Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**

# Regression


Jordan Boyd-Graber
University of Colorado Boulder

LECTURE 11

# Content Questions

# Content Questions

# Content Questions

# Content Questions

# Content Questions

# Content Questions

# Content Questions

# Content Questions

# Content Questions

# Content Questions

# Administrivia

- Learnability HW due Friday
- SVM HW next week
- I'm out of town next week

## Project

- Default project or choose your own
- I'll meet with your group early in the week of Oct 12
- Project proposal
- First deliverable due Nov 6: data / baseline

## Plan

Basics

Regularization

Sklearn

## Predictions

| dimension | weight |
|---|---|
| $b$ | 1 |
| $w_1$ | 2.0 |
| $w_2$ | $-1.0$ |
| $\sigma$ | 1.0 |

1. $\mathbf{x}_1 = \{0.0, 0.0\}$; $y_1 =$
2. $\mathbf{x}_2 = \{1.0, 1.0\}$; $y_2 =$
3. $\mathbf{x}_3 = \{.5, 2\}$; $y_3 =$

## Predictions

| dimension | weight |
|---|---|
| $b$ | 1 |
| $w_1$ | 2.0 |
| $w_2$ | $-1.0$ |
| $\sigma$ | 1.0 |

1. $\mathbf{x}_1 = \{0.0, 0.0\}$; $y_1$=1.0
2. $\mathbf{x}_2 = \{1.0, 1.0\}$; $y_2$=
3. $\mathbf{x}_3 = \{.5, 2\}$; $y_3$=

## Predictions

| dimension | weight |
|---|---|
| $b$ | 1 |
| $w_1$ | 2.0 |
| $w_2$ | $-1.0$ |
| $\sigma$ | 1.0 |

1. $\mathbf{x}_1 = \{0.0, 0.0\}$; $y_1 = 1.0$
2. $\mathbf{x}_2 = \{1.0, 1.0\}$; $y_2 = 2.0$
3. $\mathbf{x}_3 = \{.5, 2\}$; $y_3 =$

## Predictions

| dimension | weight |
|---|---|
| $b$ | 1 |
| $w_1$ | 2.0 |
| $w_2$ | $-1.0$ |
| $\sigma$ | 1.0 |

1. $\mathbf{x}_1 = \{0.0, 0.0\}$; $y_1$=1.0
2. $\mathbf{x}_2 = \{1.0, 1.0\}$; $y_2$=2.0
3. $\mathbf{x}_3 = \{.5, 2\}$; $y_3$=0.0

## Probabilities

| dimension | weight |
|---|---|
| $w_0$ | 1 |
| $w_1$ | 2.0 |
| $w_2$ | $-1.0$ |
| $\sigma$ | 1.0 |

$$p(y \mid x) = y \sim N\left(b + \sum_{j=1}^{p} w_j x_j, \sigma^2\right)$$

$$p(y \mid x) = \frac{\exp\left\{-\frac{(y-\hat{y})^2}{2}\right\}}{\sqrt{2\pi}}$$

1. $p(y_1 = 1 \mid \mathbf{x}_1 = \{0.0, 0.0\}) =$
2. $p(y_2 = 3 \mid \mathbf{x}_2 = \{1.0, 1.0\}) =$
3. $p(y_3 = -1 \mid \mathbf{x}_3 = \{.5, 2\}) =$

## Probabilities

| dimension | weight |
|---|---|
| $w_0$ | 1 |
| $w_1$ | 2.0 |
| $w_2$ | $-1.0$ |
| $\sigma$ | 1.0 |

$$p(y \mid x) = y \sim N\left(b + \sum_{j=1}^{p} w_j x_j, \sigma^2\right)$$

$$p(y \mid x) = \frac{\exp\left\{-\frac{(y-\hat{y})^2}{2}\right\}}{\sqrt{2\pi}}$$

1. $p(y_1 = 1 \mid \mathbf{x}_1 = \{0.0, 0.0\}) =$ 0.399
2. $p(y_2 = 3 \mid \mathbf{x}_2 = \{1.0, 1.0\}) =$
3. $p(y_3 = -1 \mid \mathbf{x}_3 = \{.5, 2\}) =$

