Regularization

Afternoon

Goals

- Be able to…
  - explain ridge regression and lasso regression
  - tune the bias/variance of a regression model
  - choose the best regularization hyperparameter for regression

Issues with Ordinary Linear Regression

- High dimensions -> high variance
  - High variance -> overfitting -> :(…
  - And yet we may want to include dimensions/ features/interactions, if they're helpful

Ordinary Linear Regression

Find betas to minimize RSS/loss function:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2$$

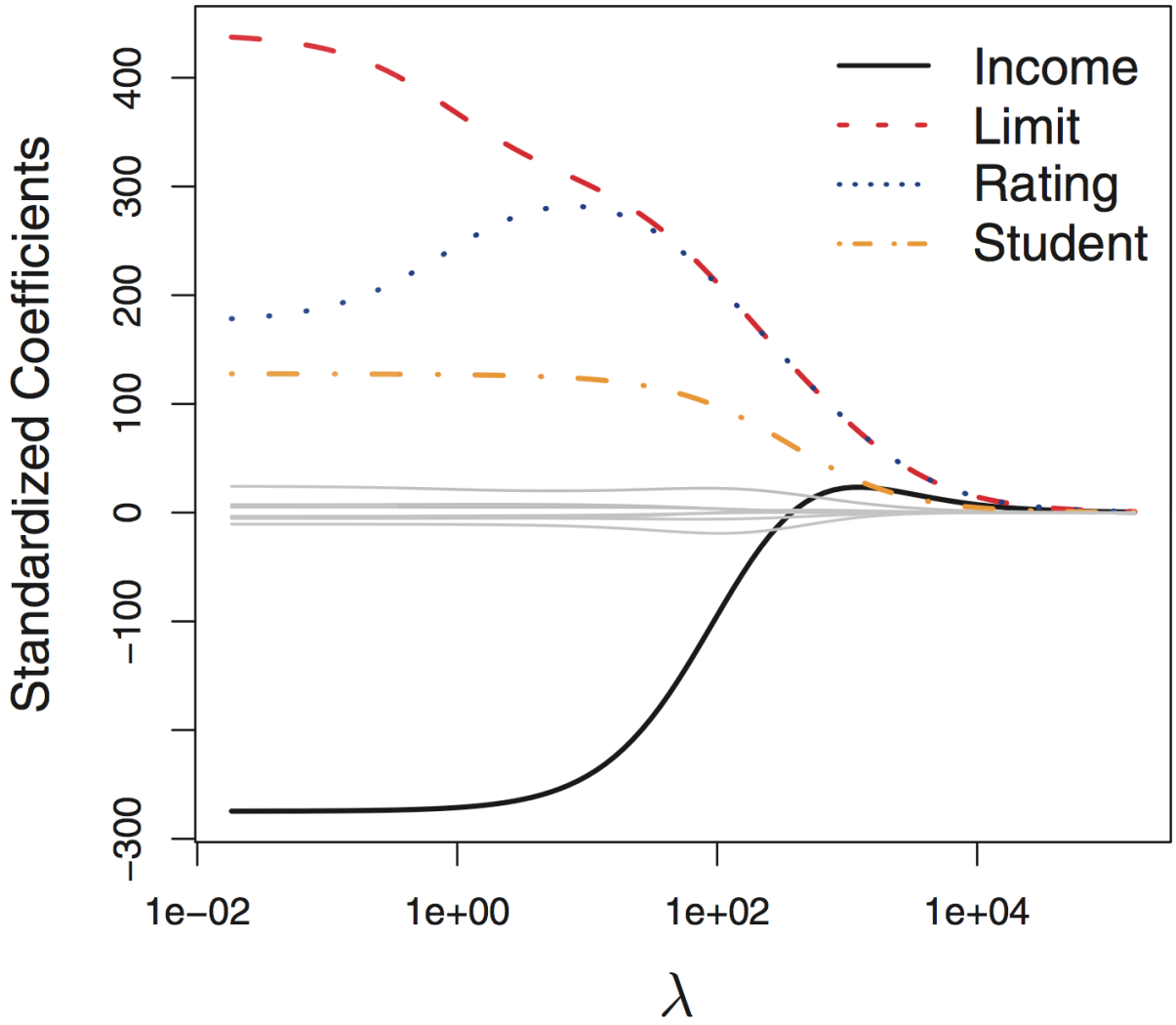
Ridge Regression

Find betas to minimize loss function:

