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

<table>
<thead>
<tr>
<th>dimension</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b )</td>
<td>1</td>
</tr>
<tr>
<td>( w_1 )</td>
<td>2.0</td>
</tr>
<tr>
<td>( w_2 )</td>
<td>-1.0</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>1.0</td>
</tr>
</tbody>
</table>

1. \( x_1 = \{0.0, 0.0\}; \ y_1 = \)
2. \( x_2 = \{1.0, 1.0\}; \ y_2 = \)
3. \( x_3 = \{.5, 2\}; \ y_3 = \)
## Predictions

<table>
<thead>
<tr>
<th>dimension</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>1</td>
</tr>
<tr>
<td>$w_1$</td>
<td>2.0</td>
</tr>
<tr>
<td>$w_2$</td>
<td>-1.0</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.0</td>
</tr>
</tbody>
</table>

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 = \{0.5, 2\}; \ y_3 = \_\_\_\_\_\_

$\sigma_{\mathbf{x}}$
## Predictions

<table>
<thead>
<tr>
<th>dimension</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>1</td>
</tr>
<tr>
<td>$w_1$</td>
<td>2.0</td>
</tr>
<tr>
<td>$w_2$</td>
<td>$-1.0$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.0</td>
</tr>
</tbody>
</table>

1. $x_1 = \{0.0, 0.0\}; y_1 = 1.0$
2. $x_2 = \{1.0, 1.0\}; y_2 = 2.0$
3. $x_3 = \{0.5, 2\}; y_3 =$
Predictions

<table>
<thead>
<tr>
<th>dimension</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>1</td>
</tr>
<tr>
<td>$w_1$</td>
<td>2.0</td>
</tr>
<tr>
<td>$w_2$</td>
<td>$-1.0$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.0</td>
</tr>
</tbody>
</table>

1. $x_1 = \{0.0, 0.0\}; \ y_1=1.0$
2. $x_2 = \{1.0, 1.0\}; \ y_2=2.0$
3. $x_3 = \{.5, 2\}; \ y_3=0.0$
### Probabilities

<table>
<thead>
<tr>
<th>dimension</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_0$</td>
<td>1</td>
</tr>
<tr>
<td>$w_1$</td>
<td>2.0</td>
</tr>
<tr>
<td>$w_2$</td>
<td>−1.0</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.0</td>
</tr>
</tbody>
</table>

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

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

1. \( p(y_1 = 1 \mid x_1 = \{0.0, 0.0\}) = \)
2. \( p(y_2 = 3 \mid x_2 = \{1.0, 1.0\}) = \)
3. \( p(y_3 = -1 \mid x_3 = \{.5, 2\}) = \)
Probabilities

<table>
<thead>
<tr>
<th>dimension</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_0$</td>
<td>1</td>
</tr>
<tr>
<td>$w_1$</td>
<td>2.0</td>
</tr>
<tr>
<td>$w_2$</td>
<td>−1.0</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.0</td>
</tr>
</tbody>
</table>

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

\[
p(y | x) = \exp \left\{ -\frac{(y - \hat{y})^2}{2} \right\}
\]

\[
p(y | x) = \frac{1}{\sqrt{2\pi}}
\]

1. $p(y_1 = 1 | x_1 = \{0.0, 0.0\}) = 0.399$
2. $p(y_2 = 3 | x_2 = \{1.0, 1.0\}) = \ldots$
3. $p(y_3 = -1 | x_3 = \{.5, 2\}) = \ldots$
## Probabilities

<table>
<thead>
<tr>
<th>dimension</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_0$</td>
<td>1</td>
</tr>
<tr>
<td>$w_1$</td>
<td>2.0</td>
</tr>
<tr>
<td>$w_2$</td>
<td>-1.0</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.0</td>
</tr>
</tbody>
</table>

