

HW3

Jay Lee

September 24, 2017

6.2

- (a) iii, because the λ term constrains the model (making it less flexible), increasing its potential bias at the benefit of sharply decreasing variance.
- (b) iii, for the same reasons as part (a).
- (c) ii, because fitting a non-linear model allows the model to follow the data more closely, reducing potential bias at the expense of an increase in variance.

6.3

- (a) iv, because OLS provides the minimum RSS possible in linear regression, and increasing s would decrease RSS from the naive model ($y_i = \bar{y}$) to OLS.
- (b) ii, because as we increase s through the model with the best test fit our test RSS will be at a low, then as the training model gets more complex our test RSS will increase again.
- (c) iii, because at $s = 0$ our model will never change based on different training data (zero variance), and as $s \rightarrow \infty$ our model will be as variant as OLS.
- (d) iv, because at $s = 0$ our model is as biased as possible (assuming $y_i = \bar{y}$), and as $s \rightarrow \infty$ our model makes no assumptions other than a linear form.
- (e) v, because the irreducible error doesn't change with the model (is non-reducible).

6.4

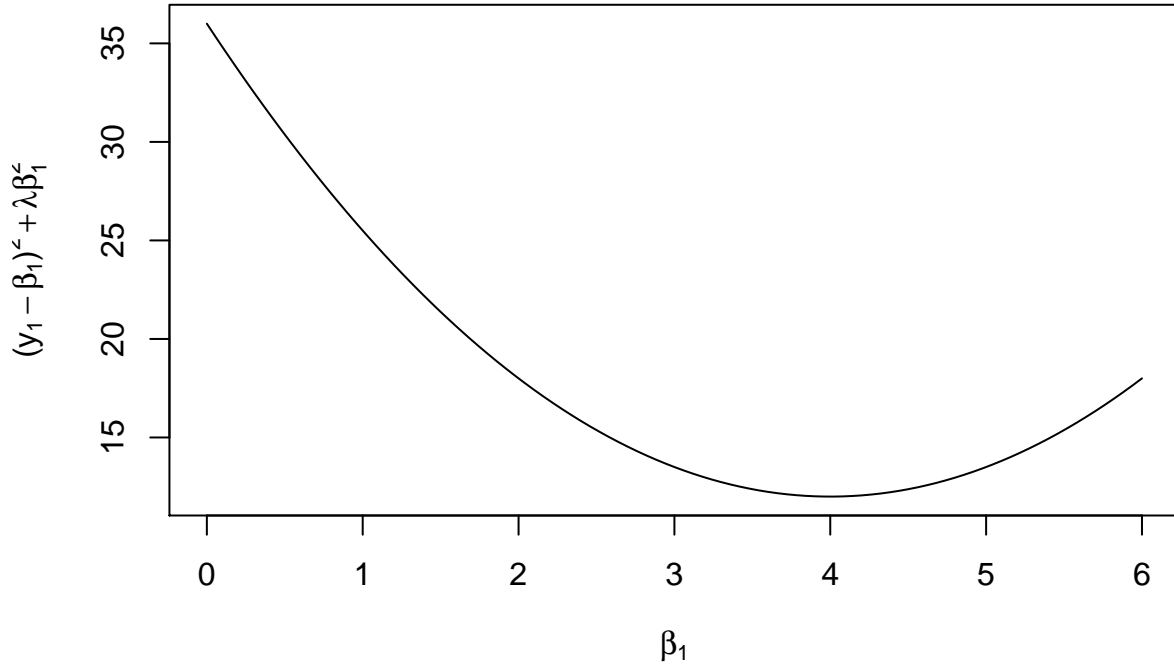
- (a) iv, because increasing λ constrains our model more, making it less flexible to fit the training data.
- (b) ii, because at first our model will grow less flexible and better fit test data, then as we pass some cutoff our bias will increase more and more even past the test data.
- (c) iv, because increasing λ constrains our model more, making it less variable to changes in training data.
- (d) iii, because an OLS model is less biased than any constrained model where we assume a penalty on the coefficients.
- (e) v, because the irreducible error doesn't change with the model (is non-reducible).