## Probabilities

| dimension | weight |
|---|---|
| $w_0$ | 1 |
| $w_1$ | 2.0 |
| $w_2$ | $-1.0$ |
| $\sigma$ | 1.0 |

$$p(y \mid x) = y \sim N\left(b + \sum_{j=1}^{p} w_j x_j, \sigma^2\right)$$

$$p(y \mid x) = \frac{\exp\left\{-\frac{(y-\hat{y})^2}{2}\right\}}{\sqrt{2\pi}}$$

1. $p(y_1 = 1 \mid \mathbf{x}_1 = \{0.0, 0.0\}) =$ 0.399
2. $p(y_2 = 3 \mid \mathbf{x}_2 = \{1.0, 1.0\}) =$ 0.242
3. $p(y_3 = -1 \mid \mathbf{x}_3 = \{.5, 2\}) =$

## Probabilities

| dimension | weight |
|---|---|
| $w_0$ | 1 |
| $w_1$ | 2.0 |
| $w_2$ | $-1.0$ |
| $\sigma$ | 1.0 |

$$p(y \mid x) = y \sim N\left(b + \sum_{j=1}^{p} w_j x_j, \sigma^2\right)$$

$$p(y \mid x) = \frac{\exp\left\{-\frac{(y-\hat{y})^2}{2}\right\}}{\sqrt{2\pi}}$$

1. $p(y_1 = 1 \mid \mathbf{x}_1 = \{0.0, 0.0\}) =$ 0.399
2. $p(y_2 = 3 \mid \mathbf{x}_2 = \{1.0, 1.0\}) =$ 0.242
3. $p(y_3 = -1 \mid \mathbf{x}_3 = \{.5, 2\}) =$ 0.242

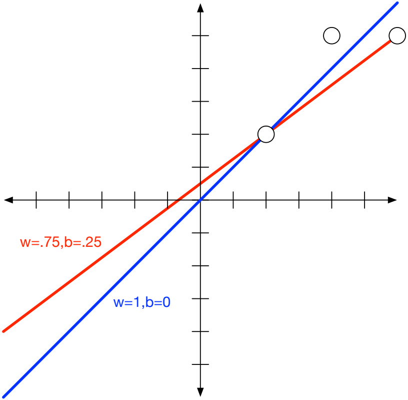
## Plan

Basics

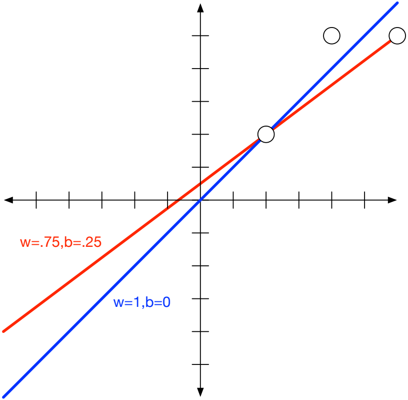
Regularization

Sklearn

## Consider these points and data

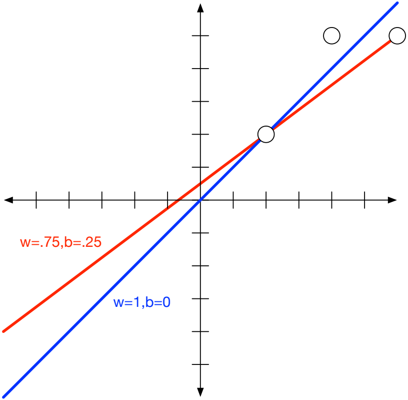


## Consider these points and data



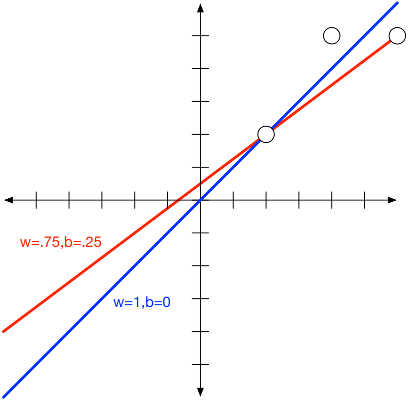

Which is the better OLS solution?