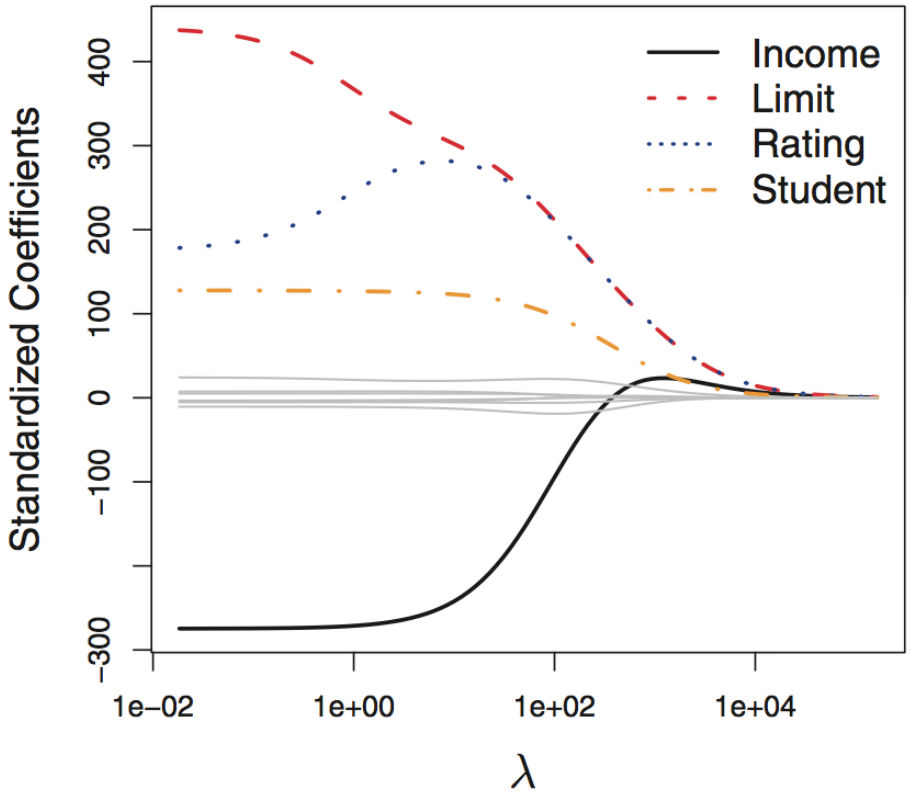
$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2$$

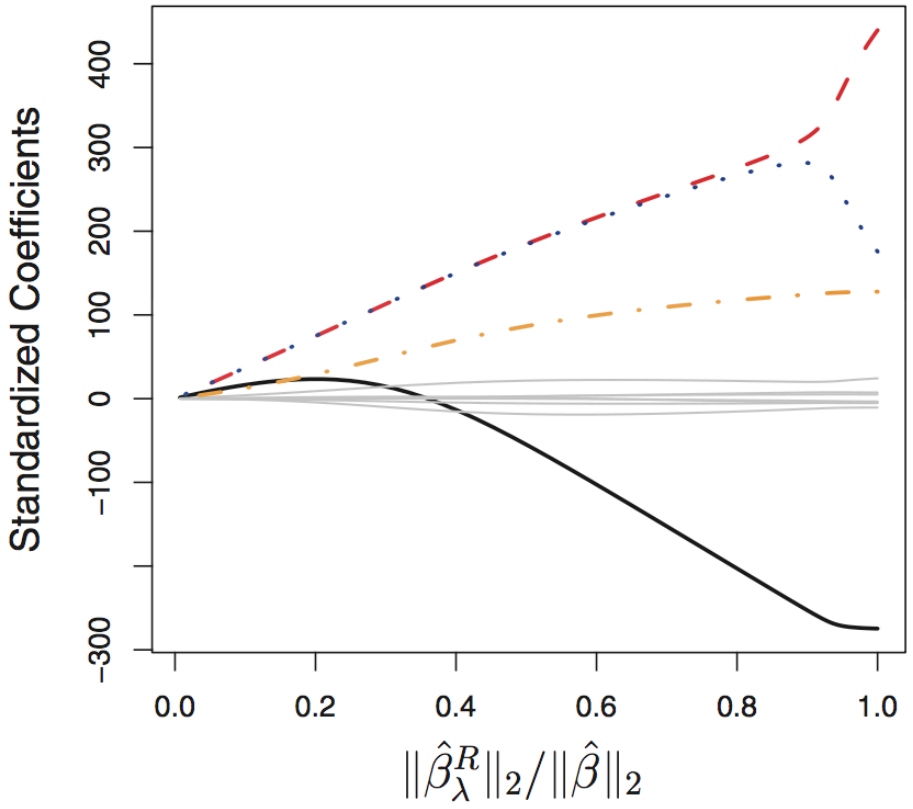
your friend, the hyperparameter

Ridge Regression



Ridge Regression





l2 ("ell two") norm: $\|\beta\|_2 = \sqrt{\sum_{j=1}^{p} \beta_j{}^2}$

x axis on the right: amount that the ridge regression coefficient estimates have been shrunken towards zero; a small value indicates that they have been shrunken very close to zero