6.6

- (a) 6.12, with $p = 1$: Minimize

$$(y_1 - \beta_1)^2 + \lambda \beta_1^2.$$

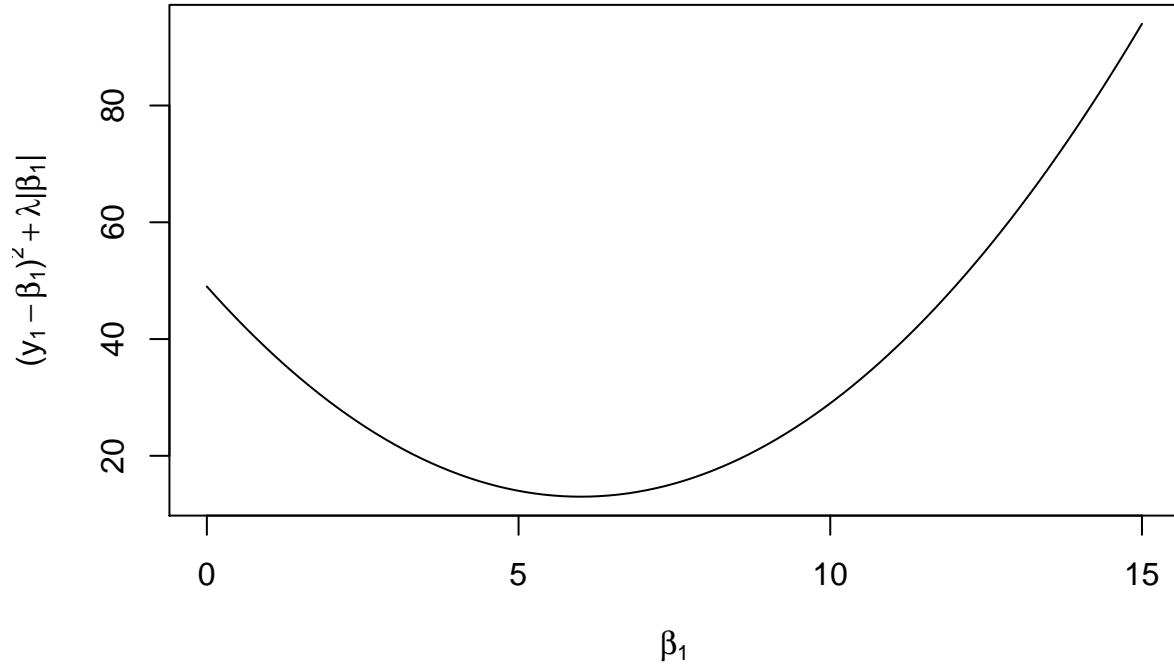
Because this quantity is always positive, it forms an turned-up parabola plotted against β_1 . With $y_1 = 6$ and $\lambda = 0.5$, we minimize at $\beta_1 = 4 = y_1/(1 + \lambda)$.



(b) 6.13, with $p = 1$: Minimize

$$(y_1 - \beta_1)^2 + \lambda|\beta_1|.$$

With $y_1 = 7$ and $\lambda = 2$, we minimize at $\beta_1 = 6 = y_1 - \lambda/2$.



6.5

(a) Let $x_{11} = x_{12} = a$, $x_{21} = x_{22} = -a$, $y_1 = b$, and $y_2 = -b$. As given, $\beta_0 = 0$. We want to minimize

$$(b - \beta_1 a + \beta_2 a)^2 + (-b - \beta_1 a + \beta_2 a)^2 + \lambda(\beta_1^2 + \beta_2^2).$$

(b) When we let $\beta_1 = \beta_2$, then the expression above simplifies to $2b^2 + \lambda\beta_1^2 + \lambda\beta_2^2$. For any choice of $\beta_1 \neq \beta_2$, the difference in $\beta_1 a - \beta_2 a$ does not cancel out in the first two terms. So, the expression is minimized when $\beta_1 = \beta_2$.

(c) We want to minimize

$$(b - \beta_1 a + \beta_2 a)^2 + (-b - \beta_1 a + \beta_2 a)^2 + \lambda|\beta_1| + \lambda|\beta_2|.$$

(d) The above expression simplifies to ?????