# Regression

Computational Linguistics: Jordan Boyd-Graber & Philip Resnik

University of Maryland

September 21, 2021

# Content Questions

# Linear Regression 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. $\mathbf{x}_3 = \{.5, 2\}$; $y_3$=

# Linear Regression 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. $\mathbf{x}_3 = \{.5, 2\}$; $y_3$=

# Linear Regression 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. $\mathbf{x}_3 = \{.5, 2\}$; $y_3$=

# Linear Regression 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 |

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

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

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

# Probabilities

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

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

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

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

# 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

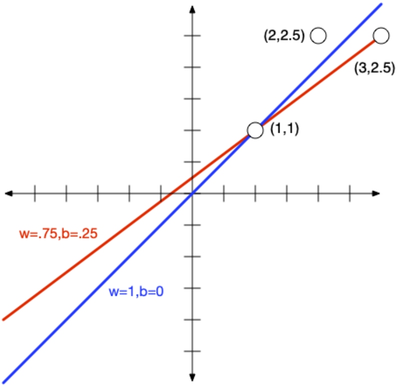
| dimension | weight |
|---|---|
| $w_0$ | 1 |
| $w_1$ | 2.0 |
| $w_2$ | $-1.0$ |
| $\sigma$ | 1.0 |

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

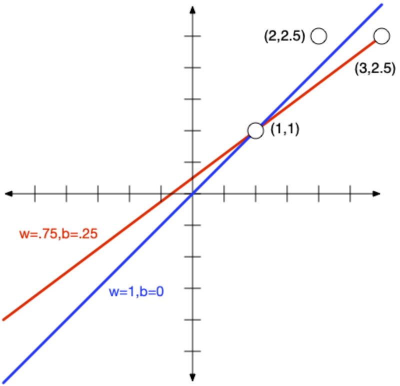
$$p(y \mid x) = \frac{\exp\left\{-\frac{(y-\hat{y})^2}{2}\right\}}{\sqrt{2\pi}}$$

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

# Consider these points and data

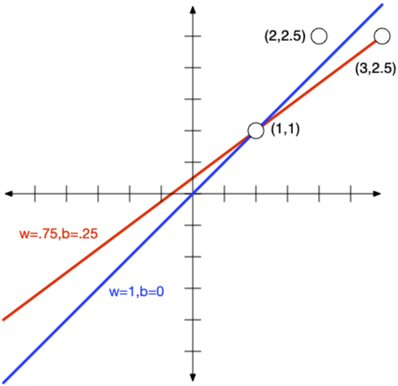


# Consider these points and data



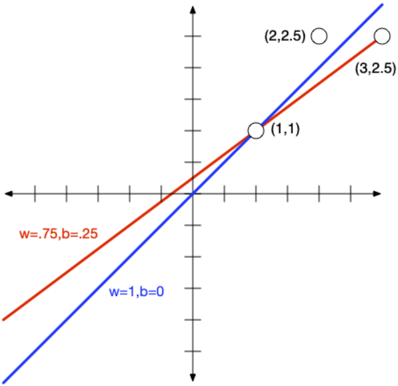

Which is the better OLS solution?

# Consider these points and data



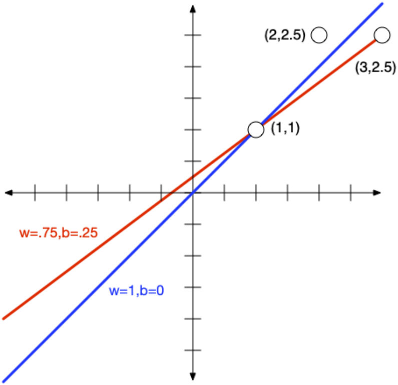

Blue! It has lower RSS.

# Consider these points and data



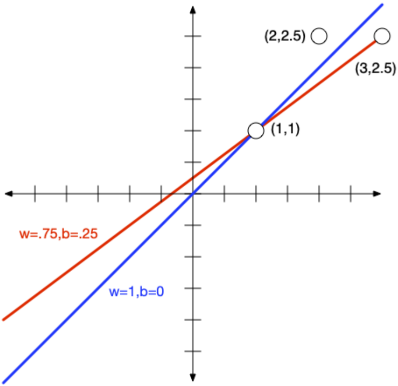

What is the RSS of the better solution?

# Consider these points and data



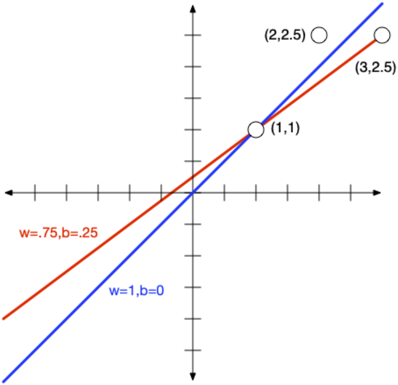

$\frac{1}{2}\sum_i r_i^2 = \frac{1}{2}\left((1-1)^2+(2.5-2)^2+(2.5-3)^2\right) = \frac{1}{4}$

# Consider these points and data




What is the RSS of the red line?

# Consider these points and data




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