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

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

1. $p(y_1 = 1 \mid x_1 = \{0.0, 0.0\}) = 0.399$
2. $p(y_2 = 3 \mid x_2 = \{1.0, 1.0\}) = 0.242$
3. $p(y_3 = -1 \mid x_3 = \{.5, 2\}) =$
Probabilities

<table>
<thead>
<tr>
<th>dimension</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_0$</td>
<td>1</td>
</tr>
<tr>
<td>$w_1$</td>
<td>2.0</td>
</tr>
<tr>
<td>$w_2$</td>
<td>−1.0</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.0</td>
</tr>
</tbody>
</table>

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

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

1. $p(y_1 = 1 \mid x_1 = \{0.0, 0.0\}) = 0.399$
2. $p(y_2 = 3 \mid x_2 = \{1.0, 1.0\}) = 0.242$
3. $p(y_3 = -1 \mid x_3 = \{0.5, 2\}) = 0.242$
Plan

Basics

Regularization

Sklearn
Consider these points and data

\[ w = 1, b = 0 \]
\[ w = 0.75, b = 0.25 \]
Consider these points and data

\[ w = 1, b = 0 \]
\[ w = 0.75, b = 0.25 \]

Which is the better OLS solution?
Consider these points and data

Blue! It has lower RSS.

\[ w=1, b=0 \]
\[ w=0.75, b=0.25 \]
Consider these points and data

\[ w=1, b=0 \]
\[ w=0.75, b=0.25 \]

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

$L_2$

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

$L_1$
When Regularization Wins

$L_2$

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

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

$L_1$
### When Regularization Wins

<table>
<thead>
<tr>
<th>$L_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{4} + \lambda_1 &gt; \frac{3}{8} + \lambda \frac{9}{16}$</td>
</tr>
<tr>
<td>$\frac{7}{16}\lambda &gt; \frac{1}{8}$</td>
</tr>
</tbody>
</table>

| $L_1$ |
When Regularization Wins

\[ L_2 \]

\[
\lambda > \frac{1}{8} \quad \frac{7}{16} \\
\lambda > \frac{2}{7}
\]

\[ L_1 \]
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 $w$
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?
If $w = 0$, what’s the intercept?
23.4
Simple Linear Regression

What are the coefficients for OLS?

<table>
<thead>
<tr>
<th>Feature</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>-0.329859</td>
</tr>
<tr>
<td>y</td>
<td>0.753367</td>
</tr>
<tr>
<td>l</td>
<td>-0.007678</td>
</tr>
<tr>
<td>d</td>
<td>-0.000391</td>
</tr>
<tr>
<td>i</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td></td>
</tr>
<tr>
<td>0.000391</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
</tr>
<tr>
<td>w</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>0.085273</td>
</tr>
<tr>
<td>c</td>
<td></td>
</tr>
<tr>
<td>l</td>
<td></td>
</tr>
<tr>
<td>0.085273</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
</tr>
<tr>
<td>0.006795</td>
<td></td>
</tr>
<tr>
<td>Intercepts: -14.5</td>
<td></td>
</tr>
</tbody>
</table>
Simple Linear Regression

What are the coefficients for OLS?

\[
\begin{align*}
\text{cyl} & \quad -0.329859 \\
\text{dis} & \quad 0.007678 \\
\text{hp} & \quad -0.000391 \\
\text{wgt} & \quad -0.006795 \\
\text{acl} & \quad 0.085273 \\
\text{yr} & \quad 0.753367
\end{align*}
\]
Simple Linear Regression

What are the coefficients for OLS?

- \text{cyl} = -0.329859
- \text{dis} = 0.007678
- \text{hp} = -0.000391
- \text{wgt} = -0.006795
- \text{acl} = 0.085273
- \text{yr} = 0.753367

Intercept: -14.5
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

\[ \text{mpg} = 46 - 0.01 \text{ Weight} \]
How is ridge different?

![Graph showing how ridge differs](image)
Regression isn’t special

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