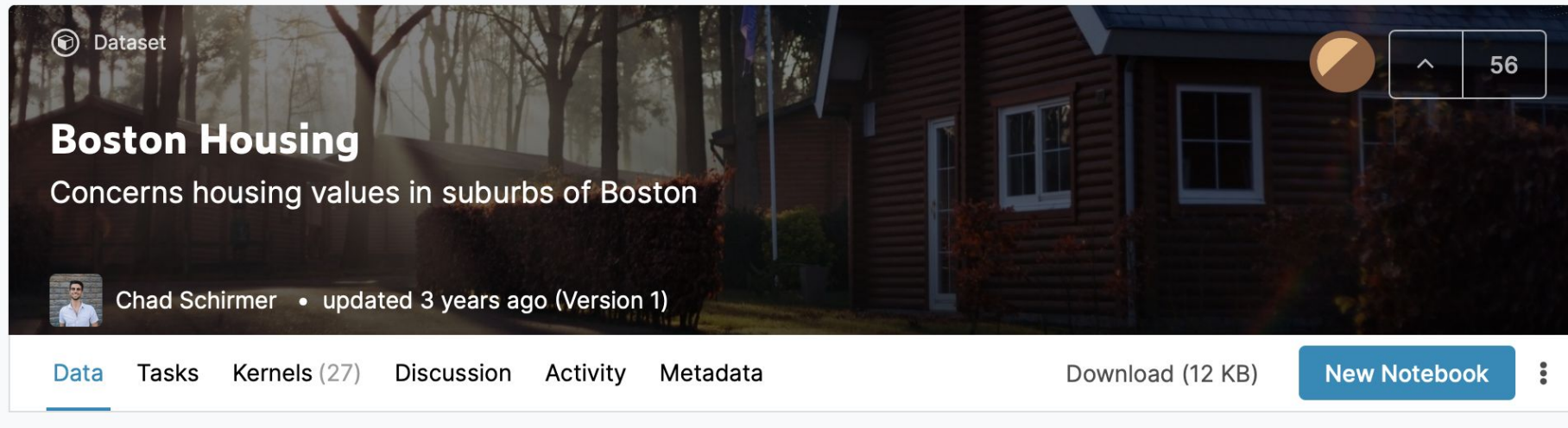


MINI PROJECT

What now?


kaggleTM


BOSTON HOUSING PRICE


The image shows the Kaggle dataset page for 'Boston Housing'. The background is a photograph of a brown wooden house with white window frames and a chimney, surrounded by trees and a lawn. In the top left corner, there is a 'Dataset' icon and the word 'Dataset'. The title 'Boston Housing' is prominently displayed in white, followed by the subtitle 'Concerns housing values in suburbs of Boston'. Below this, a user profile picture of Chad Schirmer is shown next to his name and the text 'updated 3 years ago (Version 1)'. At the bottom, there is a navigation bar with links for 'Data', 'Tasks', 'Kernels (27)', 'Discussion', 'Activity', and 'Metadata'. On the right side of the navigation bar, there is a 'Download (12 KB)' link and a blue 'New Notebook' button. A small box in the top right corner shows an upward arrow and the number '56'.

<https://www.kaggle.com/schirmerchad/bostonhousingmlnd>

US ACCIDENTS

 Dataset




 Sobhan Moosavi • updated 13 days ago (Version 3)

US Accidents (3.0 million records)
A Countrywide Traffic Accident Dataset (2016 - 2019)

410

^

[Data](#) [Tasks \(6\)](#) [Kernels \(43\)](#) [Discussion \(6\)](#) [Activity](#) [Metadata](#) [Download \(1 GB\)](#) [New Notebook](#) 

<https://www.kaggle.com/sobhanmoosavi/us-accidents/kernels>

ENRON

 Dataset



^

396

The Enron Email Dataset

500,000+ emails from 150 employees of the Enron Corporation

 William Cukierski • updated 4 years ago (Version 2)

Data

Tasks

Kernels (161)

Discussion (4)

Activity

Metadata

Download (1 GB)

New Notebook

⋮

<https://www.kaggle.com/wcukierski/enron-email-dataset>

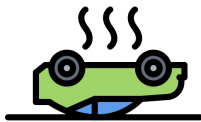
FREE CHOICE

Solve a real-world problem of your choice

Solve a problem using Machine Learning



**BOSTON HOUSE
PRICES**



US ACCIDENTS



ENRON SCANDAL



FREE CHOICE



Teaming Up is allowed (up to 2)



FINAL THANK YOU

ECE STUDENTS GET 50% OFF ALL COURSES!

<https://jeremycohen.podia.com>

COUPON
ECE20



See you soon!