

Introduction to Machine Learning

Day 3: Bias, Variance, and Regularization

Joanna Bieri DATA301

Day 3

Important Information

- Email: joanna_bieri@redlands.edu
- Office Hours: Duke 209, [Click Here for Joanna's Schedule](#)
- Class Website

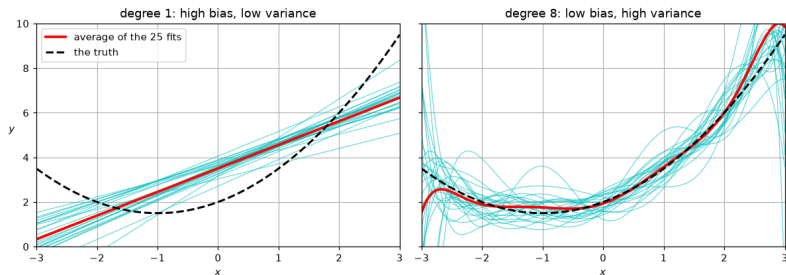
The Four Things From Today

- 1 **Bias** is being wrong on average. **Variance** is moving around when the data changes.
- 2 With a fixed amount of data you trade one for the other.
- 3 **Ridge** charges for big weights. **Lasso** deletes features. **Elastic Net** is both.
- 4 **Scale your features first**, always.

Get out your handwritten notes!

Bias and Variance

25 Datasets, 25 Fits, Two Model Types



The Numbers

	bias squared	variance
degree 1	1.884	0.220
degree 8	0.189	3.261

Degree 1 to degree 8: bias down by about 10x, variance up by about 15x.

Neither model is good. One is wrong, the other is unstable.

Discuss

Left panel: all 25 lines sit on top of each other, but their average is a **line** and the truth is a **parabola**. Right panel: the average sits almost on the truth, but the individual curves wander, worst at the edges.

- **Q2a.** What is the difference between high bias and high variance? Do not use the words underfitting or overfitting.
- **Q2b.** Why are the degree 8 curves worst near $x = -3$ and $x = 3$?

Ridge

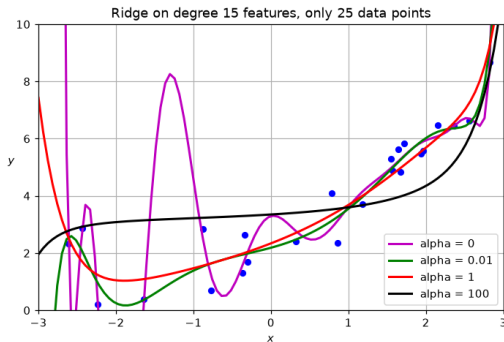
Charge The Model For Big Weights

$$\text{minimize} \quad MSE + \alpha \sum_{i=1}^n w_i^2$$

Nothing in plain MSE says the weights have to be reasonable. If a weight of 2000 lowers the error a little, the model will use 2000. A curve that swings violently between points is a curve with enormous weights.

α is a **hyperparameter**: a number you choose, not one the model learns.

Four Values Of Alpha, Only 25 Data Points



alpha 0 runs off the top of the plot twice. alpha 1 lands on the parabola.
alpha 100 has gone too far and is nearly a straight line.

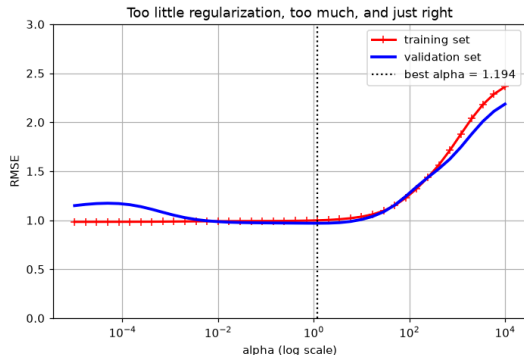
Discuss

We handed the model **15 features it did not need**, and regularization let it decline to use them. Nobody told it the answer was a parabola.

- **Q4a.** What happens as α gets very large, say a million? Why does the math force it?
- **Q5b.** The alpha 100 curve is underfitting. Which number do you change, and which direction?

Choosing Alpha

Too Little, Too Much, And Just Right



Best alpha about **1.19**, validation RMSE **0.970**. The noise has standard deviation 1.0, so that is about as good as anything can do here.

Discuss

On the far left, training 0.984 and validation 1.149: a gap, which is overfitting.
On the far right, training 2.362 and validation 2.186: both high and close, which is underfitting.

- **Q6a.** Why does the training error curve only ever go **up** as α increases?
- **Q6c.** We chose α by whatever scored best on validation. Day 1 said the test set gets used once. Why is this allowed?

Lasso and Elastic Net

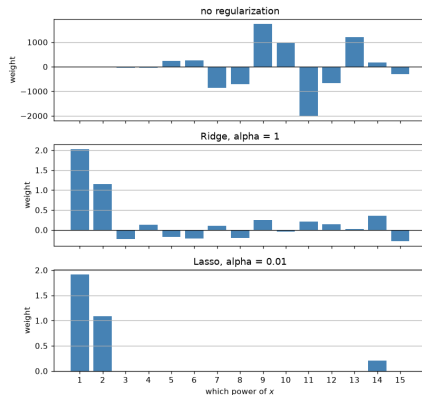
One Small Change To The Penalty

$$\text{minimize} \quad MSE + \alpha \sum_{i=1}^n |w_i|$$

Ridge shrinks all weights toward zero, almost never all the way.

Lasso pushes some weights **exactly** to zero, which drops those features entirely. That is a **sparse** model.

Look At The Top Axis



No regularization: weights in the **thousands**, total 9261. Ridge alpha 1: total 5.5, all 15 kept. Lasso alpha 0.01: total 3.2, keeps **three**.

Discuss

Lasso found the parabola on its own from 140 points, keeping x , x^2 and a little x^{14} . Every Elastic Net setting we tried scored between **0.962 and 0.970**.

- **Q7c.** Square footage, bedrooms, bathrooms. All matter, all related. Ridge or Lasso? Why?
- **Q8a.** A classmate says `l1_ratio = 1.0` is best because its RMSE is lowest. What is wrong with that?

Wrap-Up

What To Take Away

- **Bias** is wrong on average. **Variance** is unstable across datasets.
- α is a hyperparameter, and you pick it by scoring on a validation set.
- **Ridge** keeps everything small. **Lasso** throws features out.
- A dropped feature is **not** evidence the feature does not matter.
- **Scale before you regularize**, or the penalty depends on your units.

Before Next Class

- 1 **Weekly Homework 2, due Sunday 9/13 at 11:59pm.** Covers Day 3 and Day 4.
- 2 Read Geron ch. 2 **Create a Test Set**, and ch. 3 **Measuring Accuracy Using Cross-Validation**.
- 3 Watch the Day 4 video.

Day 4

Today we chose alpha from one validation set of 60 points.

That worked, and it is fragile.

Day 4: cross validation, and then all the ways a train/test split can hand you a number you should not trust.