## Consider these points and data



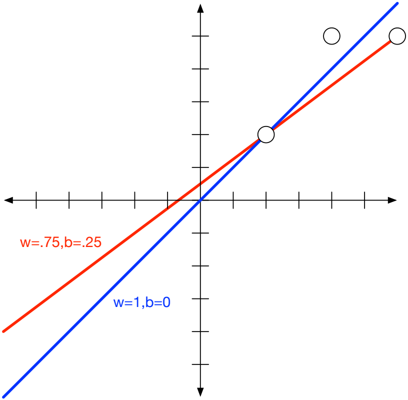

Blue! It has lower RSS.

## Consider these points and data



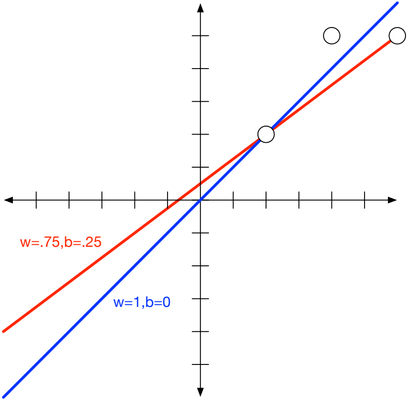

What is the RSS of the better solution?

## Consider these points and data



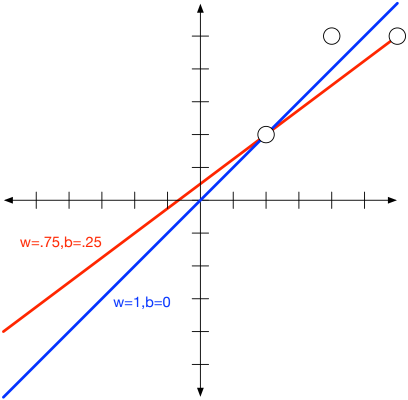

$$\frac{1}{2}\sum_i r_i^2 = \frac{1}{2}\left((1-1)^2 + (2.5-2)^2 + (2.5-3)^2\right) = \frac{1}{4}$$

## Consider these points and data



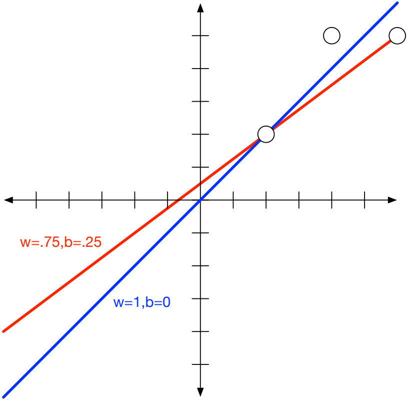

What is the RSS of the red line?

## Consider these points and data




$$\frac{1}{2}\sum_i r_i^2 = \frac{1}{2}\left((1-1)^2 + (2.5-1.75)^2 + (2.5-2.5)^2\right) = \frac{3}{8}$$

## Consider these points and data




For what $\lambda$ does the blue line have a better regularized solution with $L_2$ and $L_1$?

## When Regularization Wins

**$L_2$**

**$L_1$**

## When Regularization Wins

### $L_2$

$$\mathrm{RSS}(x, y, w) + \lambda \sum_d w_d^2 > \mathrm{RSS}(x, y, w) + \lambda \sum_d w_d^2$$

### $L_1$

## When Regularization Wins

$L_2$

$$\mathrm{RSS}(x, y, w) + \lambda \sum_d w_d^2 > \mathrm{RSS}(x, y, w) + \lambda \sum_d w_d^2$$

$$\frac{1}{4} + \lambda 1 > \frac{3}{8} + \lambda \frac{9}{16}$$

$L_1$

## When Regularization Wins

### $L_2$

$$\frac{1}{4} + \lambda 1 > \frac{3}{8} + \lambda \frac{9}{16}$$
$$\frac{7}{16}\lambda > \frac{1}{8}$$

### $L_1$

## When Regularization Wins

### $L_2$

$$\frac{7}{16}\lambda > \frac{1}{8}$$
$$\lambda > \frac{2}{7}$$

### $L_1$

## When Regularization Wins

### $L_2$

$$\lambda > \frac{2}{7}$$

### $L_1$

$$\text{RSS}(x, y, w) + \lambda \sum_d |w_d| > \text{RSS}(x, y, w) + \lambda \sum_d |w_d|$$

