Regularization: LASSO and Ridge Regression

Variable Selection

Variable Selection

Reasons to dislike forward/backward/subset selection:

- Computationally intensive - too many variables!
- Hard to use cross-validation; have to choose a penalized metric.
- "Best subset" is almost never feasible; forward/backward might miss a better option!

Regularization

Instead: We adjust our loss function that we use to estimate coefficients.

Ordinary Linear Regression:

minimize squared error:

sum of (predicted - truth)^2

Regularization

We would like to make it "harder" to allow variables into the model.

Regularized Regression:

minimize squared error plus penalty:

sum of (predicted - truth)^2 + (penalty on betas)

LASSO

The LASSO (least absolute shrinkage and selection operator) says "big coefficients are bad"

$$\sum(\hat{y}_i - y_i)^2 + \lambda \sum |\beta_j|$$

RSS + (penalty)*(sum of coefficients)

LASSO

```
lasso_spec <- linear_reg(penalty = 0.1, mixture = 1) %>%
  set_engine("glmnet") %>%
  set_mode("regression")
```

penalty: λ mixture: We use the absolute value of the betas.

Ridge Regression

Ridge Regression says "big coefficients are bad, and bigger coefficients are REALLY bad"

$$\sum(\hat{y}_i - y_i)^2 + \lambda \sum \beta_j^2$$

RSS + (penalty)*(sum of coefficients squared)

Ridge Regression

```r
ridge_spec <- linear_reg(penalty = 0.1, mixture = 0) %>%
  set_engine("glmnet") %>%
  set_mode("regression")
```

Try it!

Open Activity-Variable-Selection

Fit a LASSO model to the cannabis data with lambda = 0.1. Then fit one with lambda = 0.5. What is different?

Fit a Ridge Regression model to the cannabis data with lambda = 0.1. Then fit one with lambda = 0.5. What is different?

Which model do you prefer?

(Bonus) What is the best choice of lambda?