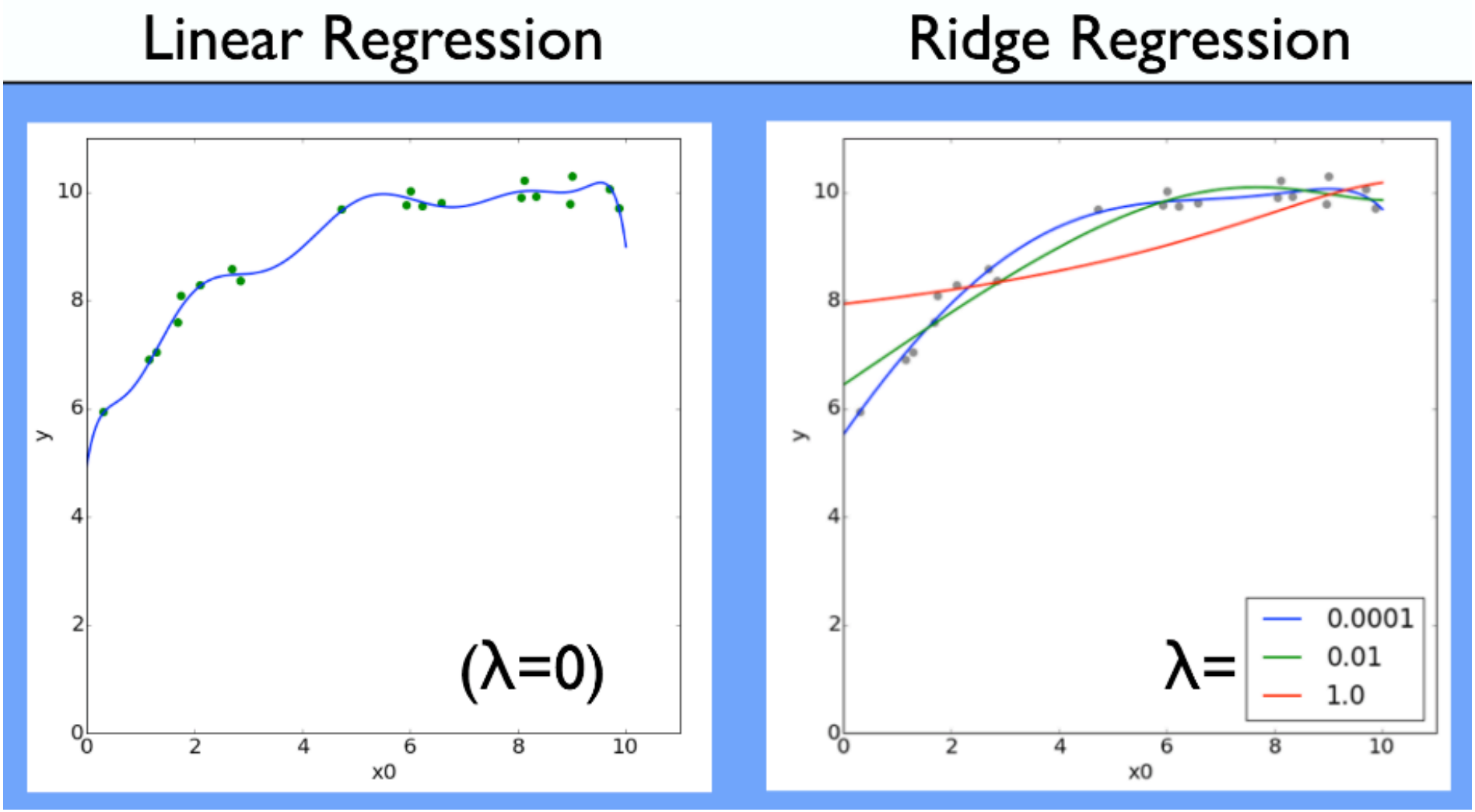
Ridge Regression

Warning: when using ridge regression, scale matters!
Why? (Consider units to measure salary)