## When Regularization Wins

### $L_2$

$$\lambda > \frac{2}{7}$$

### $L_1$

$$\text{RSS}(x, y, w) + \lambda \sum_d |w_d| > \text{RSS}(x, y, w) + \lambda \sum_d |w_d|$$

$$\frac{1}{4} + \lambda 1 > \frac{3}{8} + \lambda \frac{3}{4}$$

## When Regularization Wins

**$L_2$**

$$\lambda > \frac{2}{7}$$

**$L_1$**

$$\frac{1}{4} + \lambda 1 > \frac{3}{8} + \lambda \frac{3}{4}$$

## When Regularization Wins

### $L_2$

$$\lambda > \frac{2}{7}$$

### $L_1$

$$\frac{1}{4}\lambda > \frac{1}{8}$$

## When Regularization Wins

**$L_2$**

$$\lambda > \frac{2}{7}$$

**$L_1$**

$$\frac{1}{4}\lambda > \frac{1}{8}$$
$$\lambda > \frac{1}{2}$$

## When Regularization Wins

**$L_2$**

$$\lambda > \frac{2}{7}$$

**$L_1$**

$$\lambda > \frac{1}{2}$$

Bigger $\lambda$: preference for lower weights $w$

## Plan

Basics

Regularization

Sklearn

## MPG Dataset

- Predict mpg from features of a car
  1. Number of cylinders
  2. Displacement
  3. Horsepower
  4. Weight
  5. Acceleration
  6. Year
  7. Country (ignore this)

## Simple Regression

If $w = 0$, what's the intercept?

## Simple Regression

If $w = 0$, what's the intercept?
23.4

## Simple Linear Regression

What are the coefficients for OLS?

## Simple Linear Regression

What are the coefficients for OLS?

| | |
|---|---|
| cyl | −0.329859 |
| dis | 0.007678 |
| hp | −0.000391 |
| wgt | −0.006795 |
| acl | 0.085273 |
| yr | 0.753367 |

## Simple Linear Regression

What are the coefficients for OLS?

| | |
|---|---|
| cyl | $-0.329859$ |
| dis | $0.007678$ |
| hp | $-0.000391$ |
| wgt | $-0.006795$ |
| acl | $0.085273$ |
| yr | $0.753367$ |

Intercept: -14.5

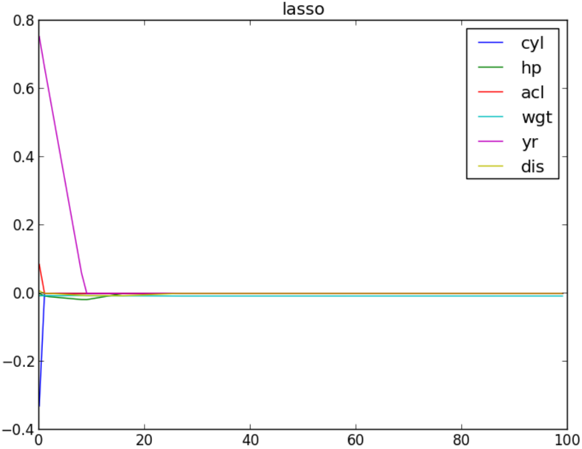
## Simple Linear Regression

```
from sklearn import linear_model
linear_model.LinearRegression()
fit = model.fit(x, y)
```

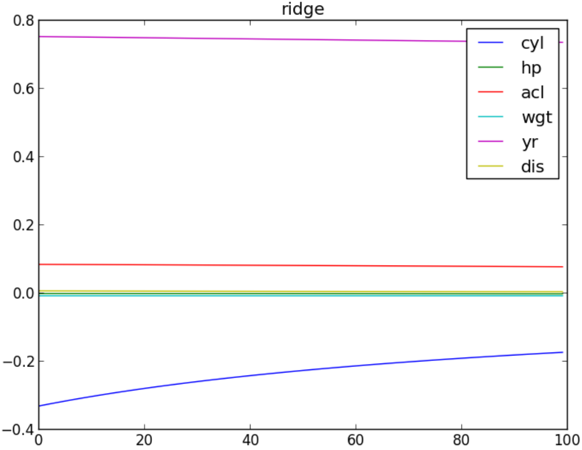
## Lasso

- As you increase the weight of alpha, what feature dominates?
- What happens to the other features?

# Weight is Everything



## How is ridge different?



## Regression isn't special

- Feature engineering
- Regularization
- Overfitting
- Development / Test Data