Standardize your predictors:

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}}$$

Ridge Regression



Lasso Regression

Find betas to minimize loss function:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|$$

your friend, the hyperparameter

remember, ridge:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2$$

Ridge and Lasso

Ridge:

minimize RSS:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2$$

l2 ("ell two") norm:

$\|\beta\|_2 = \sqrt{\sum_{j=1}^{p}{\beta_j}^2}$

Lasso:

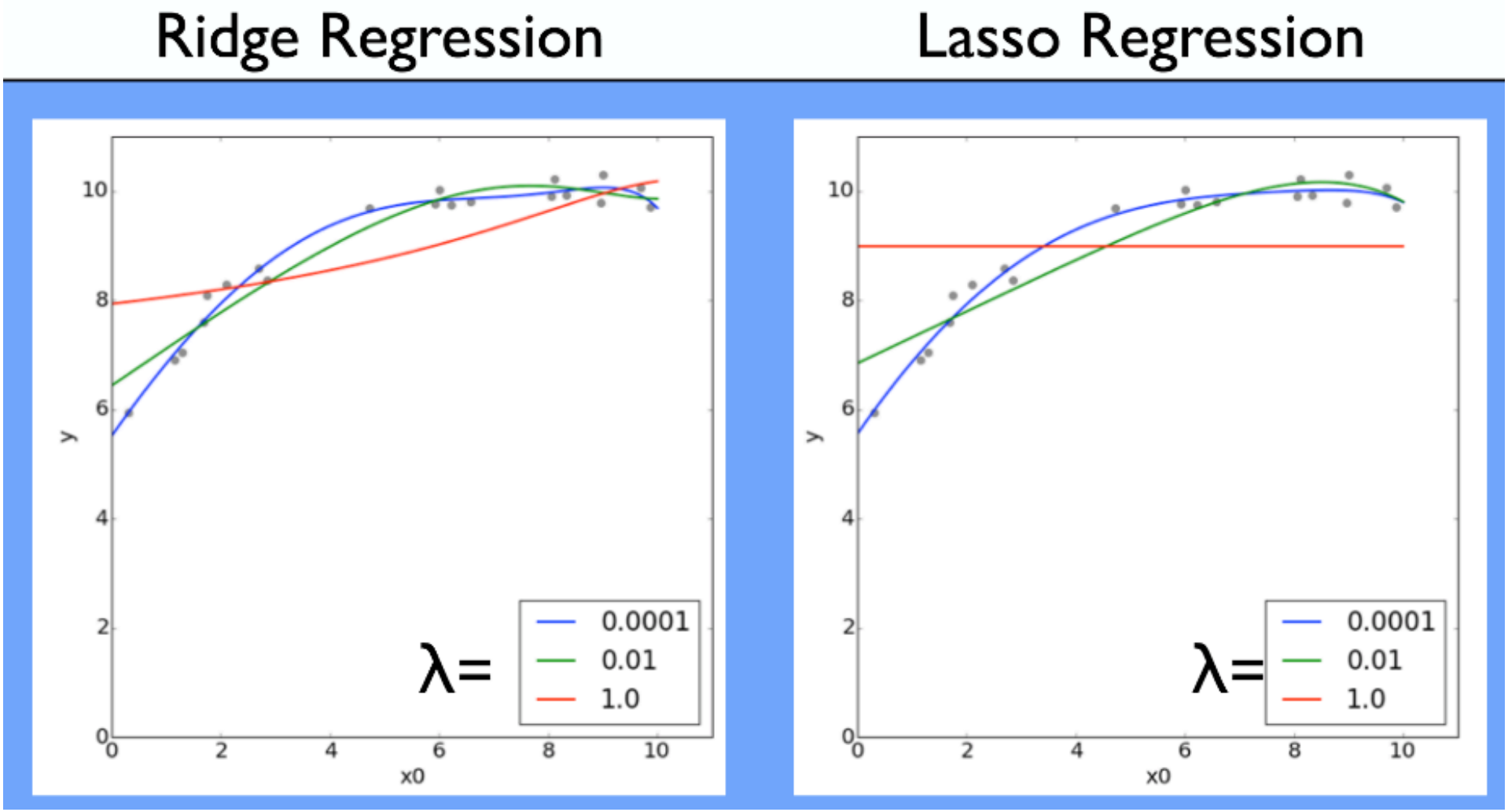
minimize RSS:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|$$

l1 ("ell one") norm:

$\|\beta\|_1 = \sum|\beta_j|$

Ridge and Lasso



Ridge vs. Lasso